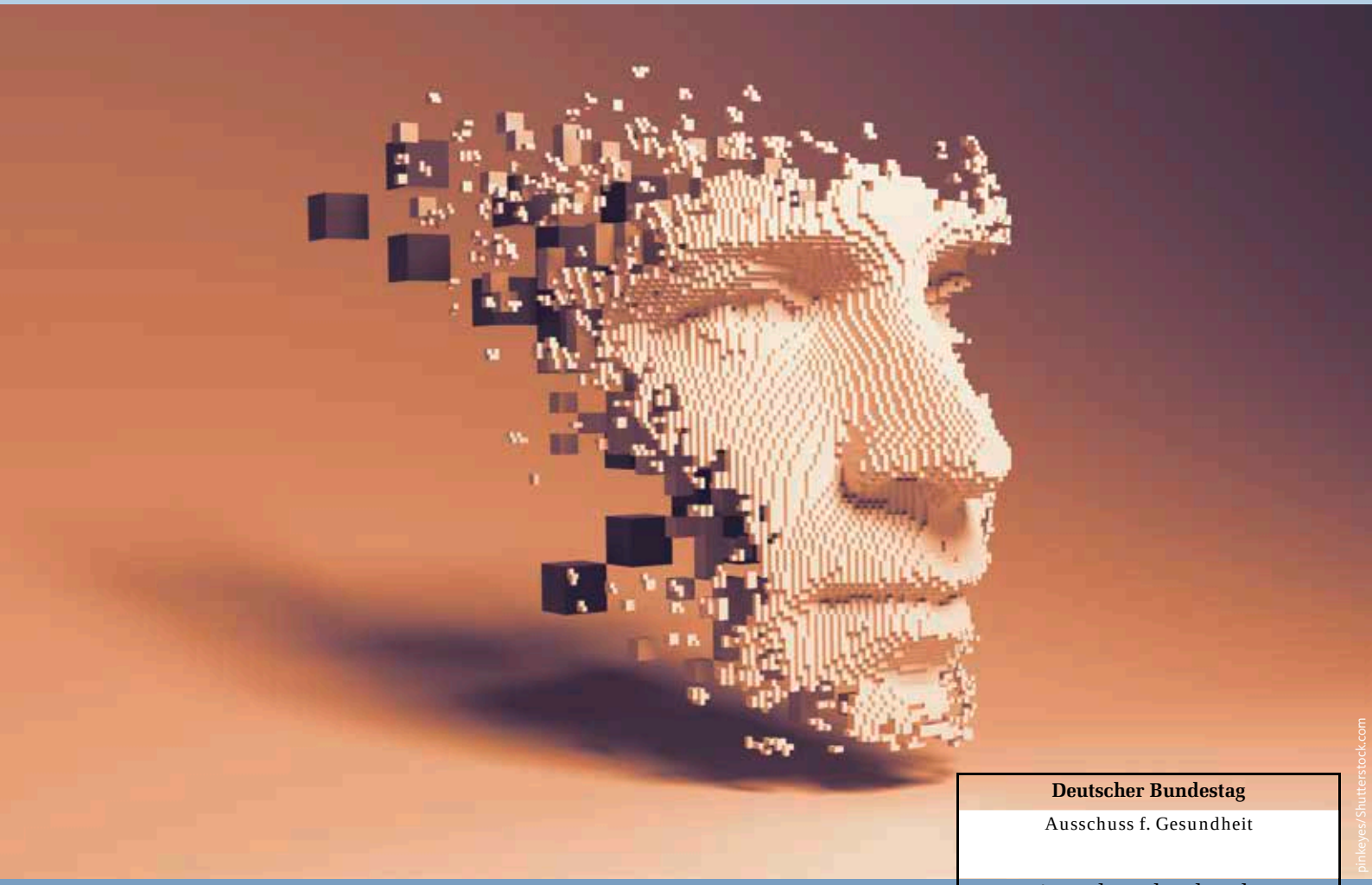
Anlagenkonvolut zum Kurzprotokoll der 77. Sitzung des Ausschusses für Gesundheit am 5. Juli 2023

Deutscher Ethikrat



pinkeyes/Shutterstock.com

Deutscher Bundestag
Ausschuss f. Gesundheit

Ausschussdrucksache
20(14)125
Zur Sitzung am 5. Juli 2023

21.06.2023

Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz

STELLUNGNAHME

20. März 2023

+++ Vorabfassung! Wird durch die lektorierte Fassung ersetzt. +++

Deutscher Ethikrat
Jägerstraße 22/23
D-10117 Berlin

Telefon: +49/30/20370-242
Telefax: +49/30/20370-252
E-Mail: kontakt@ethikrat.org
Web: www.ethikrat.org

Danksagung

An dieser Stellungnahme haben zahlreiche Personen unterstützend mitgewirkt:

Ulrike von Luxburg, Matthias Bethge, Tanja Schultz und Stefan Remy boten zum Projektbeginn bei einer öffentlichen Anhörung am 25. Februar 2021 wertvolle Orientierung zu den aktuellen und absehbaren Entwicklungen im Bereich Künstlicher Intelligenz. Inge Molenaar und Ute Schmid ergänzten dies im Rahmen einer weiteren, nichtöffentlichen Anhörung sowie mit schriftlicher Expertise zum Einsatz von KI in der Schule und Matthias May gewährte hilfreiche Einblicke in die Anwendung algorithmischer Systeme in der medizinischen Diagnostik.

Wolfgang Schulz, Indra Spiecker genannt Döhmann und noch einmal Ute Schmid brachten ihren Sachverstand zu juristischen und informatikspezifischen Passagen der Stellungnahme ein. Carl Friedrich Gethmann begleitete die Arbeit der Stellungnahme auch nach dem Ende seiner Amtszeit im Deutschen Ethikrat im Februar 2021 als nunmehr externes Mitglied der Arbeitsgruppe „Mensch und Maschine“ und auch Andreas Kruse wirkte bis zu seinem Ausscheiden aus dem Ethikrat im März 2022 an der Stellungnahme mit.

Der Deutsche Ethikrat dankt allen Mitwirkenden sehr herzlich für ihre Unterstützung!

+++ Vorabfassung! Wird durch die lektorierte Fassung ersetzt. +++

Inhalt

Zusammenfassung **7**

1 Einleitung **54**

TEIL I: TECHNISCHE UND PHILOSOPHISCHE GRUNDLEGUNGEN

2 Zentrale Entwicklungen und technische Grundlagen Künstlicher Intelligenz **60**

2.1 Historischer Kontext 60

2.2 Künstliche Intelligenz im 21. Jahrhundert: Big Data, Algorithmen und soziotechnische Ökosysteme 64

2.2.1 Digitale Durchdringung der menschlichen Lebenswelt 64

2.2.2 Daten und digitale Infrastrukturen 66

2.2.3 Algorithmen und Datenverarbeitung 68

2.2.4 Einsatzbereiche algorithmischer Systeme und Künstlicher Intelligenz 74

2.2.5 Ethische Leitlinien und regulativer Rahmen für algorithmische Systeme und KI 76

3 Zentrale Begriffe und philosophische Grundlagen **83**

3.1 Künstliche Intelligenz: Begriffliche Analyse 83

3.2 Intelligenz und Vernunft 88

3.2.1 Intelligenz 88

3.2.2 Vernunft 93

3.3 Handlung und Verantwortung 97

3.3.1 Handlung 97

3.3.2 Verantwortung 101

3.4 Anthropologische Aspekte des Mensch-Maschine-Verhältnisses 105

3.4.1 Philosophische Grundbestimmung des Menschseins 105

3.4.2 Der Mensch als Maschine – die Maschine als Mensch? 108

3.4.3 Verleiblichte Vernunft 114

3.5 Fazit 117

4 Mensch-Technik-Relationen **120**

4.1 Einleitung 120

4.2 Technikdeterminismus versus Sozialkonstruktivismus 121

4.3 Mehrstufige Mensch-Technik-Wechselwirkungen 125

4.4 Erweitern und Vermindern menschlicher Autonomie und Autorschaft 130

4.5 Fazit 134

TEIL II: AUSGEWÄHLTE ANWENDUNGEN UND SEKTORSPEZIFISCHE EMPFEHLUNGEN

5 Medizin 140
5.1 Einleitung 140
5.2 Einsatz von. KI-Systemen in der Medizin 141
5.2.1 Entwicklung von KI-Systemen 142
5.2.2 KI in der medizinischen Forschung 144
5.2.3 KI in der medizinischen Versorgung 146
5.3 Fazit und Empfehlungen 159
6 Bildung 163
6.1 Einleitung 163
6.2 Zum Bildungsbegriff 164
6.3 Einsatzmöglichkeiten datenbasierter, KI-gestützter Lehr-Lern-Systeme 167
6.4 Mensch-Maschine-Relationen in der schulischen Bildung: Ersetzen, Erweitern, Vermindern 170
6.5 Grundsätzliche Diskussion: KI im schulischen Bildungsprozess 178
6.6 Fazit und Empfehlungen 182
7 Öffentliche Kommunikation und Meinungsbildung 187
7.1 Einleitung 187
7.2 Das Internet als soziotechnisches System: Funktionsweise Sozialer Medien 188
7.2.1 Neue soziotechnische Infrastrukturen 188
7.2.2 Informationsauswahl und Kuratierung 189
7.2.3 Moderation von Inhalten 192
7.2.4 Auswirkungen auf die Erweiterung und Verminderung menschlicher Handlungsfähigkeiten 195
7.3 Informationsqualität 197
7.3.1 Falschnachrichten und Verschwörungstheorien 198
7.3.2 Filterblasen und Echokammern 200
7.3.3 Moralische und emotionale Aufladung 202
7.3.4 Relevanz der beobachteten Effekte 203
7.4 Diskursqualität 204
7.4.1 Politische Polarisierung 205
7.4.2 Politische Werbung und Manipulation 208
7.4.3 Spannungsfeld Diskursverrohung und Äußerungsfreiheit 212

7.4.4 Erweiternde und vermindernde Rückwirkungen auf den öffentlichen Vernunftgebrauch ... 215
7.5 Fazit und Empfehlungen ... 216
8 Öffentliche Verwaltung ... 223
8.1 Einleitung ... 223
8.2. Ethische Fragen algorithmischer Automatisierung im Verwaltungshandeln ... 224
8.3 Automatische Entscheidungssysteme am Beispiel des Sozialwesens ... 229
8.3.1 Erweiterung professioneller Handlungskompetenz ... 230
8.3.2 Verminderung von Entscheidungskompetenz, Handlungsoptionen und Autorschaft ... 236
8.4 Predictive Policing – KI in der Kriminalitätsbekämpfung ... 241
8.5 Fazit und Empfehlungen ... 247

TEIL III: QUERSCHNITTSTHEMEN UND ÜBERGREIFENDE EMPFEHLUNGEN
9 Zusammenfassung der bisherigen Analyse ... 253
9.1 Anthropologische und ethische Grundorientierung ... 253
9.2 Einsichten aus den Anwendungsfeldern ... 256
10 Entfaltung von Querschnittsthemen und Empfehlungen ... 261
10.1 Querschnittsthema 1: Erweiterung & Verminderung von Handlungsmöglichkeiten ... 261
10.2 Querschnittsthema 2: Wissenserzeugung durch KI und der Umgang mit KI-gestützten Voraussagen ... 263
10.3 Querschnittsthema 3: Die Gefährdung des Individuums durch statistische Stratifizierung ... 265
10.4 Querschnittsthema 4: Auswirkungen von KI auf menschliche Kompetenzen und Fertigkeiten ... 267
10.5 Querschnittsthema 5: Schutz von Privatsphäre und Autonomie versus Gefahren durch Überwachung und Chilling-Effekte ... 270
10.6 Querschnittsthema 6: Datensouveränität und gemeinwohlorientierte Datennutzung ... 272
10.7 Querschnittsthema 7: Kritische Infrastrukturen, Abhängigkeiten und Resilienz ... 276
10.8 Querschnittsthema 8: Pfadabhängigkeiten, Zweitverwertung und Missbrauchgefahren ... 279
10.9 Querschnittsthema 9: Bias und Diskriminierung ... 281
10.10 Querschnittsthema 10: Transparenz und Nachvollziehbarkeit – Kontrolle und Verantwortung ... 284
10.11 Fazit ... 286

Zusammenfassung

Einleitung

1. Digitale Technologien und Künstliche Intelligenz (KI) haben mittlerweile in nahezu allen Bereichen des öffentlichen und privaten Lebens Einzug gehalten. Für die ethische Bewertung solcher Entwicklungen und ihres Einsatzes in verschiedenen Bereichen ist es nötig, nicht nur die Technologien zu verstehen, sondern auch ihre Wechselwirkungen mit den Personen, die sie verwenden oder von ihrer Anwendung betroffen sind. Zentral ist dabei die Frage, welche Auswirkungen damit verbunden sind, wenn Tätigkeiten, welche zuvor Menschen vorbehalten waren, an Maschinen delegiert werden. Werden menschliche Autorschaft und Handlungsmöglichkeiten durch den Einsatz von KI erweitert oder vermindert?

2. In der vorliegenden Stellungnahme geht der Deutsche Ethikrat dieser Frage nach und schreibt damit Themen fort, die bereits in den Stellungnahmen zu Big Data und Gesundheit (2017) sowie Robotik und Pflege (2019) angeschnitten wurden. Der Deutsche Ethikrat reagiert mit der Stellungnahme auf eine im Oktober 2020 vom Präsidenten des Deutschen Bundestages formulierte Bitte, eine multidisziplinäre Stellungnahme zu den ethischen Fragen des Verhältnisses von Mensch und Maschine zu erarbeiten.

3. Die Stellungnahme gliedert sich in drei Teile. Im ersten Teil geht es um die *technischen, philosophischen und methodischen Grundlagen* des Themas. Im zweiten Teil werden die zuvor angestellten Überlegungen anhand von ethischen Analysen in vier ausgewählten Anwendungsfeldern exemplarisch konkretisiert: der *Medizin*, der *schulischen Bildung*, der *öffentlichen Kommunikation und Meinungsbildung* sowie der *öffentlichen Verwaltung*. Im dritten Teil werden zehn in allen Anwendungsbereichen relevante *Querschnittsthemen* entfaltet, welche jeweils auch übergreifende Empfehlungen enthalten.

TEIL I: TECHNISCHE UND PHILOSOPHISCHE GRUNDLEGUNGEN

Zentrale Entwicklungen und technische Grundlagen Künstlicher Intelligenz

4. Die Idee von Maschinen, deren Fähigkeiten in bestimmten, für das menschliche Wesen besonders prägenden Kernbereichen wie dem Erkennen, Lernen oder Handeln menschlichen Fähigkeiten ähneln oder diese sogar übertreffen, lässt sich bis in die griechische

Mythologie zurückverfolgen, Jahrtausende vor der Erfindung von Softwaresystemen. Mit dem Bau der ersten Computer im 20. Jahrhundert rückte die Existenz maschineller Intelligenz erstmals in greifbare Nähe. Der englische Mathematiker Alan Turing formulierte 1950 ein später als Turing-Test bezeichnetes Kriterium für KI, nachdem maschinelle Intelligenz dann vorläge, wenn das Verhalten einer Maschine für menschliche Beobachter nicht von dem eines Menschen unterscheidbar erscheint.

5. Die frühe Forschung zur KI ging davon aus, dass man menschliches Lernen oder menschliche Intelligenz so genau beschreiben könne, dass eine Maschine dazu gebracht werden kann, sie zu simulieren. Konkrete KI-Forschungsthemen, die bis heute eine Rolle spielen, waren von Anfang an zum Beispiel Mustererkennung, Sprachverarbeitung, Abstraktionsfähigkeit, Kreativität und flexibles Problemlösen. Flankiert von Fortschritten bei der Entwicklung von Computerhardware und Programmiersprachen entstand bald großer Optimismus über die Potenziale maschineller Intelligenz. In den folgenden Jahrzehnten wechselten sich enthusiastische Phasen mit sogenannten „KI-Wintern“ ab, während derer Enttäuschungen über vermeintlich ausbleibende praktische Erfolge im Vordergrund standen und Fördermittel gekürzt wurden.

6. Weiterentwicklungen in einzelnen KI-Kerngebieten, die Entstehung paralleler Datenverarbeitungsmethoden und des Internets sowie das wachsende Engagement von Forschungsorganisationen, Militär und Industrie prägten die KI-Forschung im späten 20. Jahrhundert. Parallel dazu entstand ein kritischer Diskurs, im Zuge dessen sich auch die Computerethik als eigene Disziplin etablierte. Dabei kamen zunehmend auch philosophische Zweifel auf, ob insbesondere die von einigen Forschenden in Aussicht gestellten Visionen einer generellen oder starken Künstlichen Intelligenz jemals realisiert werden könnten – oder sollten.

7. Um die Jahrtausendwende nahmen drei Entwicklungen Fahrt auf, die der Entwicklung von KI zu einer bis heute anhaltenden Dynamik verhalfen: erstens eine deutliche Leistungssteigerung und Miniaturisierung von Computern; zweitens eine zunehmende Vernetzung digitaler Systeme und drittens damit verbundene neue Möglichkeiten der Datenzusammenführung und -auswertung.

8. Diese Entwicklungen haben zu einer intensivierten Durchdringung der Alltagswelt mit Computern geführt, darunter auch zahlreiche vernetzte und mit Sensoren versehene „smarte“ Alltagsbegleiter wie Mobiltelefone, Uhren und Haushaltgeräte. Es entstehen so-

ziotechnische Datenökosysteme, in denen über Geräte und die sie verknüpfenden Datennetzwerke zunehmend akkurate und umfangreiche digitale Repräsentationen der Bewegungen, Handlungen, Eigenschaften und Präferenzen vieler Personen entstehen. Solche digitalen Abbilder können nicht nur ausgewertet werden, sondern wirken auch auf menschliches Verhalten zurück, indem auf ihrer Grundlage Menschen Informationen oder Handlungsempfehlungen angeboten werden.

9. Das Fundament solcher digitalen Operationen und Interaktionen bilden Daten und die sie begleitenden Metadaten, die jeweils von höchst unterschiedlicher Natur wie Qualität sein können. Die Qualität eines Datensatzes hängt dabei nicht nur davon ab, wie genau, vollständig, aktuell oder detailliert seine Daten sind, sondern auch vom Verhältnis zwischen den Erhebungs- und den Anwendungskontexten. Daten können einer jeweiligen Frage oder Aufgabenstellung mehr oder weniger angemessen sein. Werden solche Fragen von Qualität und Passung nicht rechtzeitig und angemessen berücksichtigt, sind Fehler, Verzerrungen (Bias) und irreführende Analysen möglich.

10. Entscheidend für die Leistungsfähigkeit datengetriebener Anwendungen ist auch die Hardware und Infrastruktur, die für die Handhabung und Nutzung von Daten zur Verfügung steht. Hier kommen aktuell sowohl über das Internet zugängliche Dienste zum Einsatz (*Cloud Computing*), hinter denen Großeinrichtungen stehen, die auf Datenspeicherung und/oder Datenanalyse spezialisiert sind, wie auch immer leistungsfähigere Möglichkeiten, Daten zumindest teilweise bereits lokal in den Geräten, die sie erheben, zu verarbeiten (*Edge Computing*).

11. Herzstück jeglicher Datenverarbeitung sind Algorithmen: Verarbeitungsanweisungen, die vorgeben, wie eingegebene Daten meist schrittweise nach klar definierten Regeln umgeformt werden, bis der gesuchte Ausgabewert erreicht ist. Im Kontext aktueller KI-Forschung sind statistische Analysen, mit denen Regelmäßigkeiten in Daten erkannt sowie Zusammenhänge zwischen einzelnen Merkmalen identifiziert werden, von besonderer Bedeutung. Auf dieser Grundlage können Vorhersagen für ähnliche Datensätze oder künftige Entwicklungen abgeleitet werden. Geht es darum, kausale Mechanismen nachzuweisen, sind in der Regel weitere Überlegungen und Untersuchungen nötig, die eine plausible Erklärung für den vermuteten Wirkzusammenhang zwischen Merkmalen anbieten, die sich auch empirisch – zum Beispiel in Experimenten – überprüfen lässt.

12. Statistische Analysen enthalten Unsicherheiten, die sich in der Regel nicht ganz ausmerzen lassen. Mit der Minimierung bestimmter Fehlerquellen können zudem andere Fehlerquellen verstärkt werden. Welche Fehler in statistischen Analysen am ehesten in Kauf zu nehmen sind, hängt daher auch immer von der konkreten Fragestellung und Zwecksetzung ab und ist in zahlreichen Bereichen nicht nur eine technisch-methodische, sondern auch eine ethische Frage.

13. Die in KI-Systemen verwendeten algorithmischen Verfahren und Systeme werden vielfach unter dem Stichwort maschinelles Lernen zusammengefasst und zeichnen sich dadurch aus, dass sie ihre Mustersuche, Modellbildung und sonstige Funktionsweise datenbasiert optimieren können. Dabei gibt es anfangs eine Trainingsphase, in der ein Algorithmus sein Modell zur Mustererkennung durch wiederholte Analyse von Trainingsdaten aufbaut und verfeinert.

14. Maschinelles Lernen umfasst unterschiedliche Ansätze. Beim *überwachten Lernen* sind die Zuordnungen zwischen den Eingabe- und den gesuchten Ausgabedaten im Trainingsdatensatz bereits bekannt, beispielsweise Bilder von gesunder Haut und Hautkrebs, deren gesicherte Zuordnung zu einer dieser beiden Kategorien in einem Etikett (Label) vermerkt ist. *Unüberwachtes Lernen* hingegen funktioniert ohne vorherige Etikettierung der Trainingsdaten; stattdessen „sucht“ der Algorithmus eigenständig nach Mustern in Daten. Beim *Verstärkungslernen* optimiert der Algorithmus seine Operationen auf bestimmte Ziele hin und erhält dabei in der Trainingsphase für jeden Versuch eine Rückmeldung, ob dieser Schritt das System dem Ziel nähergebracht, oder es davon entfernt hat.

15. Deep Learning ist ein Teilbereich des maschinellen Lernens, der besonders für den Umgang mit großen Datenmengen geeignet ist und in den letzten Jahren zu einem wichtigen Treiber für viele KI-Anwendungen geworden ist. Hier kommen sogenannte neuronale Netze zum Einsatz, deren Funktionsweise entfernt an Netzwerkstrukturen im Gehirn angelehnt ist.

16. Die algorithmischen Strategien, die im Laufe des Trainings zur Bewältigung der jeweiligen Aufgaben entwickelt werden, sind in der Regel selbst für geschultes Personal, das den Code vollständig einsehen kann, nicht vollständig nachvollziehbar (Blackbox). Es gibt verschiedene Lösungsansätze, um trotzdem eine für die jeweilige Zielgruppe angemessene Transparenz, Interpretierbarkeit oder Erklärbarkeit algorithmischer Prozesse zu erreichen (*explainable AI*). Deren Auswahl und Anwendung ist jedoch technisch anspruchsvoll.

17. Die kombinierten Entwicklungen von Hardware und Software, Vernetzung und Datenproduktion haben vielfältig einsetzbare Anwendungsmöglichkeiten von KI hervorgebracht. KI-Systeme können beispielsweise inzwischen Menschen in anspruchsvollen Strategiespielen wie Schach und Go schlagen (MuZero) oder komplexe Texte produzieren, deren maschineller Ursprung oftmals nicht mehr zu erkennen ist (ChatGPT).

18. Der Deutsche Ethikrat nimmt in dieser Stellungnahme vier Handlungsfelder in den Blick, in denen der Einsatz von KI entweder schon besonders weitreichende Veränderungen mit sich bringt oder dies in näherer Zukunft bewirken könnte. In der *Medizin* stellt maschinelles Lernen beispielsweise Fortschritte bei Diagnostik und individualisierten Präventions- und Therapieempfehlungen in Aussicht. In der *schulischen Bildung* entstehen vielfältige Ansätze, die Vermittlung von Wissen und Kompetenzen in der Schule mithilfe von KI effektiver zu gestalten. In der *öffentlichen Kommunikation und Meinungsbildung* läuft inzwischen ein Großteil des Informationsaustauschs über algorithmisch gestützte digitale Plattformen und Soziale Medien ab. In der *öffentlichen Verwaltung* berührt der Einsatz algorithmisch gestützter Entscheidungshilfen und Prognosen das Leben vieler Menschen, beispielsweise bei der Beurteilung oder Überwachung von Personen im Bereich des Sozial- oder Polizeiwesens.

19. Um auf die mit solchen Veränderungen verbundenen Herausforderungen für das menschliche Miteinander zu reagieren, sind bereits eine Reihe von Regularien entstanden oder aktuell in der Entwicklung. Dazu gehört zum einen eine Fülle an Leitlinien, die von Codizes einzelner Unternehmen über Richtlinien von Fachgesellschaften bis hin zu Werken auf nationaler oder internationaler Ebene reicht. Parallel entwickelt sich zum anderen auch der regulative Rahmen weiter, beispielsweise in den Mediengesetzen in Deutschland. Da soziotechnische Entwicklungen im Bereich KI häufig von international agierenden IT-Unternehmen vorangetrieben werden, gewinnen Regelungen auf übernationaler Ebene zunehmend an Bedeutung, in der Europäischen Union beispielsweise die Datenschutz-Grundverordnung und der Vorschlag für einen Artificial Intelligence Act.

20. Diese Entwicklungen berücksichtigend fokussiert der Deutsche Ethikrat nicht auf den rechtlichen Rahmen, sondern gründet seine Analyse der Konsequenzen digitaler Entwicklungen für das menschliche Zusammenleben auf einer philosophischen Auseinandersetzung mit den anthropologischen Grundbegriffen, die im Mittelpunkt des menschlichen Selbstverständnisses stehen. Darauf aufbauend entwickelt er ein Verständnis von

Mensch-Technik-Relationen, in dem es entscheidend darauf ankommt, wie die Delegation menschlicher Tätigkeiten an Maschinen und algorithmische Systeme auf zentrale anthropologische Konzepte zurückwirkt und dabei insbesondere menschliche Autorschaft erweitert oder vermindert.

Zentrale Begriffe und philosophische Grundlagen

21. Das Verständnis des Begriffs der Künstlichen Intelligenz (KI) hat sich im Laufe der Jahre verändert und unterscheidet sich sowohl innerhalb als auch zwischen verschiedenen Berufsgruppen und Disziplinen. Eine große Rolle spielt die Unterscheidung zwischen sogenannter *schwacher* und *starker KI,* wobei letztere Vision eine menschenähnliche oder gar menschliche Fähigkeiten übertreffende KI beschreibt. Weitere Begriffspaare, mit denen unterschiedliche Formen oder Grade der Annäherung künstlicher an menschliche Intelligenz erfasst werden sollen, sind *spezielle* versus *allgemeine* KI sowie *enge* versus *breite* KI.

22. Die Charakterisierung als spezifische, enge, oder schwache KI einerseits, sowie allgemeine, breite oder starke KI andererseits, verweist nicht nur jeweils auf Differenzen zwischen zwei Polen. Dahinter verstecken sich, insbesondere beim Begriffspaar der schwachen und starken KI, vielmehr auch unterschiedliche Verständnisse von Intelligenz sowie unterschiedliche Positionen hinsichtlich der Kernfrage, ob es qualitative und kategorische, oder nur quantitative und prinzipiell überwindbare Unterschiede zwischen menschlicher und Künstlicher Intelligenz gibt.

23. Wichtig ist zum einen die Differenz hinsichtlich der Breite bzw. Enge des Fähigkeitenspektrums der Künstlichen Intelligenz. Die meisten Anwendungen Künstlicher Intelligenz entfalten ihre jeweilige Leistung auf klar umrissenen, engen Gebieten oder Domänen. Zum anderen geht es jedoch auch darum, ob Intelligenz an bestimmte mentale Voraussetzungen geknüpft ist, welche über die bloße *Simulation* von Verständnis hinausgehen. Es ergibt sich also die Frage, ob Intelligenz in allgemeiner oder starker Form jemals vollumfänglich Maschinen zukommen kann, oder ob dafür spezifisch menschliche Eigenschaften Voraussetzung sind.

24. Antworten auf diese Frage variieren in unterschiedlichen anthropologischen Theoriemodellen. Aus behavioristischer Sicht würden manche in einem humanoiden Roboter mit perfekten Bewegungsfähigkeiten und einer menschenähnlichen Mimik und Gestik ein Beispiel breiter oder gar starker KI sehen, wenn er in der Lage wäre, alle menschlichen

kognitiven Fähigkeiten perfekt zu simulieren. Nach anderen Konzeptionen wäre hingegen zu bestreiten, dass damit eine Form starker KI vorliege, da auch eine perfekte Simulation nicht garantiere, dass ein solcher Roboter mentale Zustände aufweise, über Einsichts- und Urteilsfähigkeit, sowie über emotive Einstellungen, wie Hoffnungen und Ängste verfüge.

25. In dieser Stellungnahme wird vorausgesetzt, dass die Unterscheidung zwischen enger und breiter KI quantitativer, bzw. gradueller Natur ist, die Entstehung einer starken KI jedoch einen qualitativen Sprung bedeuten würde. Als e*nge KI* gelten dabei Anwendungen, welche menschliche Fähigkeiten in einer Domäne simulieren, um spezifische Aufgaben zu erfüllen. *Breite KI* erweitert das Spektrum ihrer Anwendbarkeit über einzelne Domänen hinaus. Der Begriff der *starken KI* wird für die Vision einer Künstlichen Intelligenz verwendet, die jenseits der möglicherweise perfekten Simulation menschlicher Kognition auch über mentale Zustände, Einsichtsfähigkeit und Emotionen verfügen würde.

26. Eine wichtige Grundlage für Diskussionen über die aktuellen und möglichen künftigen Potenziale von KI sind Vorstellungen zu *menschlicher* Intelligenz. Aus psychologischer Perspektive ist Intelligenz als ein hypothetisches Konstrukt aufzufassen, das als solches zwar verbal umschrieben werden kann, zum Beispiel im Sinne von Verstehen, Urteilen und Schlussfolgern oder zielgerichtetem Handeln, rationalem Denken und effektiver Auseinandersetzung mit der Umwelt, aber nicht direkt beobachtbar ist. Intelligenztests ermöglichen hier eine Operationalisierung, in dem sie Situationen anbieten, in denen Menschen Verhalten zeigen können, das vor dem Hintergrund eines theoretischen Vorverständnisses als mehr oder weniger „intelligent“ bezeichnet werden kann.

27. Die Frage, ob Intelligenz eine einheitliche Fähigkeit ist oder viele Fähigkeiten umfasst, die gegebenenfalls auch voneinander unabhängig sein können, ist empirisch nicht eindeutig zu klären. Viel diskutiert und mit Blick auf KI von Relevanz ist auch der Zusammenhang von Intelligenz und Kreativität. Eine wichtige Rolle spielt hierbei die Unterscheidung zwischen konvergentem Denken, das durch logische Schlussfolgerungen zu einer einzigen oder besten Lösung gelangt, und dem für Kreativität charakteristischen divergenten Denken, das mehrere alternative Lösungen finden kann, die jeweils den gegebenen Anforderungen entsprechen.

28. In jüngerer Zeit hat sich der Blick auf Intelligenz sukzessive erweitert, beispielsweise mit Konzepten wie die der *sozialen* bzw. *emotionalen Intelligenz*. Darüber hinaus entwickelte sich rund um die Stichworte *embodied*, *embedded*, *enactive* und *extended cognition* ein

Forschungsfeld, welches in Philosophie, Psychologie und Robotik, die Rolle des Körpers einerseits sowie der Umwelt andererseits für Intelligenz und kognitive Leistungen erforscht. Spätestens diese Erweiterungen werfen die grundsätzliche Frage auf, wie die Übertragung des Intelligenzbegriffs auf technische Artefakte zu verstehen ist. Man sollte daher die Verwendung des Ausdrucks „Intelligenz“ in der Wortverbindung „Künstliche Intelligenz“ eher als eine Metapher einordnen, deren Beschreibungs- und Erklärungsfunktion genauerer Aufklärung bedarf.

29. Der Begriff der Vernunft wurde bereits lange vor der Einführung des Begriffs der Intelligenz verwendet, um die spezifische menschliche Fähigkeit zu kennzeichnen, sich in der Welt orientieren, selbstverantwortlich handeln und so der eigenen Lebenspraxis eine kohärente Struktur geben. Intelligenz ist für Vernunft eine wichtige Voraussetzung, aber keine hinreichende Bedingung.

30. Der Vernunftbegriff ist überaus komplex und umfasst ein mehrdimensionales Beziehungsgefüge von Denk-, Reflexions- und Operationsformen, das in seiner Gesamtheit im Dienste einer möglichst adäquaten Wirklichkeitserschließung steht und in einen komplexen sozialen und kulturellen Kontext verwoben ist. Von grundlegender Bedeutung ist dabei die Gegenüberstellung von *theoretischer Vernunft*, die sich auf den Erkenntnisgewinn richtet, um zu wahren empirischen oder apriorischen Urteilen zu gelangen, und *praktischer Vernunft*, die auf ein kohärentes, verantwortliches Handeln abzielt, um ein gutes Leben zu ermöglichen.

31. Vor allem im Blick auf theoretische Vernunft scheinen sich einige Parallelen zur Arbeitsweise von KI-Systemen aufzudrängen. In beiden Bereichen spielen Fähigkeiten der Informationsverarbeitung, des Lernens, des logischen Schlussfolgerns und konsistenten Regelfolgens sowie der sinnvollen Verknüpfung gespeicherter Informationen eine zentrale Rolle. Bei näherer Betrachtung zeigen sich jedoch insofern gravierende Differenzen als sich nicht nur die Arbeitsweise des menschlichen Gedächtnisses in mehrfacher Hinsicht vom technischen Speicher eines Computers unterscheidet, sondern auch die menschliche Urteilspraxis technisch nicht substituierbar ist. Zumindest die bislang verfügbaren KI-Systeme verfügen nicht über die dafür relevanten Fähigkeiten des Sinnverstehens, der Intentionalität und der Referenz auf eine außersprachliche Wirklichkeit.

32. Dies bestätigt sich erst recht für die praktische Vernunft, die insofern noch von weit komplexerer Natur ist, als ihr Ziel nicht nur in wohlbegründeten praktischen Einzelurteilen, sondern in einem möglichst richtigen und verantwortlichen Handeln besteht, das über

einen langen Zeitraum aufrechterhalten wird, eine kohärente Ordnung der Praxis garantiert und damit ein insgesamt gutes Leben ermöglicht. Dazu bedarf es mehrerer Einzelkompetenzen, deren Simulationsmöglichkeiten durch technische Artefakte kontrovers diskutiert werden.

33. Zu diesen Einzelkompetenzen gehört erstens ein *Verständnis* der für unsere Moralsprache bedeutsamen Begriffe zur Bezeichnung moralisch relevanter Güter, Werte und Haltungen; zweitens ein *Unterscheidungs- und Einfühlungsvermögen*; drittens die Fähigkeit zur *Abwägung* konfligierender Güter und Werte; viertens die Befähigung zum *reflektierten Umgang mit Regeln* unterschiedlicher Reichweite; fünftens die Kompetenz zum *intuitiven Erfassen komplexer Handlungssituationen und Umstände*; sechstens *Urteilsvermögen*; siebtens die Fähigkeit zur *Begründung* der eigenen moralischen Urteile und der ihnen korrespondierenden Praxis; und achtens *Affekt-* und *Impulskontrolle*, um die jeweils gefällten praktischen Urteile auch handlungswirksam werden zu lassen.

34. Während partielle Überschneidungen des Kompetenzprofils moderner KI-Systeme mit dem komplexen Phänomen menschlicher Vernunft durchaus möglich sind, ist zu berücksichtigen, dass die hier genannten Einzelfähigkeiten nicht beziehungslos nebeneinanderstehen. Vielmehr ist von vielfältigen Wechselwirkungen, Rückkopplungen und Bedingungsverhältnisse zwischen ihnen auszugehen. Sie bilden einen integralen Bestandteil einer komplexen menschlichen Natur, die als leib-seelischen Einheit zu verstehen ist. Menschliche Vernunft ist stets als *verleiblichte Vernunft* zu begreifen. Praktische Vernunft ist zudem nicht aus einer rein individualistischen Perspektive zu verstehen. Da jeder Mensch Teil einer sozialen Mitwelt und kulturellen Umgebung ist, die sich nachhaltig auf seine Sozialisation auswirkt, müssen auch überindividuelle kulturelle Faktoren in die Deutung der praktischen Vernunft einbezogen werden.

35. Ein angemessenes Verständnis insbesondere des praktischen Vernunftgebrauchs ist eng mit unserem basalen Selbstverständnis als handlungsfähige Personen verbunden. Nicht jedes menschliche Tun, das auf die Umwelt einwirkt, ist als *Handlung* zu verstehen, sondern nur solches, das zweckgerichtet, beabsichtigt und kontrolliert ist. Unterstellt man, dass Maschinen nicht zweckgerichtet operieren, also keine Absichten haben, dann ist die Zuschreibung von Handlungen in Bezug auf Maschinen in diesem engen Sinne nicht möglich.

36. Im KI-Diskurs kommt allerdings seit der Jahrtausendwende zunehmen die Frage auf, in welchem Sinne Maschinen außerhalb des obigen engen Handlungsbegriffs doch in bestimmten Kontexten in einem weiteren Sinne handeln können, zum Beispiel wenn Entscheidungen komplett an Softwaresysteme delegiert werden. Daran anknüpfend gibt es einen Diskurs, ob und inwieweit zunehmend eigenständige, d.h. ohne menschliches Zutun funktionierende maschinelle Systeme als „Agenten" in der Folge für ihr „Handeln" verantwortlich gemacht werden können, etwa mit Blick auf Fragen der Haftung.

37. Selbst wenn Maschinen komplexe Vollzüge oder Operationen durchführen, damit Veränderungen in der Welt bewirken und flexibel mit anspruchsvollen Herausforderungen der menschlichen Lebenswelt umgehen können, führen sie diese Veränderungen aber nicht absichtlich herbei und haben sie diese daher auch nicht in einem moralischen und rechtlichen Sinne zu verantworten. Vor diesem Hintergrund scheint es sinnvoll, den Handlungsbegriff im engen Sinne Menschen vorzubehalten, um inflationären Ausweitungen des Akteur-Status zu vermeiden und konzeptionelle Grenzziehungen zu ermöglichen.

38. Entscheidend ist demnach das Konzept der Handlungsurheberschaft bzw. Autorschaft, das auf die universelle menschliche Erfahrung verweist, sich selbst und andere im Hinblick auf bestimmte Ereignisse und Zustände als Urheber anzusehen. Die Fähigkeit zur Handlungsurheberschaft kann als Grundlage von Autonomie betrachtet werden, also dafür, dass handelnde Menschen ihre Handlungen nach Maximen ausrichten können, die sie sich selber setzen.

39. Die Umstände und Folgen von Handlungen können für deren moralische und rechtliche Bewertung von Bedeutung sein. So können aus einer Handlung beispielsweise neben den beabsichtigten Folgen auch nicht beabsichtigte, aber dem Handelnden erkennbare Folgen erwachsen. Dies ist relevant für das Konzept von Fahrlässigkeit, das im Kontext von KI eine große Rolle spielt. Und auch wenn primär einzelne Menschen handeln, schließt dies ein Konzept kollektiver Handlungen nicht aus, bei denen mehrere Personen von vornherein in einem Kontext der Koordination agieren.

40. Auch Technologie kann erheblichen Einfluss auf menschliches Handeln oder die menschliche Handlungserfahrung haben. Die zunehmende Durchdringung der menschlichen Lebenswelt mit informationstechnisch immer leistungsfähigeren Maschinen führt zu hybriden, sozio-technischen Konstellationen, in denen Menschen und Maschinen eng verwoben sind und auf komplexe Weise interagieren. Zudem können manche maschinel-

len Systeme menschliches Tun zum Teil so gut imitieren, dass die Simulation wie intentionales menschliches Handeln erscheint. Vor diesem Hintergrund ist es sinnvoll, an einem engen Handlungsbegriff, der an das zentrale Kriterium der Intentionalität gebunden ist, festzuhalten.

41. Das Intentionalitätskriterium ist zudem entscheidend für die Möglichkeit der Zuschreibung von *Verantwortung* im Kontext von Mensch-Maschine-Interaktionen in zunehmend komplexer sozio-technischer Vernetzung. Verantwortung kann als Konzept einer fünffachen Relation gefasst werden: Wer (*Verantwortungssubjekt*) ist für was (*Verantwortungsobjekt*) gegenüber wem (*Betroffenen*), vor wem (*Instanz*) und unter welcher *Norm* verantwortlich?

42. In der Verantwortungsdiskussion zu wissenschaftlich-technischem Fortschritt ist zu bedenken, dass die Handlungsfolgen neuer Entwicklungen oft nur unter hohen und nicht eliminierbaren Wissensunsicherheiten abgeschätzt werden können. Verantwortungszuschreibung muss daher die Dimension des Handelns unter Unsicherheit berücksichtigen.

43. Moralische Verantwortung können nur natürliche Personen übernehmen, die über Handlungsfähigkeit verfügen, d.h. in der Lage sind, aktiv, zweckgerichtet und kontrolliert auf die Umwelt einzuwirken und dadurch Veränderungen zu verursachen. Träfe dies auch auf Maschinen zu, wären auch diese verantwortungsfähig. Dann müsste Maschinen der Personenstatus zugeschrieben werden, was jedoch weder aktuell noch angesichts der in absehbarer Zukunft erwartbaren qualitativen Entwicklungen maschineller Systeme angemessen wäre. Verantwortung kann daher nicht direkt von maschinellen Systemen übernommen werden, sondern nur von den Menschen, die in je unterschiedlichen Funktionen hinter diesen Systemen stehen, gegebenenfalls im Rahmen institutioneller Verantwortung.

44. Wer jeweils konkret wie viel Verantwortung trägt, ist häufig schwierig zu bestimmen. Die facettenreichen Verantwortungsgefüge zwischen Individuen, Organisationen und Staat werden noch komplexer, wenn die Wechselwirkungen zwischen diesen Beteiligten zumindest teilweise von algorithmischen Systemen gestützt oder vermittelt werden, welche mitunter kaum durchschaubar sind oder autonom zu agieren scheinen. Vor einem solchen Hintergrund ist eine angemessene Gestaltung von *Multiakteursverantwortung* zentral.

45. Handlung, Vernunft und Verantwortung stehen im Zentrum humanistischer Philosophie. Menschen sind befähigt zur Handlungsurheberschaft und somit zur Autorschaft ihres Lebens. Sie sind frei und tragen daher Verantwortung für die Gestaltung ihres Handelns.

Freiheit und Verantwortung sind zwei sich wechselseitig bedingende Aspekte menschlicher Autorschaft. Autorschaft ist wiederum an Vernunftfähigkeit gebunden.

46. Im Mittelpunkt dieser Trias aus Vernunft, Freiheit und Verantwortung steht das Phänomen der Affektion durch Gründe. Praktische Gründe sprechen für Handlungen, theoretische Gründe sprechen für Überzeugungen. In der Regel gibt es Gründe das Eine zu tun und das Andere zu lassen, die gegeneinander abgewogen werden müssen. Der Konflikt von Gründen zwingt dann zur Abwägung und zur Systematisierung dieser Abwägung in Gestalt ethischer Theoriebildung.

47. Die menschliche Lebensform ist von reaktiven Einstellungen und moralischen Gefühlen geprägt, die von normativen Gründen begleitet sind. Freiheit kommt insofern ins Spiel, als wir diese zurückstellen, wenn wir erfahren, dass eine Person in ihrem Handeln nicht frei war. Diese Praxis der Zuschreibung von Freiheit und Verantwortung ist essenziell für die Grundlegung moralischer Beurteilung. Die Normen von Moral und Recht sind ohne die Annahme menschlicher Verantwortung und damit Freiheitsfähigkeit und Vernunftfähigkeit unbegründet.

48. Eine Herausforderung dieser humanistischen Perspektive kommt aus den Neurowissenschaften. Empirische Studien, nach denen beispielsweise das motorische Zentrum des Gehirns schon mit der Vorbereitung einer Bewegung beginnt, bevor man sich bewusst für die Ausführung der Bewegung entschieden hat, werden mitunter als Beleg dafür interpretiert, dass es Freiheit und damit menschliche Verantwortlichkeit nicht gebe. Tatsächlich lassen solche Befunde jedoch unterschiedliche Interpretationen zu und eignen sich nicht als Widerlegung menschlicher Freiheit und Verantwortung.

49. Eine zweite Kritik der humanistischen Anthropologie wird von der KI-Debatte inspiriert. Sie changiert zwischen einer Überwindung des Menschen in Gestalt des Transhumanismus, der mit neuen Mensch-Maschinen-Symbiosen die Reichweite menschlichen Wirkens in neue Dimensionen heben möchte und einem Maschinenparadigma, das den menschlichen Geist auf das Modell eines algorithmischen Systems reduziert. Gerade letzteres entfaltet besondere Relevanz im Kontext dieser Stellungnahme, da es großen Einfluss auf die Interpretation der Wechselwirkungen zwischen Mensch und Maschine und deren Rückwirkungen auf das menschliche Selbstverständnis hat.

50. In Maschinenparadigmen werden Menschen materialistisch als Maschinen oder Maschinen animistisch mit mentalen Zuständen ausgestattet und als menschengleich gedeutet. Die in KI-Diskursen teilweise verbreitete Tendenz, eine äußerliche Ununterscheidbarkeit

von menschlicher und maschineller *Performanz* pauschal mit der Annahme von Intelligenz und Denkvermögen solcher Maschinen gleichzusetzen, ist das Ergebnis bestimmter theoretischer Vorannahmen insbesondere *behavioristischer* und *funktionalistischer* Art.

51. Der Behaviorismus versucht, menschliches Verhalten auf der Grundlage präzise beschreibbarer Reiz-Reaktions-Schemata zu erklären und die Psychologie damit in eine exakte Wissenschaft zu verwandeln; das Innenleben derart beschriebener Organismen wird dabei komplett ausgeblendet. Der *Funktionalismus* beruht auf der Annahme, dass mentale Zustände funktional vollständig erfasst werden können und die Frage nach der *Seinsart* mentaler Zustände zugunsten der genauen Beschreibung ihrer *Funktion* aufgehoben werden kann und sollte. Durch die These der *multiplen Realisierung*, nach der bestimmte mentale Ereignisse, Eigenschaften oder Zustände durch ganz unterschiedliche physikalische Ereignisse, Eigenschaften oder Zustände realisiert werden können, scheint es zudem möglich, auch Computern mentale Zustände zuzuschreiben, obwohl sie keine biologischen Strukturen besitzen.

52. Kritik am Funktionalismus verweist auf *phänomenales Bewusstsein*, nach dem die mentalen Zustände eines Wesens entscheiden von *Empfindungsqualitäten* abhängen, die allein aufgrund *äußeren* Verhaltens nicht zugänglich sind. Dieses phänomenale Bewusstsein setzt dem Vermögen, die Qualität des Erlebens oder die mentalen Zustände anderer Lebewesen zu beurteilen, gewisse Grenzen und lässt die funktionalistisch inspirierte Mensch-Computer-Analogie als eine fragwürdige Reduktion erscheinen.

53. Ein weiteres Argument gegen den Funktionalismus stammt aus John Searles Gedankenexperiment zum „chinesischen Zimmer“, in dem eine Person aus einer Kammer anhand einer genauen Gebrauchsanleitung chinesische Antworten auf Fragen herausreicht. Nicht diese Person beherrscht die chinesische Sprache, und auch kein Übersetzungscomputer, sondern diejenigen, die die Gebrauchsanleitung bzw. den Algorithmus zur Beantwortung der Fragen verfasst haben.

54. In der Zurückweisung funktionalistischer Maschinenparadigmen wird die Bedeutung der gesamten Lebenserfahrung für die Vernunft deutlich. Menschliche Vernunft ist *leibliche Vernunft*. Der Leib ist Ausgangspunkt und Bestandteil jeder Wahrnehmung und Empfindung und Voraussetzung für menschliches In-der-Welt-Sein und die Herstellung von Beziehungen zu anderen. Kognitive Fähigkeiten sind in ihrem Entstehungs- und Vollzugsprozess also an Sinnlichkeit und Leiblichkeit, Sozialität und Kulturalität gebunden.

55. Daraus ergeben sich auch Grenzen der Formalisierbarkeit und Simulierbarkeit menschlicher Vernunft. Die Aneignung menschlicher Erfahrung ist immer mit Deutungsprozessen verbunden und setzt immer ein Beteiligtsein, ein Engagement voraus. Auch hier spielt der Leib eine wichtige Rolle, denn er ermöglicht ein Handeln, das allein mittels bewusster Planung und Berechnung so nicht möglich wäre. Darin gründet eine Nicht-Simulierbarkeit des Denkens, vor deren Hintergrund die Entwicklung von Künstlicher Intelligenz an Grenzen stößt.

56. Aus den bisherigen Überlegungen lassen zusammenfassen, dass menschliche Intelligenz unauflöslich mit den vielfältigen Dimensionen der menschlichen Lebenswelt verbunden ist. Sie operiert gründegeleitet und ist Ausdruck von akzeptierten Werten und Normen. Es ist fraglich, ob eine derart gründegeleitete, multidimensional bestimmte und soziokulturell eingebettete kohärente Praxis selbst für komplexe maschinelle Systeme jemals plausibel sein könnte.

Mensch-Technik-Relationen

57. Menschen entwickeln, gestalten und nutzen Technik als Mittel zum Zweck. Die mehr oder minder umfassende Delegation menschlicher Tätigkeiten an Maschinen – bis hin zur vollständigen Ersetzung menschlicher Handlungen durch maschinelle Vollzüge – wirkt allerdings häufig zurück auf menschliche Handlungsmöglichkeiten, Fertigkeiten, Autorschaft und Verantwortungsübernahme und kann diese jeweils erweitern oder vermindern. Die drei Begriffe des *Erweiterns*, *Verminderns* und *Ersetzens* dienen in diese Stellungnahme als analytische Matrix.

58. Technikgestaltung wird im *sozialem Konstruktivismus* als Prozess beschrieben der eher menschlich gesetzten, durch die jeweiligen gesellschaftlichen Prioritäten geprägten Zwecken folgt. Im *technologische Determinismus* wird eine insbesondere nach ökonomischen Verhältnissen bestimmende Eigendynamik als maßgeblich gesehen, der sich Mensch und Gesellschaft letztlich unterordnen und anpassen müssen. Tatsächlich spielen beide Ansätze zusammen und unterliegt die Mensch-Technik-Relation von Grund auf einem Ko-Konstruktions-Verhältnis und kann als Ko-Evolution beschrieben werden. Soziale Kontexte und normative Kriterien auf der einen und Technologien auf der anderen Seite entwickeln sich weiter in gegenseitiger Wechselwirkung.

59. In ihrer Gesamtheit kann Technik dabei zu einer *zweiten Natur* werden, die Randbedingungen und Erfolgsbedingungen für weiteres menschliches Leben setzt und auch Weltsicht und das Problemlösen beeinflusst. Somit ist neue Technologie oft bereits das Ergebnis einer technologischen Art und Weise, wie Menschen die Welt sehen und sich zu ihr in Beziehung setzen. Die zunehmende Komplexität der Mensch-Technik bzw. Mensch-Maschine-Relation verändert auch deren Wahrnehmung. In KI-gesteuerten Systemen scheinen die vormals klaren Unterscheidungen von Mensch und Technik weniger eindeutig zu werden. Auch in der Umgangssprache ist die Anthropomorphisierung digitaler Technik weit fortgeschritten, zum Beispiel in der Zuschreibung von Fähigkeiten wie Denken, Lernen, Entscheiden oder Zeigen von Emotion an KI und Roboter.

60. Subjekt-Objekt-Verhältnisse zwischen Mensch und Technik verändern sich ebenfalls. In vernetzten Systemen haben Menschen teils die Subjekt-, teils aber auch die Objektrolle inne. Wenn beispielsweise Entscheidungen über Menschen an Softwaresysteme delegiert werden, etwa hinsichtlich der Gewährung von Sozialleistungen, werden Menschen zu Objekten der „Entscheidungen“ dieser Systeme, die hier auftreten, als ob sie Subjekte seien.

61. Verschiedene Ansätze versuchen, diese Entwicklungen in Konzepten zu mehrstufigen Mensch-Technik-Wechselwirkungen zu beschreiben.

62. Die Zuschreibung von Verantwortung bleibt in diesen Ansätzen jeweils beim Menschen. Moralisch problematische Resultate können dennoch durch KI-Systeme verursacht werden und sie haben Einfluss auf menschliches Handeln. Dieses ist also weder völlig autonom noch völlig sozial oder technisch determiniert, sondern in zunehmendem Maß soziotechnisch situiert.

63. KI zeigt vielen Fällen eindeutig positive Folgen im Sinne der *Erweiterung* der Möglichkeiten menschlicher Autorschaft Im Rahmen der Diffusion von Technik und Innovationen in die Gesellschaft, ihrer Nutzung und Veralltäglichung kommt es jedoch auch zu *Verminderungen* menschlicher Entfaltungsmöglichkeiten. Durch den Einsatz digitaler Technologien können Abhängigkeiten von diesen oder Anpassungsdruck entstehen – und andere, bis dahin etablierte Optionen verschlossen werden.

64. Solche Effekte können schleichend und teilweise unbewusst durch Verhaltensänderungen entstehen auftreten, ohne dass Intentionen von Akteuren dahinterstehen. Das *Ersetzen* als Endpunkt des Delegierens vormals menschlich ausgeübter Tätigkeiten an technische Systeme erfolgt jedoch *intentional*. Eine derartige Übertragung ist für sich genommen ein

Ausdruck der Wahrnehmung menschlicher Autorschaft. Die zentrale ethische Frage ist, ob und wie diese Übertragung die Möglichkeiten *anderer* Menschen beeinflusst, vor allem von jenen, *über* die entschieden wird. Daraus ergibt sich ein Bedarf, die Übertragung menschlicher Tätigkeiten auf KI-Systeme auch gegenüber den davon Betroffenen transparent zu gestalten und bei der Beurteilung von KI zu berücksichtigen, *für wen* eine Anwendung jeweils Chancen oder Risiken, und Erweiterungen oder Verminderungen der Autorschaft mit sich bringt. Damit sind Aspekte sozialer Gerechtigkeit und Macht involviert.

65. Weiterhin sind psychologische Effekte im Zusammenhang mit KI-Systemen zu beachten, vor allem der *Automation Bias.* Menschen vertrauen algorithmisch erzeugten Ergebnissen und automatisierten Entscheidungsprozeduren häufig mehr als menschlichen Entscheidern. Damit wird Verantwortung – zumindest unbewusst – an diese „Quasi-Akteure" delegiert. Selbst wenn ein KI-System normativ strikt auf die Rolle der *Entscheidungsunterstützung* begrenzt wird, kann Automation Bias dazu führen, dass ein KI-System allmählich in die Rolle des eigentlichen „Entscheiders" gerät und menschliche Autorschaft und Verantwortung ausgehöhlt werden.

TEIL II: AUSGEWÄHLTE ANWENDUNGEN UND SEKTORSPEZIFISCHE EMPFEHLUNGEN

Medizin

66. KI-gestützte digitale Produkte kommen zunehmend im Gesundheitssystem zum Einsatz. Die Betrachtung der mit ihnen verbundenen Chancen und Risiken bedarf einer wenigstens dreifachen Differenzierung. Erstens sind mehrere *Akteursgruppen* zu unterscheiden, die bezüglich eines KI-Einsatz unterschiedliche Funktionen und Verantwortlichkeiten besitzen. Zweitens umfasst das Gesundheitswesen von der Forschung bis zur konkreten Patientenversorgung unterschiedliche *Anwendungsbereiche* für KI-Produkte. Drittens sind unterschiedliche *Grade der Ersetzung* menschlicher Handlungssegmente zu beobachten.

67. Bereits die *Entwicklung* geeigneter KI-Komponenten für die medizinische Praxis erfordert enge interdisziplinäre Zusammenarbeit verschiedener Sachverständiger und stellt hohe Anforderungen an die Qualität der verwendeten Trainingsdaten, um vermeidbare Verzerrungen der Ergebnisse von vorneherein zu minimieren. Systeme sind so zu konzipieren, dass die Plausibilitätsprüfungen in der Nutzungsphase vorsehen, um den Gefahren

eines Automation Bias zu entgehen. Mittels geeigneter Prüf-, Zertifizierungs- und Auditierungsmaßnahmen sollte gewährleistet werden, dass nur hinreichend geprüfte KI-Produkte zum Einsatz kommen, deren grundlegenden Funktionsweise zumindest bei Systemen, die Entscheidungsvorschläge mit schwerwiegenden Konsequenzen für Betroffene unterbreiten, auch für diejenigen, die ein Produkt später verwenden, hinreichend erklär- sowie interpretierbar ist.

68. In der medizinischen *Forschung* kann der Einsatz von KI in mehrfacher Hinsicht vorteilhaft sein, sofern der Schutz der an den Studien teilnehmenden Personen und ihrer Daten gewährleistet ist. KI kann hier beispielsweise hilfreiche Vor- und Zuarbeiten bei Literaturrecherchen oder der Durchsuchung großer Datenbanken leisten, neue Korrelationen zwischen bestimmten Phänomenen entdecken und auf dieser Grundlage treffsichere Vorhersagen machen, etwa zur Ausbreitung eines Virus oder zur Struktur komplexer Moleküle.

69. In der medizinischen *Versorgung* werden KI-Instrumente zunehmend auch zur Diagnostik und Therapie eingesetzt, beispielsweise bei Brust- und Prostatakrebserkrankungen. Entscheidungsunterstützungssysteme modellieren und automatisieren hier Entscheidungsprozesse mittels Analyse verschiedener Parameter der Labordiagnostik, der Bildbearbeitung sowie der automatisierten Durchsicht von Patientenakten und wissenschaftlichen Datenbanken. Gerade Fortschritte in der KI-gestützten Bilderkennung eröffnen dabei neue Möglichkeiten einer frühzeitigen Detektion, Lokalisation und Charakterisierung pathologischer Veränderungen. In der Therapie kommt KI beispielsweise in Operationsrobotern zum Einsatz.

70. Wenn ärztliche Tätigkeiten in derart engem bis mittleren Ausmaß an Technik delegiert werden, können beispielsweise Tumore früher erkannt, Therapieoptionen erweitert und die Chancen auf eine erfolgreiche Therapie somit erhöht werden. Für ärztliches Personal eröffnet die Technik zudem die Chance, von monotonen Routinearbeiten entlastet zu werden und mehr Zeit für den Austausch mit der jeweiligen Patientin zu gewinnen. Diesen Chancen stehen aber auch Risiken gegenüber, beispielweise wenn Fachkräfte durch die fortschreitende Delegation bestimmter Aufgaben an technische Systeme eigene Kompetenzen verlieren oder Sorgfaltspflichten im Umgang mit KI-gestützter Technik aufgrund eines Automation Bias vernachlässigen.

71. Um die Chancen des KI-Einsatzes in klinischen Situation zu realisieren und Risiken zu minimieren, sind mehrere Ebenen zu berücksichtigen. So bedarf es unter anderem einer

flächendeckenden und möglichst einheitlichen technischen Ausrüstung, Personalschulung und kontinuierlichen Qualitätssicherung ebenso wie Strategien, die gewährleisten, dass auch in KI-gestützten Protokollen Befunde auf Plausibilität geprüft werden, die persönliche Lebenssituation von Erkrankten umfassend berücksichtigt und vertrauensvoll kommuniziert wird. Auch der bei den meisten medizinischen KI-Anwendungen große Datenbedarf bringt Herausforderungen mit sich, sowohl hinsichtlich des Schutzes der Privatsphäre Betroffener als auch mit Blick auf eine teils sehr restriktive individuelle Auslegungspraxis geltender Datenschutzbestimmungen, die der Realisierung von Potenzialen des KI-Einsatzes in der klinischen Praxis im Wege stehen kann.

72. Einer der wenigen medizinischen Handlungsbereiche, in denen KI-basierte Systeme zum Teil ärztliches bzw. anderes Gesundheitspersonal mitunter weitgehend oder vollständig ersetzen können, ist die Psychotherapie. Hier kommen seit einigen Jahren Instrumente zum Einsatz, meist in Form von Bildschirm-basierten Apps, die auf algorithmischer Basis eine Art von Therapie anbieten. Solche Apps können einerseits angesichts ihrer Niedrigschwelligkeit und ständigen Verfügbarkeit Menschen in Erstkontakt mit therapeutischen Angeboten bringen, die sonst zu spät oder gar nicht eine Therapie erhalten. Andererseits gibt es Bedenken etwa hinsichtlich mangelnder Qualitätskontrollen, dem Schutz der Privatsphäre oder wenn Menschen eine Art emotionale Beziehung zur therapeutischen App aufbauen. Kontrovers diskutiert wird auch, ob die zunehmende Nutzung solcher Apps weiterem Abbau von therapeutischem Fachpersonal Vorschub leistet.

73. Auf Grundlage dieser Überlegungen formuliert der Deutsche Ethikrat neun Empfehlungen für den Einsatz von KI im Gesundheitssektor:

- *Empfehlung Medizin 1:* Bei der Entwicklung, Erprobung und Zertifizierung medizinischer KI-Produkte bedarf es einer engen Zusammenarbeit mit den relevanten Zulassungsbehörden sowie insbesondere mit den jeweils zuständigen medizinischen Fachgesellschaften, um Schwachstellen der Produkte frühzeitig zu entdecken und hohe Qualitätsstandards zu etablieren.
- *Empfehlung Medizin 2:* Bei der Auswahl der Trainings-, Validierungs- und Testdatensätze sollte über bestehende Rechtsvorgaben hinaus mit einem entsprechenden Monitoring sowie präzise und zugleich sinnvoll umsetzbaren Dokumentationspflichten sichergestellt werden, dass die für die betreffenden Patientengruppen relevanten Faktoren (wie z. B. Alter, Geschlecht, ethnische Einflussfaktoren, Vorerkrankungen und Komorbiditäten) hinreichend berücksichtigt werden.

- *Empfehlung Medizin 3:* Bei der Gestaltung des Designs von KI-Produkten zur Entscheidungsunterstützung ist sicherzustellen, dass die Ergebnisdarstellung in einer Form geschieht, die Gefahren etwa von Automatismen (Automation Bias) transparent macht, ihnen entgegenwirkt und die die Notwendigkeit einer reflexiven Plausibilitätsprüfung der jeweils vom KI-System vorgeschlagenen Handlungsweise unterstreicht.
- *Empfehlung Medizin 4:* Bei der Sammlung, Verarbeitung und Weitergabe von gesundheitsbezogenen Daten sind generell strenge Anforderungen und hohe Standards in Bezug auf Aufklärung, Datenschutz und Schutz der Privatheit zu beachten. In diesem Zusammenhang verweist der Deutsche Ethikrat auf seine 2017 in Kontext von Big Data und Gesundheit formulierten Empfehlungen, die sich am Konzept der Datensouveränität orientieren, das für den Bereich von KI-Anwendungen im Gesundheitsbereich gleichermaßen Gültigkeit entfaltet.
- *Empfehlung Medizin 5:* Bei durch empirische Studien sorgfältig belegter Überlegenheit von KI-Anwendungen gegenüber herkömmlichen Behandlungsmethoden ist sicherzustellen, dass diese allen einschlägigen Patientengruppen zur Verfügung stehen.
- *Empfehlung Medizin 6:* Für erwiesen überlegene KI-Anwendungen sollte eine rasche Integration in die klinische Ausbildung des ärztlichen Fachpersonals erfolgen, um eine breitere Nutzung vorzubereiten und verantwortlich so gestalten zu können, dass möglichst alle Patientinnen und Patienten davon profitieren und bestehende Zugangsbarrieren zu den neuen Behandlungsformen abgebaut werden. Dazu ist die Entwicklung einschlägiger Curricula/Module in Aus-, Fort- und Weiterbildung notwendig. Auch die anderen Gesundheitsberufe sollten entsprechende Elemente in die Ausbildung aufnehmen, um die Anwendungskompetenz bei KI-Anwendungen im Gesundheitsbereich zu stärken.
- *Empfehlung Medizin 7:* Bei routinemäßiger Anwendung von KI-Komponenten sollte nicht nur gewährleistet werden, dass bei denjenigen, die sie klinisch nutzen, eine hohe methodische Expertise zur Einordnung der Ergebnisse vorhanden ist, sondern auch strenge Sorgfaltspflichten bei der Datenerhebung und -weitergabe sowie bei der Plausibilitätsprüfung der maschinell gegebenen Handlungsempfehlungen eingehalten werden. Besondere Aufmerksamkeit erfordert die Gefahr eines Verlustes von theoretischem wie haptisch-praktischem Erfahrungswissen und entsprechenden Fähigkeiten (*deskilling*); dieser Gefahr sollte mit geeigneten, spezifischen Fortbildungsmaßnahmen entgegengewirkt werden.

- *Empfehlung Medizin 8:* Bei fortschreitender Ersetzung ärztlicher, therapeutischer und pflegerischer Handlungssegmente durch KI-Komponenten ist nicht nur sicherzustellen, dass Patientinnen und Patienten über alle entscheidungsrelevanten Umstände ihrer Behandlung vorab informiert werden. Darüber hinaus sollten auch gezielte kommunikative Maßnahmen ergriffen werden, um dem drohenden Gefühl einer zunehmenden Verobjektivierung aktiv entgegenzuwirken und das Vertrauensverhältnis zwischen den beteiligten Personen zu schützen. Je höher der Grad der technischen Substitution menschlicher Handlungen durch KI-Komponenten ist, desto stärker wächst der Aufklärungs- und Begleitungsbedarf der Patientinnen und Patienten. Die verstärkte Nutzung von KI-Komponenten in der Versorgung darf nicht zu einer weiteren Abwertung der sprechenden Medizin oder einem Abbau von Personal führen.
- *Empfehlung Medizin 9:* Eine vollständige Ersetzung der ärztlichen Fachkraft durch ein KI-System gefährdet das Patientenwohl und ist auch nicht dadurch zu rechtfertigen, dass schon heute in bestimmten Versorgungsbereichen ein akuter Personalmangel besteht. Gerade in komplexen Behandlungssituationen bedarf es eines personalen Gegenübers, das durch technische Komponenten zwar immer stärker unterstützt werden kann, dadurch selbst als Verantwortungsträger für die Planung, Durchführung und Überwachung des Behandlungsprozesses aber nicht überflüssig wird.

Bildung

74. Auch in der schulischen Bildung kommen zunehmend digitale Technologien und algorithmische Systeme zum Einsatz. Dies kann sowohl zur Standardisierung von Lernprozessen führen, als auch mehr Personalisierung ermöglichen. Die Einsatzmöglichkeiten reichen von sehr eng umrissenen punktuellen Angeboten bis hin zu Szenarien, in denen KI-gestützte Lehr-Lernsysteme zeitweise oder gänzlich eine Lehrkraft ersetzen können.

75. Das hier zugrunde gelegte Verständnis von Bildung orientiert sich an der Fähigkeit des Menschen zu freiem und vernünftigem Handeln, das nicht auf behavioristische oder funktionalistische Modelle zu reduzieren ist. Bildung erfordert den Erwerb von Orientierungswissen als Bedingung von reflexiver Urteilskraft und Entscheidungsstärke. Dieser Prozess umfasst auch kulturelles Lernen sowie emotionale und motivationale Aspekte. Das Lehr- und Lerngeschehen ist als dynamische Interaktion mit anderen Personen zu begreifen. Der Einsatz von KI-gestützten Instrumenten in der Schule ist daraufhin zu überprüfen,

ob er einem solchen Verständnis des Menschen als einer zur Selbstbestimmung und Verantwortung fähigen Person entspricht und solche Prozesse fördert, oder ob er diesen entgegensteht.

76. Ausgangspunkt der meisten KI-Anwendungen in der Bildung ist die Sammlung und Auswertung vieler Daten der Lernenden und mitunter auch der Lehrkräfte. Hier stellen sich Fragen nach dem sinnvollen Grad und Ausmaß der Datenerhebung sowie deren wünschenswerten Verwendungsweisen. Es geht darum, Lernende in ihrem individuellen Lernprozess durch Datennutzung bestmöglich zu unterstützen und gleichzeitig zu verhindern, dass diese Daten zur Überwachung oder Stigmatisierung von einzelnen Lernenden missbraucht werden können.

77. Auf Grundlage der erhobenen Daten können individualisierte Rückmeldungen über Lern- und Lehrprozesse erfolgen sowie entsprechende Reaktionen oder Empfehlungen des Softwaresystems. Durch Auswertung von zum Beispiel Lerngeschwindigkeit, typischen Fehlern, Stärken und Schwächen kann die Software das Lernprofil der Lernenden erkennen und die Lerninhalte entsprechend anpassen. Subjektive Eindrücke der Lehrkräfte können dadurch datenbasiert untermauert, aber auch korrigiert werden.

78. Auch in der Schule kann es durch KI zu engen, mittleren und weiten Ersetzungen bestimmter Handlungssegmente und Interaktionen kommen. Eine enge Ersetzung liegt etwa vor, wenn ein Softwaresystem für einen genau bestimmten Lernabschnitt eingesetzt wird. Aufwändigere und datenintensivere Intelligente Tutorsysteme können auch komplexere Lerninhalte in unterschiedlichen Fächern im Zusammenwirken mit Lernenden vermitteln und so breitere Teilaspekte des Unterrichtsgeschehens ersetzen oder im Einzelfall die Funktion einer Lehrkraft vollständig übernehmen.

79. Darüber hinaus gibt es mittlerweile auch Bestrebungen, KI zur Analyse des Verhaltens im Klassenraum einzusetzen (*classroom analytics*), um die Dynamik ganzer Lerngruppen umfassend zu dokumentieren und auszuwerten. Solche Ansätze sind aufgrund der für sie notwendigen Erfassung vielfältiger Daten u.a. über das Verhalten von Schülerinnen und Schülern sowie Lehrkräften umstritten. Chancen auf verbesserte Pädagogik und Didaktik stehen potenziell negative Auswirkungen umfangreicher Datensammlungen auf die Privatsphäre und Autonomie aller Beteiligten gegenüber.

80. Ein besonders kontrovers diskutierter Aspekt von Classroom Analytics betrifft die mögliche Erfassung von Aufmerksamkeit (*attention monitoring*) oder emotionaler Verfasst-

heit (*affect recognition*) der im Klassenraum interagierenden Personen, insbesondere basierend auf der Analyse von Video- oder Audiodaten aus Klassenräumen. Auch wenn dies durchaus mit dem Ziel einer Verbesserung von Lernergebnissen verbunden sein kann, wird bezweifelt, dass Aufmerksamkeit und Emotionen jedenfalls mit aktueller Technik hinreichend genau, zuverlässig und ohne systematische Verzerrung gemessen werden können. Außerdem werden die vorstehend genannten Risiken der notwendigen Datenerfassung hier als besonders gravierend eingeschätzt.

81. Zusammenfassend lassen sich auf Seite der Chancen von KI in der Schule personalisiertes Lernen und Entlastung von Lehrkräften anführen, genau wie eine potenziell objektivere und fairere Bewertung von Lernergebnissen sowie mitunter verbesserte Zugangschancen und Möglichkeiten zur Inklusion von Lernenden mit besonderen Bedürfnissen. Zu den Risiken gehören neben den bereits erwähnten Bedenken hinsichtlich Verzerrungen und Beeinträchtigungen der Privatsphäre und der Autonomie auch Gefahren der Isolation und Vereinsamung von Lernenden sowie möglicherweise qualitative Veränderungen des Lernverhaltens. qualitativ verändert. So könnten sich etwa grundsätzliche Auswirkungen auf die Motivation und Fähigkeit von Schülerinnen und Schülern zu Lösung komplexerer Aufgaben ergeben.

82. KI-gestützte Lehr-Lernsysteme können den jeweiligen Lernprozess unterstützen, ersetzen aber nicht die personale Vermittlung und die personalen Aspekte von Bildung. Die Relevanz der Schule als Sozialraum der Interaktion zwischen Menschen ist dabei nicht zu unterschätzen. Da Bildung nicht nur in optimierbarer und berechenbarer Anhäufung von Wissen, sondern vor allem in einem konstruktiven und verantwortlichen Umgang mit erlerntem Wissen besteht, ist bei der Delegation von Elementen des Lehr-Lern-Geschehens an Maschinen besonders darauf zu achten, dass Lernprozesse, die zentral für die Persönlichkeitsbildung des Menschen sind, dadurch nicht vermindert werden.

83. Vor dem Hintergrund dieser Überlegungen legt der Deutsche Ethikrat elf Empfehlungen für den Einsatz von KI in der schulischen Bildung vor:

- *Empfehlung Bildung 1:* Digitalisierung ist kein Selbstzweck. Der Einsatz sollte nicht von technologischen Visionen, sondern von grundlegenden Vorstellungen von Bildung, die auch die Bildung der Persönlichkeit umfassen, geleitet sein. Die vorgestellten Tools sollten deshalb im Bildungsprozess kontrolliert und-als ein Element innerhalb der Beziehung zwischen Lehrenden und Lernenden eingesetzt werden.

- *Empfehlung Bildung 2:* Für jedes Einsatzgebiet gilt es, eine angemessene Abwägung von Chancen und Risiken vorzunehmen. Insbesondere sollten Autonomie und Privatheit von Lehrenden und Lernenden hohen Schutz erfahren. Besondere Chancen ergeben sich im Bereich der Inklusion und Teilhabe, wo das Potenzial dieser Systeme genutzt werden sollte, um etwa sprachliche oder räumliche Barrieren abzubauen.
- *Empfehlung Bildung 3:* Tools, die einzelne Elemente des Lehr- und Lernprozesses ersetzen bzw. ergänzen (enge Ersetzung) und nachweislich Fähigkeiten, Kompetenzen oder soziale Interaktion der Personen, die sie nutzen, erweitern, wie etwa einige intelligente Tutor-Systeme oder Telepräsenz-Roboter für externe Lehrbeteiligung, sind prinzipiell weniger problematisch als solche, die umfassendere bzw. weitere Teile des Bildungsprozesses ersetzen. Je höher der Ersetzungsgrad, desto strenger müssen Einsatzbereiche, Umgebungsfaktoren und Nutzenpotenziale evaluiert werden.
- *Empfehlung Bildung 4:* Es gilt standardisierte Zertifizierungssysteme zu entwickeln, die anhand transparenter Kriterien des Gelingens von Lernprozessen im genannten umfassenden Sinne Schulämter, Schulen und Lehrkräfte dabei unterstützen können, sich für oder gegen die Nutzung eines Produkts zu entscheiden. Hier kann sich auch der Empfehlung zur dauerhaften Einrichtung länderübergreifender Zentren für digitale Bildung, wie es im jüngsten Gutachten „Digitalisierung im Bildungssystem. Handlungsempfehlungen von der Kita bis zur Hochschule“ von der Ständigen Wissenschaftlichen Kommission der Kultusministerkonferenz angesprochen wurde, angeschlossen werden.
- *Empfehlung Bildung 5:* Bei der Entwicklung, Erprobung und Zertifizierung entsprechender KI-Produkte bedarf es einer engen Zusammenarbeit mit den relevanten Behörden, mit den jeweils zuständigen pädagogischen Fachgesellschaften sowie der Partizipation von Beteiligten, um Schwachstellen der Produkte frühzeitig zu entdecken und hohe Qualitätsstandards zu etablieren. Bekannte Herausforderungen KI-getriebener Technologien wie beispielsweise Verzerrungen bzw. Bias oder Anthropomorphisierungstendenzen sollten bei der Entwicklung und Standardisierung berücksichtigt werden.
- *Empfehlung Bildung 6:* Um den verantwortlichen Einsatz von KI-Technologien im Bildungsprozess zu gewährleisten, muss die Nutzungskompetenz insbesondere der Lehrkräfte erhöht werden; es bedarf der Entwicklung und Etablierung entsprechender Module und Curricula in der Aus-, Fort- und Weiterbildung. Insbesondere die Gefah-

ren eines verengten pädagogischen Ansatzes und eines Deskillings in der Lehre sollten dabei aktiv in den Blick genommen werden. Ebenso sollte die digitale Nutzungskompetenz von Lernenden sowie Eltern gestärkt und um KI-Aspekte erweitert werden.

- *Empfehlung Bildung 7:* Im Sinne der Beteiligungsgerechtigkeit sollten KI-basierte Tools Lernenden grundsätzlich auch für das Eigenstudium zur Verfügung stehen.
- *Empfehlung Bildung 8:* Die Einführung von KI-Tools im Bildungsbereich erfordert ferner den Ausbau verschiedener flankierender Forschungsbereiche. Sowohl theoretische Fundierung als auch empirische Evidenz zu Effekten, etwa auf die Kompetenzentwicklung (z. B. Problemlösen) oder zur Beeinflussung der Persönlichkeitsentwicklung bei Kindern und Heranwachsenden, müssen weiter ausgebaut werden. Dabei sollte nicht nur stärker in Forschung und entsprechende Produktentwicklung investiert, sondern vor allen Dingen auch die praktische Erprobung und Evaluation im schulischen Alltag verstärkt werden.
- *Empfehlung Bildung 9:* Des Weiteren stellt sich hier die Problematik der Datensouveränität. Zum einen sind bei der Sammlung, Verarbeitung und Weitergabe von bildungsbezogenen Daten strenge Anforderungen an den Schutz der Privatsphäre zu beachten. Zum anderen sollte die gemeinwohlorientierte, verantwortliche Sammlung und Nutzung von großen Daten, etwa in der prognostischen lehrunterstützenden Anwendung, ermöglicht werden.
- *Empfehlung Bildung 10:* Eine vollständige Ersetzung von Lehrkräften läuft dem hier skizzierten Verständnis von Bildung zuwider und ist auch nicht dadurch zu rechtfertigen, dass schon heute in bestimmten Bereichen ein akuter Personalmangel und eine schlechte (Aus-)Bildungssituation herrschen. In der komplexen Situation der schulischen Bildung bedarf es eines personalen Gegenübers, das mithilfe technischer Komponenten zwar immer stärker unterstützt werden kann, dadurch selbst als Verantwortungsträger für die pädagogische Begleitung und Evaluation des Bildungsprozesses aber nicht überflüssig wird.
- *Empfehlung Bildung 11:* In Anbetracht der erkenntnistheoretischen und ethischen Herausforderungen und unter Abwägung potenzieller Nutzen und Schäden stehen die Mitglieder des Deutschen Ethikrates dem Einsatz von Audio- und Videomonitoring im Klassenzimmer insgesamt skeptisch gegenüber. Insbesondere erscheint die Analyse von Aufmerksamkeit und Emotionen per Audio- und Videoüberwachung des Klassenraums mittels aktuell verfügbarer Technologien nicht vertretbar. Ein Teil des

Ethikrates schließt den Einsatz von Technologien zur Aufmerksamkeits- und Affekterkennung zukünftig jedoch nicht vollständig aus, sofern sichergestellt ist, dass die erfassten Daten eine wissenschaftlich nachweisbare Verbesserung des Lernprozesses bieten und das hierfür notwendige Monitoring von Lernenden und Lehrkräften keine inakzeptablen Auswirkungen auf deren Privatsphäre und Autonomie hat. Ein anderer Teil des Ethikrates hingegen hält die Auswirkungen auf Privatsphäre, Autonomie und Gerechtigkeit hingegen generell für nicht akzeptabel und befürwortet daher ein Verbot von Technologien zu Aufmerksamkeitsmonitoring und Affekterkennung in Schulen.

Öffentliche Kommunikation und Meinungsbildung

84. Durch die digitale Transformation verändern sich auch politisch relevante Kommunikationsprozesse. Die rasante Verbreitung digitaler Plattformen und Sozialer Medien mit ihren algorithmisch vermittelten Informations- und Kommunikationsangeboten wirkt sich nicht nur auf einzelne gesellschaftliche Sphären aus, sondern potenziell auch auf große Teile der öffentlichen Kommunikation und Meinungsbildung – mit Konsequenzen für das demokratische Legitimationsgefüge.

85. Viele Plattformen bieten inzwischen sich ähnelnde Möglichkeiten an, multimediale Inhalte zu erstellen und zu verbreiten, auf die Inhalte anderer zu reagieren, sich mit anderen Personen auszutauschen und die Plattform nach Inhalten zu durchsuchen oder diese zu abonnieren. Auch Optionen, eigene Inhalte gezielt zu bewerben und Produkte und Dienstleistungen direkt anzubieten oder zu kaufen, sind vielfach vorhanden. Fast alle weiter verbreiteten Plattformen und Dienste werden von privaten Unternehmen aus den USA oder China betrieben, und die größten Sozialen Netzwerke gehören nur wenigen Firmen. Aufgrund dieser Marktmacht sowie der Vielseitigkeit und Integration der Dienste funktionieren viele Angebote inzwischen als reichhaltige soziotechnische Infrastrukturen, in denen sich ein Großteil des Online-Nutzungsverhaltens nach den Vorgaben weniger Konzerne abspielt.

86. Mit der Fülle der Informationen und Interaktionsmöglichkeiten in Sozialen Medien gehen technische Herausforderungen und ökonomische Potenziale einher, die gemeinsam zur Gestaltung aktueller Funktionsweisen und Geschäftsmodelle beigetragen haben. Die Fülle der Inhalte stellt Plattformen wie auch Kundschaft vor das Problem der Informationsauswahl. Diese wird aktuell überwiegend an Algorithmen delegiert, die dafür sorgen,

dass jeder Person beim Besuch einer Plattform auf sie persönlich zugeschnittene Inhalte in einer bestimmten Reihenfolge angezeigt werden.

87. Die Kriterien, nach denen solche Algorithmen ihre Auswahl treffen, sind eng mit ökonomischen Faktoren verknüpft. Die meisten Plattformen und Dienste folgen einem werbebasierten Geschäftsmodell, das am besten funktioniert, wenn die Interessen der einzelnen Nutzerinnen und Nutzer erstens möglichst präzise bekannt sind und Menschen zweitens möglichst viel Zeit auf der Plattform verbringen, während derer ihnen auf persönliche Interessen zugeschnittene Werbung präsentiert wird. Daher lohnt es sich für Plattformen, so viele Datenspuren wie möglich über den persönlichen Hintergrund, die Interessen, das Nutzungsverhalten und das soziale Netzwerk der Personen, die es nutzen, zu sammeln und für die Auswahl personalisierter Inhalte zu verwenden (Profiling).

88. Eine algorithmisch gesteuerte personalisierte Informationsauswahl, in der ökonomische und aufmerksamkeitsbasierte Faktoren derart eng verbunden sind und die sich anhand des Nutzungsverhaltens ständig weiterentwickelt, führt dazu, dass Inhalte, die besonders sensationell erscheinen oder intensive emotionale Reaktionen auslösen, sich überproportional schnell und weit verbreiten. Dies begünstigt unter anderem Falschnachrichten und Inhalte wie Hassrede, Beleidigungen und Volksverhetzungen.

89. In Reaktion auf die Herausforderung, wie mit solchen potenziell problematischen und gleichzeitig verbreitungsstarken Inhalte umzugehen ist, bemühen sich Plattformen darum, ihre Inhalte nach verschiedenen Kriterien zu moderieren (Content Moderation). Hierbei kommen sowohl Menschen als auch algorithmische Systeme zum Einsatz. Grundlage für die Moderation sind rechtliche Vorgaben sowie plattformeigene Kommunikationsregeln, auf deren Grundlage auch rechtlich zulässige Inhalte gelöscht, gesperrt oder in ihrer Reichweite eingeschränkt werden können.

90. Menschliche Moderation erfolgt typischerweise durch Personen, die bei Drittanbietern angestellt sind, mit denen eine Plattform vertraglich zusammenarbeitet. Diese Personen werden bei oftmals prekären Arbeitsbedingungen mit häufig extrem belastendem Material wie Tötungen, Kindesmissbrauch, Tierquälerei und Suizid konfrontiert. Zudem müssen sie innerhalb weniger Sekunden sprachlich und kulturell komplexe Nuancen berücksichtigen, von denen die Zulässigkeit eines Beitrags entscheidend abhängen kann.

91. Algorithmen können im Gegensatz dazu anstößige Inhalte herausfiltern, ohne dass diese durch Menschen angesehen werden müssen, und darüber hinaus mit der unübersichtli-

chen Menge an Daten und Inhalten im Netz besser umgehen. Allerdings sind automatisierte Methoden jedenfalls bislang häufig unzureichend, um den kulturellen und sozialen Zusammenhang einer Äußerung einzubeziehen und diese damit adäquat zu beurteilen. Aufgrund der aktuellen rechtlichen Anreizstruktur besteht die Gefahr, dass systematisch auch Inhalte gelöscht oder unzugänglich gemacht werden, die nicht gegen Regeln verstoßen (*overblocking*).

92. Durch die beschriebenen Funktionsweisen von Plattformen und die sich dabei entfaltenden soziotechnischen Verquickungen können menschliche Handlungsfähigkeiten in unterschiedlicher Weise erweitert oder vermindert werden. Die Delegation von Kuratierungs- und Moderationsprozessen an Algorithmen ist mit Komfort- und Effizienzgewinnen verbunden und kann eine Erweiterung von Handlungsmöglichkeiten bedeuten, wenn beispielsweise Informationen und persönliche Ziele besser oder schneller erreicht werden können oder aufgrund der effektiven Delegation der Inhaltsauswahl an Algorithmen Entlastungseffekte auftreten, die Freiräume für andere Aktivitäten schaffen.

93. Eine Verminderung menschlicher Handlungsspielräume und persönlicher Freiheit kann sich ergeben, wo es Menschen schwerfällt, sich dem Sog von Plattformangeboten zu entziehen und ihre Nutzung dieser Angebote auf ein für sie gesundes Maß zu beschränken. Zudem kann eine algorithmische Kuratierung Autorschaft vermindern, wenn eine rationale Auseinandersetzung mit Alternativen durch die algorithmische Vorwegnahme bestimmter Relevanzentscheidungen nur noch eingeschränkt stattfinden kann.

94. Neben diesen allgemeinen Auswirkungen der Funktionsweisen von Plattformen und Sozialen Medien verändern sich durch sie auch die Informationsqualität und Diskursqualität, welche wichtige Grundlagen der öffentlichen Meinungsbildung sind – mit potenziell weitreichenden Konsequenzen für Prozesse der politischen Willensbildung. Wie weit verbreitet und wirkmächtig die nachfolgend genannten Effekte sind, lässt sich aktuell zwar noch nicht abschließend beurteilen, auch weil die Datenlage mitunter unklar oder widersprüchlich ist. Ein genauerer Blick auf die postulierten Mechanismen lohnt jedoch schon deswegen, weil die von ihnen berührten Prozesse grundlegend für unsere Demokratie sind.

95. Mit Blick auf die Informationsqualität ist zunächst auf die positive Erweiterung vieler Informationsmöglichkeiten zu verweisen. Demgegenüber wird vielfach die Sorge geäußert, dass die derzeit gängigen Praktiken algorithmischer Kuratierung auch negative Auswirkungen haben und die Verbreitung von Falschnachrichten und Verschwörungstheorien fördern, zur Entstehung von Filterblasen und Echokammern beitragen und Inhalte

priorisieren, die negative emotionale und moralische Reaktionen und Interaktionen provozieren.

96. Trotz der genannten Unsicherheiten über das Ausmaß der beschriebenen Effekte erscheint es plausibel, dass Falschnachrichten, Filterblasen und Echokammern sowie eine emotional-moralische Zuspitzung vieler Inhalte negative Auswirkung auf die Informationsqualität entfalten können. Die Freiheit, qualitativ hochwertige Informationen zu finden, wird unter diesen Umständen durch die Wirkmacht der zum Einsatz kommenden Algorithmen praktisch vermindert.

97. Änderungen in der Qualität, Darbietung und Verbreitung algorithmisch vermittelter Informationen betreffen auch die Diskursqualität in ethisch wie politisch relevanter Hinsicht. Auch hier sind zunächst wieder positive Entwicklungen und Potenziale zu benennen, die sich insbesondere aus den auf Plattformen und in Sozialen Medien wesentlich erhöhten Möglichkeiten zu Teilhabe und direkter Vernetzung ergeben. Gegenüber den genannten Chancen werden jedoch auch mit Blick auf die Diskursqualität negative Entwicklungen diskutiert. Dabei geht es vor allem um drei Themen: politische Polarisierung öffentlicher Diskurse; politische Werbung und Manipulation; und das Spannungsfeld von Diskursverrohungen und überbordenden Eingriffen in die Äußerungs- und Meinungsfreiheit.

98. Es gibt viele Hinweise, dass die beschriebene Verbreitungsfähigkeit emotional und moralisch aufgeladener Inhalte zu Tonfallverschiebungen geführt hat, auch und insbesondere auf Kanälen, die aktiv zur Gestaltung des politischen Diskurses beitragen. Nachdem beispielsweise geänderte Auswahlkriterien auf Facebook dazu führten, dass sich künftig vor allem solche Inhalte erfolgreich verbreiteten, auf die Menschen besonders aufgebracht reagieren, verschärften viele politische Kommunikationsteams den Tonfall ihrer Beiträge, um diesen Kriterien gerecht zu werden.

99. Auf Plattformen ergibt sich zudem viel Potenzial für besonders wirkmächtige Kommunikationskampagnen, die eingebettet in den digitalen Alltag ablaufen, ohne dass Nutzerinnen und Nutzer sich dessen gewahr werden. Die reichhaltigen datenbasierten Profile, die sich aus dem Nutzungsverhalten auf Plattformen erstellen lassen, können auch genutzt werden, um zielgenaue politische Werbung zu schalten (*targeted advertisement*) oder um Menschen strategisch zu desinformieren oder von der Wahl abzuhalten. Wie erfolgreich solche auch als *Microtargeting* bezeichneten Ansätze sind, ist zwar noch nicht hinreichend erforscht, doch allein das Wissen darum, dass versucht wird, auf Grundlage sehr

persönlicher psychologischer Merkmale politische Präferenzen zu manipulieren, kann plausibel negative Effekte auf den politischen Diskurs und das Vertrauen in politische Meinungsbildungsprozesse entfalten.

100. Vertrauensschädigend kann sich weiterhin der Umstand auswirken, dass zur strategischen Beeinflussung des öffentlichen politischen Diskurses auch vielfach unechte Profile (Fake-Accounts) eingesetzt werden, die teilweise automatisiert betriebenen werden (Bots). Kommunikationskampagnen, die solche unechten Profile nutzen, können damit Botschaften effektiv verstärken, ihnen somit mehr Überzeugungskraft verleihen und Diskurse mitunter problematisch verzerren.

101. Der bereits beschriebene Trend zur Verschärfung von Tonlagen auf Plattformen und in Sozialen Medien geht mit der Sorge einher, dass eine Zunahme stark negativ und aggressiv geprägter Kommunikationsstile bis hin zu Hassrede, Drohungen und Gewaltaufforderungen zu einer Verrohung des politischen Diskurses beitragen kann. Selbst wenn online verbreitete Hetze nicht in Handlungen in der realen Welt umschlägt, kann es auch hier zu Chilling-Effekten kommen. Wo nämlich solche Äußerungen so viel Unbehagen und Angst schüren, dass dies Personen davon abhält, sich am öffentlichen Diskurs zu beteiligen, wirkt dies auf die Freiheit und Handlungsmöglichkeiten Betroffener in der Online-Kommunikation vermindernd.

102. Andererseits werden auch Bemühungen, potenziell problematische Inhalte mit Moderationsmaßnahmen einzudämmen, teils kritisch beurteilt, denn solche Reaktionen werfen ihrerseits demokratietheoretische Fragen auf. Übermäßige Löschungen und Sperrungen können einen Eingriff in die Meinungs- und Pressefreiheit darstellen und selbst zu Chilling-Effekten beitragen, nämlich dann, wenn Menschen bestimmte Inhalte gar nicht erst veröffentlichen, weil sie befürchten, dass diese Inhalte gleich wieder gelöscht oder gar ihre Accounts (zeitweise) gesperrt werden könnten.

103. In der Zusammenschau können die hier aufgezeigten Phänomene und Entwicklungen, die sich in den soziotechnischen Infrastrukturen digitaler Netzwerke vollziehen, erhebliche Auswirkungen auf Prozesse der öffentlichen Kommunikation sowie der politischen Meinungs- und Willensbildung entfalten, auch und vielleicht insbesondere in demokratischen Gesellschaften. Vor diesem Hintergrund legt der Deutsche Ethikrat zu diesem Anwendungsfeld von KI zehn Empfehlungen vor:

- *Empfehlung Kommunikation 1:* Regulierung Sozialer Medien: Es bedarf klarer rechtlicher Vorgaben, in welcher Form und in welchem Ausmaß Soziale Medien und Plattformen über ihre Funktions- und Vorgehensweisen zur Kuratierung und Moderation von Inhalten informieren müssen und wie dies auf der Grundlage institutioneller Regelungen umgesetzt wird. Dies muss durch externe Kontrollen überprüfbar sein; rein freiwillige Ansätze privater Handelnder, insbesondere die unverbindliche Überprüfung durch von diesen selbst besetzen Aufsichtsgremien, sind nicht ausreichend. Hier gibt es auf Ebene der Europäischen Union im Digital Services Act bereits Ansätze, die aber noch nicht weit genug gehen.
- *Empfehlung Kommunikation 2:* Transparenz über Moderations- und Kuratierungspraktiken: Anstelle allgemeiner Moderations- und Löschungsrichtlinien und wenig aussagekräftigen Zahlen über Löschungen muss für externe Kontrollen nachvollziehbar sein, wie, unter welchen Umständen und anhand welcher Kriterien solche Entscheidungen gefällt und umgesetzt werden und welche Rolle hierbei Algorithmen bzw. menschliche Moderierende übernommen haben. Darüber hinaus, müssen auch die grundlegenden Funktionsweisen der Kuratierung von Inhalten Sozialer Medien und Plattformen in dem Ausmaß offengelegt werden, das nötig ist, um systemische Verzerrungen und möglicherweise resultierende informationelle Dysfunktionen erkennen zu können. Die Berichtspflichten und Transparenzvorgaben im Medienstaatsvertrag, im Netzwerkdurchsetzungsgesetz und im Digital Services Act stellen dies noch nicht hinreichend sicher. Die datenschutzrechtlichen Auskunftspflichten gemäß Art. 12 ff. DSGVO sind zum Teil auf nationalstaatliche Ebene beschränkt worden und erfassen oftmals diese weitergehenden Aspekte nicht.
- *Empfehlung Kommunikation 3:* Zugriff auf wissenschaftsrelevante Daten von Plattformen: Um die Wirkungsweisen von Plattformen und Sozialen Medien, ihren Einfluss auf öffentliche Diskurse, aber auch weitere Themen von hoher gesellschaftlicher Relevanz zu untersuchen, sollte sichergestellt werden, dass unabhängigen Forschenden der Zugriff auf wissenschaftsrelevante Daten von Plattformen nicht mit dem pauschalen Verweis auf Betriebs- oder Geschäftsgeheimnisse verweigert werden kann. Für den Zugang müssen sichere, datenschutzkonforme sowie forschungsethisch integre Wege gefunden werden. Netzwerkdurchsetzungsgesetz und Digital Services Act enthalten bereits Reglungen zum Datenzugang, die aber in ihrem Anwendungsbereich sehr begrenzt sind; auch der Data Act sieht vergleichbare Regelungen vor.

- *Empfehlung Kommunikation 4:* Berücksichtigung von Sicherheit, Datenschutz und Geheimhaltungsinteressen: Anforderungen an Offenlegungen und Datenzugang müssen kontextsensitiv spezifiziert werden, wobei Anforderungen an Sicherheit und Schutz vor Missbrauch, Datenschutz sowie dem Schutz von intellektuellem Eigentum und Geschäftsgeheimnissen angemessen Rechnung zu tragen ist. Je nach Kontext muss zwischen unterschiedlich klar definierten Zeitpunkten der Prüfung und Graden der Offenlegung unterschieden werden.
- *Empfehlung Kommunikation 5:* Personalisierte Werbung, Profiling und Microtargeting: Personalisierte Werbung ist das zentrale Geschäftsmodell Sozialer Medien und Plattformen. Die Praktiken des Profiling und Microtargeting können jedoch problematische Auswirkungen auf öffentliche Kommunikation und Meinungsbildung entfalten, insbesondere im Kontext politischer Werbung. Um solche negativen Auswirkungen durch effektive Regelungen zu verhindern, ist es zunächst notwendig, die Bedingungen für eine Erforschung und Überprüfung der Zusammenhänge zwischen Geschäftsmodellen und Praktiken algorithmischer Kuratierung in ihren Wirkungsweisen und Effekten zu schaffen. Der auf Ebene der Europäischen Union diskutierte Vorschlag für eine Verordnung über die Transparenz und das Targeting politischer Werbung adressiert diesen Bedarf. Hierbei zeigen sich allerdings auch die Herausforderungen, Regeln so zuzuschneiden, dass sie einerseits wirksam sind, andererseits aber die Freiheit der politischen Kommunikation nicht übermäßig beschränken.
- *Empfehlung Kommunikation 6:* Bessere Regulierung von Online-Marketing und Datenhandel: Ursache vieler der in diesem Kapitel beschriebenen informationellen und kommunikativen Dysfunktionen haben ihre Ursache im Online-Marketing, welches das grundlegende Geschäftsmodell vieler Sozialer Medien und Plattformen ist und auf der Sammlung, Analyse und dem Verkauf vielfältiger Daten über die Personen, die diese Angebote nutzen, beruht. Das Problem ist hierbei nicht die Werbefinanzierung per se, sondern der invasive Umgang mit diesen Daten. Hier gilt es einerseits, die Auswirkungen dieses Geschäftsmodells auf öffentliche Diskurse besser zu erforschen. Andererseits bedarf es besserer gesetzlicher Regelungen, um sowohl Individuen in ihren Grundrechten online effektiver schützen als auch negative systemische Effekte auf den öffentlichen Diskurs zu minimieren. In diese Richtung gehende Vorschläge hat der Deutsche Ethikrat unter dem Stichwort Datensouveränität in seiner Stellungnahme Big Data und Gesundheit vorgestellt. Europäische Regelungen wie der Digital Markets Act adressieren das Problem der Datenmacht großer Plattformen,

aber – schon aus Gründen der Regelungskompetenz – nicht mit Blick auf die Folgen für den öffentlichen Diskurs.

- *Empfehlung Kommunikation 7:* Machtbeschränkung und Kontrolle: Unternehmen, die im Bereich der öffentlichen Vorstellung von Daten bzw. Tatsachen de facto monopolartige Machtmöglichkeiten haben, sind durch rechtliche Vorgaben und entsprechende Kontrolle auf Pluralismus, Minderheiten- und Diskriminierungsschutz zu verpflichten. Ein Teil der Mitglieder des Deutschen Ethikrates ist der Auffassung, dass medienrechtliche Regelungen zur Sicherung von Pluralität, Neutralität und Objektivität generell auf Nachrichtenfunktionen von Sozialen Medien und Plattformen, ausgedehnt werden sollten, sofern sie denen traditioneller Medien ähneln.
- *Empfehlung Kommunikation 8:* Erweiterung der Nutzerautonomie: Plattformen und Soziale Medien sollten ihre Inhalte auch ohne eine personalisierte Kuratierung verfügbar machen. Darüber hinaus sollten sie für die Kriterien, nach denen Inhalte auf Plattformen und in Sozialen Medien algorithmisch ausgewählt und prioritär präsentiert werden, weitere Wahlmöglichkeiten anbieten. Dazu sollte auch die Möglichkeit gehören, bewusst Gegenpositionen angezeigt zu bekommen, die den bisher geäußerten eigenen Präferenzen zuwiderlaufen. Solche Wahlmöglichkeiten sollten gut sichtbar und leicht zugänglich sein.
- *Empfehlung Kommunikation 9:* Förderung kritischer Rezeption von Inhalten: Zur Eindämmung unreflektierter Verbreitung fragwürdiger Inhalte sollten diverse Hinweisfunktionen entwickelt und eingesetzt werden, die eine kritische Auseinandersetzung mit Material fördern, bevor man sich dafür entscheidet, es zu teilen oder öffentlich darauf zu reagieren. Dies könnten etwa Rückfragen sein, ob Texte gelesen und Videos geschaut wurden, bevor man sie teilt, oder Angaben zur Seriosität von Quellen.
- *Empfehlung Kommunikation 10:* Alternative Informations- und Kommunikationsinfrastruktur: Zu erwägen wäre, den privaten Social-Media-Angeboten im europäischen Rahmen eine digitale Kommunikationsinfrastruktur in öffentlich-rechtlicher Verantwortung zur Seite zu stellen, deren Betrieb sich nicht am Unternehmensinteresse eines möglichst langen Verweilens von Menschen auf der Plattform oder an anderen kommerziellen Interessen orientiert. Damit sollte nicht etwa der öffentlich-rechtliche Rundfunk (TV und Radio) auf eine weitere digitale Plattform ausgedehnt, sondern eine digitale Infrastruktur bereitgestellt werden, die eine Alternative zu den kommerz-

betriebenen, stark oligopolartigen Angeboten bietet. Um eine hinreichende Staatsferne zu garantieren, könnte auch an eine Trägerschaft in Gestalt einer öffentlichen Stiftung gedacht werden.

Öffentliche Verwaltung

104. Für viele Menschen und Organisationen stellt die öffentliche Verwaltung, so etwa im Finanz-, Steuer-, Melde- und Sozialwesen und in der Straffälligen- und Jugendgerichtshilfe, die unmittelbar erfahrbare Staatsgewalt dar. Funktionierende, transparente, als legitim anerkannte und bürgernahe Verwaltung ist für ein funktionierendes Gemeinwesen und die Akzeptanz von Demokratie und Staat wesentlich. Mit Digitalisierungsstrategien in diesem Bereich verbinden sich Hoffnungen auf eine Rationalisierung und Beschleunigung staatlichen Verwaltungshandelns, eine effektivere und kohärentere Datennutzung, sowie eine Ausweitung der Einbeziehung wissenschaftlichen, sowie bürgerschaftlichen Sachverstandes. Dem steht die dystopische Schreckensvision einer sogenannten „Algokratie“ gegenüber, in der autonome Softwaresysteme die staatliche Herrschaft über Menschen ausüben.

105. Vielfach und zunehmend werden in der öffentlichen Verwaltung automatisierte Entscheidungssysteme (ADM-Systeme, *automatic* bzw. *algorithmic decision making systems)* eingesetzt, etwa zur Bewertung von Arbeitsmarktchancen, bei der Prüfung und Vergabe von Sozialleistungen oder für Vorhersagen im Bereich der Polizeiarbeit. Von besonderem Interesse ist hier zum einen, inwieweit der Einsatz von KI-Systemen menschliche Handlungsfähigkeiten und Autorschaft beeinflusst. Angesichts der häufig beobachteten Tendenz, sich maschinellen Empfehlungen vorbehaltslos anzuschließen (Automation Bias) kann bereits die Nutzung von Software zur Entscheidungsunterstützung in der Verwaltung weitreichende Wirkung entfalten.

106. Andere Fragen betreffen vor allem Aspekte von Gerechtigkeit, beispielsweise wenn es darum geht, ob und in welchem Umfang die verwendeten Systeme Diagnosen und Prognosen tatsächlich verbessern, ob die Genauigkeit für verschiedene Anwendungskontexte oder für verschiedene Personengruppen gleich ist, oder ob es systematische Verzerrungen oder Diskriminierungen gibt (*algorithmic bias*). Ebenso können datenbasierte Systeme jedoch historischen Ungerechtigkeiten oder menschliche Vorurteile aufdecken und sie damit Gegenmaßnahmen zugänglich zu machen.

107. Eine grundsätzliche Grenze für die Anwendung von automatisierten Entscheidungssystemen liegt in nicht eliminierbaren normativen Ziel- oder Regelkonflikten im deutschen, deontologisch verfassten Rechtssystem, in dem die Folgenabwägung nie allein das Rechtmäßige bestimmt, sondern unbedingte Ansprüche auf Schutz der Person zu wahren sind und der Algorithmisierung ethischer und rechtlicher Entscheidungsprozesse Grenzen setzen.

108. Das Sozialwesen ist ein Bereich der Verwaltung, in dem Entscheidungen mit weitreichenden Folgen für die Betroffenen fallen, etwa über die Gewährung von Hilfen, bei Maßnahmen im Kontext einer Kindeswohlgefährdung oder bei der Abschätzung von Gefährdungspotentialen von Straftätern in der Bewährungshilfe. Algorithmenbasierte Entscheidungshilfen kommen hier zunehmend zum Einsatz und können professionelle Handlungskompetenz erweitern, wenn sie Fachkräften helfen, ihre sonst oft intuitiven Einschätzungen auf eine solidere Datengrundlage zu stellen bei Bedarf zu korrigieren und Entscheidungen so evidenzbasiert zu standardisieren. Dies ist besonders wichtig bei der Abschätzung von Gefährdungspotenzialen, beispielsweise bei Verdacht auf Kindeswohlgefährdung oder in der Bewährungshilfe.

109. Menschliche Autorschaft kann unter Zuhilfenahme sachdienlicher Ergebnisse von KI-Algorithmen allerdings auch vermindert werden. Auf der professionellen Seite kann dies beispielsweise dann der Fall sein, wenn es zu einer ungeprüften Übernahme algorithmisch vorgeschlagener Ergebnisse kommt (Automation Bias). Auch für die von den Entscheidungen betroffenen Personen sind negative Effekte möglich, etwa wenn ihnen aufgrund von durch Verzerrungen geprägten algorithmisch unterstützten Entscheidungen Handlungs- oder Entwicklungsmöglichkeiten ungerechtfertigterweise genommen werden.

110. Gerade bei der Erfassung von Hilfebedarfen birgt der Einsatz algorithmischer Systeme zudem das Risiko der Entkopplung aus einer dialogischen Beziehungsarbeit, die entscheidend für die Erfahrung von Selbstwirksamkeit Betroffener sein kann. Wird diese persönliche Ebene bei der Ermittlung des individuellen Hilfebedarfs im Zuge einer algorithmenbasierter Informatisierung des Sozialwesens vernachlässigt, können positive Effekte selbst materieller Hilfeleistungen schnell verpuffen und damit kaum nachhaltig wirken. Das österreichische AMAS-System beispielsweise, das für Arbeitssuchende Erfolgsprognosen für eine Wiedereingliederung in den Arbeitsmarkt berechnet, ist für seine Ausrichtung an den Werten, Normen und Zielen einer restriktiven Fiskalpolitik kritisiert worden,

die den Zielen eines personenorientierten Hilfesystems, welches individuelle Hilfebedarfe betroffener Personen fokussieren muss, diametral zuwiderläuft.

111. Ein anderer Bereich, in dem algorithmenbasierte Risikoanalysen zunehmend zum Einsatz kommen, ist die Kriminalitätsbekämpfung. Im *Predictive Policing* unterstützen algorithmenbasierte Anwendungen präventive Polizeiarbeit mittels Prognosen künftiger Straftaten, straffälliger Personen und Tatorte, um Verbrechen zu verhindern. Vor allem personenbezogene Verfahren werden in diesem Zusammenhang kontrovers diskutiert. Einerseits geht damit die Hoffnung besserer polizeilicher Arbeit und eines besseren Schutzes möglicher Opfer einher. Andererseits können Fehler und Verzerrungen in algorithmenbasierter Verbrechensbekämpfung mit besonders folgenschweren Konsequenzen für ungerechtfertigt klassifizierte Personen verbunden sein und besteht hier eine besonders große Gefahr, dass Fehler und Verzerrungen in der Software systembedingt besondere Breitenwirkung entfalten.

112. Ein weiteres Problem ist der Schutz der Privatsphäre im Kontext von Predictive Policing. Die für die Polizeiarbeit herangezogenen Daten sind in aller Regel besonders sensibel. Insbesondere bei sogenannten Chatkontrollen zur Prävention und Bekämpfung des sexuellen Missbrauchs von Kindern, zu denen die Kommission der Europäischen Union im Mai 2022 einen Verordnungsvorschlag vorgelegt hat, wird hinterfragt, ob eine anlasslose und flächendeckende Überwachung privater Kommunikation gerechtfertigt werden kann oder einen unverhältnismäßig intensiven Eingriff in die Grundrechte darstellt.

113. Nicht zuletzt wird die Sorge geäußert, dass mit algorithmengesteuerter Polizeiarbeit das Risiko der Verfestigung eines mechanischen Menschenbildes einhergehen könne, welches den einzelnen Menschen verobjektiviere, seine Individualität auf eine datengetriebene Klassifikation reduziere, jedoch die gesamtgesellschaftlichen Ursachen von Kriminalität unberücksichtigt lasse.

114. In der Zusammenschau führen automatisierte Entscheidungsverfahren in der Öffentlichen Verwaltung zu neuen Möglichkeiten und Herausforderungen, die erheblich weiter reichende ethische und demokratietheoretische Fragen aufwerfen, etwa in Bezug auf Nachvollziehbarkeit, Erklärbarkeit und Vertrauenswürdigkeit im Verwaltungshandeln, aber auch in Bezug auf Sorgen um Diskriminierung und Technokratie, in der menschliche Kommunikation und Abwägung hinter anonymen Datenmengen und standardisierten Benutzeroberflächen verschwindet.

115. Insofern der Rückgriff auf große Datenmengen und ihre zielgerichtete Auswertung bessere Entscheidungsgrundlagen schafft, kann der Einsatz algorithmischer Systeme zu diesem Zweck menschliche Autorschaft unterstützen und ist in ethischer Hinsicht grundsätzlich zu begrüßen. Eine unkritische Übernahme von Systemempfehlungen droht menschliche Autorschaft allerdings zu vermindern, bis im ungünstigen Fall nur noch ein automatisiertes Geschehen verbleibt, in dem technische Systeme für Betroffene weitreichende, teils existenzielle Festlegungen treffen und systemimmanente Fehler oder Verzerrungen möglicherweise nicht mehr erkannt werden.

116. Beim Einsatz von KI in der Öffentlichen Verwaltung muss daher kontextbezogen im Detail eingeschätzt und abgewogen werden, welche Auswirkungen eine entsprechende Maßnahme auf die Autorschaft unterschiedlichster Beteiligter und Betroffener hat, welche Konflikte auftreten und wie mit ihnen umgegangen werden kann oder soll. Hierzu legt der Deutsche Ethikrat neun Empfehlungen vor:

- *Empfehlung Verwaltung 1:* Die mit automatisierten Entscheidungshilfen (ADM-Systemen) einhergehende verstärkte Standardisierung und pauschale Kategorisierung von Einzelfällen muss umso stärker hinterfragt und um spezifisch einzelfallbezogene Erwägungen ergänzt werden, je intensiver die betroffene Entscheidung individuelle Rechtspositionen berührt.
- *Empfehlung Verwaltung 2:* Es müssen geeignete technische und organisatorische Instrumente zur Vorkehrung gegen die manifeste Gefahr eines Automation Bias bereitgestellt werden, die es den Fachkräften erschweren, selbst bei einer Letztentscheidungskompetenz der algorithmischen Entscheidungsempfehlung unbesehen zu folgen. Es ist zu prüfen, ob eine Umkehrung der Begründungspflicht (nicht eine Abweichung, sondern ein Befolgen ist zu rechtfertigen) hier eine geeignete Vorkehrung sein kann.
- *Empfehlung Verwaltung 3:* Aufgrund ihrer Grundrechtsbindung sind an staatliche Einrichtungen bei der Entwicklung und Nutzung algorithmischer Systeme hohe Anforderungen in Bezug auf Transparenz und Nachvollziehbarkeit zu stellen, um den Schutz vor Diskriminierung zu gewährleisten sowie Begründungspflichten erfüllen zu können.
- *Empfehlung Verwaltung 4:* Für Softwaresysteme in der Öffentlichen Verwaltung müssen Qualitätskriterien verbindlich und transparent festgelegt werden (in Bezug auf Genauigkeit, Fehlervermeidung, Unverzerrtheit und so weiter). Ebenso bedarf es einer Dokumentation der jeweils eingesetzten Methoden. Diesbezüglich sollten auch

aktuelle Beschaffungspraktiken, in deren Verlauf staatliche Behörden Softwarelösungen kaufen, einer kritischen Prüfung unterzogen werden.

- *Empfehlung Verwaltung 5:* Überall dort, wo algorithmische Systeme Einsatz in der Öffentlichen Verwaltung finden, gilt es, Sorge zu tragen, dass die Personen, die diese Systeme anwenden, über die erforderlichen Kompetenzen im Umgang damit verfügen. Dazu gehört neben Kenntnis der Verwendungsweisen auch das Wissen um die Limitationen und möglichen Verzerrungen, um Systeme angemessen einsetzen zu können.
- *Empfehlung Verwaltung 6:* Die Einsichts- und Einspruchsrechte Betroffener müssen auch beim Einsatz algorithmischer Systeme effektiv gewährleistet werden. Dazu bedarf es gegebenenfalls weiterer wirksamer Verfahren und Institutionen.
- *Empfehlung Verwaltung 7:* In Öffentlichkeit, Politik und Verwaltung sollte eine Sensibilisierung gegenüber möglichen Gefahren von Automatisierungssystemen, wie etwa Verletzungen der Privatsphäre oder Formen systematisierter Diskriminierung, erfolgen. Dazu gehört eine öffentliche Debatte darüber, ob es in bestimmten Kontexten überhaupt einer technischen Lösung bedarf.
- *Empfehlung Verwaltung 8:* Im Bereich des Sozialwesens ist sicherzustellen, dass ADM-Systeme elementare fachliche Standards von sozialprofessionellen Interaktionen (z. B. gemeinsame Sozialdiagnose oder Hilfeplanung als *Teil* therapeutischer bzw. unterstützender Hilfeleistung) nicht unterlaufen oder verdrängen. Dies beinhaltet insbesondere Maßnahmen, die Vergröberungen individueller Fallkonstellationen und -prognosen durch die ADM-induzierte grobklassifizierende Einteilung von Fall- und/oder Leistungsberechtigtengruppen verhindern. Dabei ist Sorge zu tragen, dass die Feststellung individueller Hilfebedarfe nicht erschwert wird und es zu keiner schleichenden Aushöhlung der sozialrechtlich gebotenen Identifizierung individueller Hilfebedarfe zugunsten einseitiger externer Interessen an Gefahrenminimierung oder Kostenersparnis kommt.
- *Empfehlung Verwaltung 9:* Die Arbeit von Gefahrenabwehrbehörden einschließlich der Polizei betrifft besonders grundrechtssensible Bereiche. Dies wirkt sich auf die Reichweite eines zulässigen Einsatzes von algorithmischen Systemen in der prädiktiven Polizeiarbeit aus. Risiken wie Verletzungen der Privatsphäre oder potenziell unzulässige Diskriminierungen der von dem Einsatz betroffenen Personen müssen mit Chancen auf erhebliche Verbesserungen der staatlichen Gefahrenabwehr sorgfältig abgewogen und in ein angemessenes Verhältnis gebracht werden. Hierfür erforderli-

che gesellschaftliche Aushandlungsprozesse sollten umfangreich geführt werden. Dabei ist der diffizilen Bestimmung des Verhältnisses von Freiheit und Sicherheit Rechnung zu tragen. Jegliche Gesetzesübertretung zu verhindern, wäre mit rechtsstaatlichen Mitteln nicht möglich.

TEIL III: QUERSCHNITTSTHEMEN UND ÜBERGREIFENDE EMPFEHLUNGEN

Zusammenfassung der bisherigen Analyse

117. Der Begriff der Künstlichen Intelligenz hat in der öffentlichen Debatte zunehmend an Aufmerksamkeit gewonnen und wird mit teils überzogenen Hoffnungen aber auch mit teilweise fehlgeleiteten Befürchtungen verknüpft. Der Deutsche Ethikrat geht von einem normativ grundlegenden Unterschied zwischen Mensch und Maschine aus. Softwaresysteme verfügen weder über theoretische noch über praktische Vernunft. Sie handeln oder entscheiden nicht selbst und können keine Verantwortung übernehmen. Sie sind kein personales Gegenüber, auch dann nicht, wenn sie Anteilnahme, Kooperationsbereitschaft oder Einsichtsfähigkeit simulieren.

118. Menschliche Vernunft ist immer zugleich eingebunden in die konkrete soziale Mit- und Umwelt. Nur so ist zu erklären, dass sie handlungswirksam wird. Vernünftig handelt der einzelne Mensch als Teil einer sozialen Mitwelt und einer kulturellen Umgebung. Schon deshalb kann den in dieser Stellungnahme besprochenen Softwaresystemen weder theoretische noch praktische Vernunft zugeschrieben werden.

119. Menschen entwickeln digitale Technik und nutzen sie als Mittel zu menschlichen Zwecken. Jedoch wirken diese Technologien zurück auf menschliche Handlungsmöglichkeiten. Sie können einerseits neue Optionen eröffnen aber andererseits auch Anpassungen erforderlich machen, die nicht wünschenswert sind. Auch wenn Maschinen also nicht selbst handeln, so verändern sie die Handlungsfähigkeit von Menschen tiefgreifend und können Handlungsmöglichkeiten erheblich erweitern oder vermindern.

120. Ziel der Delegation menschlicher Tätigkeiten an Maschinen sollte prinzipiell die Erweiterung menschlicher Handlungsfähigkeit und Autorschaft sein. Ihre Verminderung sowie eine Diffusion oder Evasion von Verantwortung gilt es hingegen zu verhindern. Dafür muss die Übertragung menschlicher Tätigkeiten auf KI-Systeme gegenüber den Betroffenen hinreichend transparent erfolgen, sodass wichtige Entscheidungselemente, -parameter oder -bedingungen nachvollziehbar bleiben.

121. Um über Wert und Nutzen der Delegation vormals menschlichen Handelns an Maschinen ethisch zu befinden, bedarf es daher immer eines kontextspezifischen Blicks, der die Perspektiven unterschiedlicher Beteiligter und Betroffener ebenso berücksichtigt wie die langfristigen Auswirkungen solcher Übertragungen. Die Herausforderungen stecken also wie so oft im Detail, genauer: in den Details der Technik, der Einsatzkontexte sowie der institutionellen und sozio-technischen Umgebung.

122. Um diesen kontextspezifischen Blick zu ermöglichen hat sich der Deutsche Ethikrat in dieser Stellungnahme exemplarisch mit Anwendungen in der Medizin, der schulischen Bildung, der öffentlichen Kommunikation sowie der Verwaltung beschäftigt. Es wurden bewusst Sektoren ausgewählt, in denen die Durchdringung durch KI-basierte Technologien sehr unterschiedlich ausfällt und in denen sich jeweils unterschiedliche Ausmaße des Ersetzens vormals menschlicher Handlungen durch KI veranschaulichen lassen. In allen vier Sektoren sind Einsatzszenarien durch teils erhebliche Beziehungs- und Machtasymmetrien gekennzeichnet, was einen verantwortungsvollen Einsatz von KI und die Berücksichtigung der Interessen und des Wohls insbesondere vulnerabler Personengruppen umso wichtiger macht. Diese Unterschiedlichkeit der Art und Weise des KI-Einsatzes sowie des Ausmaßes der Delegation an Maschinen in den Blick zu nehmen, erlaubt es nuancierte ethische Betrachtungen anzustellen.

Entfaltung von Querschnittsthemen und Empfehlungen

123. Die Darstellung der sozio-technischen Entwicklungen und deren ethische Analyse in den vier Anwendungsbereichen zeigen, dass es eine Reihe von Querschnittsthemen und -herausforderungen gibt, die sich durch alle vier Bereiche ziehen, wenn auch teils in unterschiedlicher Weise und Ausprägung. Um im Hinblick auf die Erweiterung menschlicher Handlungsfähigkeit und Autorschaft zukünftig einen guten gesellschaftlichen Umgang mit KI zu gewährleisten, müssen solche Querschnittsfragen nicht nur innerhalb einzelner Bereiche angegangen werden, sondern darüber hinaus auch in vernetzten, bereichsübergreifenden Ansätzen.

124. Solches gleichermaßen horizontales wie vertikales, gestaltendes Denken stellt eine Herausforderung insbesondere für die Politikgestaltung und etwaige zukünftige Regulierung dar. Die Darstellung der Querschnittsthemen in dieser Stellungnahme, die für jedes Thema in einer Empfehlung münden, soll daher als Anregung für eine breitere Debatte

dienen, wie für zukünftige Politik- und Technikgestaltung gleichzeitig und im Zusammenspiel mit sektoralen Aspekten immer auch übergreifende Fragen in den Blick genommen werden können und müssen.

125. Im *ersten Querschnittsthema* geht es noch einmal um das in dieser Stellungnahme zentrale Konzept der Erweiterung und Verminderung menschlicher Handlungsmöglichkeiten. Zwar besteht eine sektorenübergreifende Gemeinsamkeit hinsichtlich der angestrebten Erweiterung menschlicher Handlungspotenziale darin, dass die komplette Ersetzung menschlicher Akteure durch KI-Systeme sich überall dort verbietet, wo die konkrete zwischenmenschliche Begegnung eine notwendige Voraussetzung für die Erreichung der jeweiligen Handlungsziele darstellt. Darüber hinaus besteht jedoch die Notwendigkeit, die Unterschiede beim KI-Einsatz in den einzelnen Handlungsbereichen sorgfältig zu beachten.

    - *Empfehlung Querschnittsthema 1:* Da die Vor- und Nachteile von KI-Anwendungen für verschiedene Personengruppen sowie die Gefahr des Verlustes bestimmter Kompetenzen bei den Personen, die solche Systeme anwenden, erheblich variieren, bedarf es sowohl einer differenzierten Planung des KI-Einsatzes in unterschiedlichen Handlungsfeldern, welche die jeweiligen Zielsetzungen und Verantwortlichkeiten präzise benennt, als auch einer zeitnahen Evaluation der tatsächlichen Folgen eines solchen Einsatzes, um die Systeme besser an die spezifischen Handlungskontexte anzupassen und sie fortlaufend zu verbessern.

126. Das *zweite Querschnittsthema* behandelt Wissenserzeugung durch KI und der Umgang mit KI-gestützten Voraussagen. Zentral ist dabei die Prämisse, dass Korrelationen und Datenmuster nicht mit Erklärungen und Begründungen von Ursachen von Ereignissen gleichzusetzen sind, sondern auch qualitativ evaluiert und normativ beurteilt werden müssen. Bei probabilistischen Methoden bleiben immer Restunsicherheiten über deren Akzeptabilität zu entscheiden ist. Ethisch positiv zu werten ist, dass durch KI-Einsatz in allen vier hier betrachteten Anwendungsbereichen erhebliche funktionale Verbesserungen erreicht wurden und weiterhin erwartbar sind. Es wird jedoch eine grundsätzlich normativ problematische Schwelle überschritten, wenn funktionale Verbesserungen (eventuell sogar unbemerkt) in eine Ersetzung moralischer Kompetenz und damit verbundener Verantwortung hinübergleiten.

    - *Empfehlung Querschnittsthema 2:* Der Einsatz KI-gestützter digitaler Techniken ist im Sinne der Entscheidungsunterstützung und nicht der Entscheidungsersetzung zu

gestalten, um Diffusion von Verantwortung zu verhindern. Er darf nicht zulasten effektiver Kontrolloptionen gehen. Den von algorithmisch gestützten Entscheidungen Betroffenen ist insbesondere in Bereichen mit hoher Eingriffstiefe die Möglichkeit des Zugangs zu den Entscheidungsgrundlagen zu gewähren. Das setzt voraus, dass am Ende der technischen Prozeduren entscheidungsbefugte Personen sichtbar bleiben, die in der Lage und verpflichtet sind, Verantwortung zu übernehmen.

127. Das *dritte Querschnittsthema* betrachtet die Gefährdung des Individuums durch statistische Stratifizierung. Grundlage vieler KI-Anwendungen sind Korrelationen, die bei der Analyse großer Datenmengen entdeckt werden und anhand derer man Einzelpersonen Kohorten mit bestimmten Merkmalskombinationen zuordnen kann. Die Bildung solcher Kohorten und die auf ihrer Basis durch Algorithmen produzierten Voraussagen kann die Qualität und Effektivität einer Anwendung insgesamt verbessern. Sie kann aber auch Probleme für Individuen bedeuten, welche von solchen kollektiven Schlüssen betroffen sind – insbesondere dann, wenn die statistisch getroffene Diagnose oder Prognose in ihrem Fall nicht zutrifft.

- *Empfehlung Querschnittsthema 3:* Neben einer Analyse der konkreten und naheliegenden Probleme datenbasierter Software, beispielsweise in Bezug auf den Schutz der Privatsphäre oder die Verhinderung von Diskriminierung, gilt es, auch die langfristigen Auswirkungen dieser statistischen Präkonfiguration von Individuen sowie deren Rückwirkung – im Sinne einer Erweiterung oder Verminderung der Handlungsmöglichkeiten – auf Individuen wie Kollektive für alle Sektoren sorgfältig zu beleuchten.

  Darüber hinaus gilt, dass Einzelfallbeurteilungen grundsätzlich wichtig bleiben. KI-basierte Beurteilungen und Vorhersagen können unter günstigen Bedingungen ein Hilfsmittel sein, aber kein geeignetes Instrument der *definitiven* Lagebeurteilung und Entscheidung. Pragmatische und heuristische Faktoren wie Prüfung der Kohärenz mit anderen Evidenzquellen, Erfolgseinschätzungen und anderes spielen eine nicht zu vernachlässigende Rolle.

128. Im *vierten Querschnittsthema* geht es um die Auswirkungen von KI auf menschliche Kompetenzen und Fertigkeiten. Deren Erwerb und Erhalt kann durch die Delegation menschlicher Tätigkeiten an Maschinen gefährdet werden. Weil die Nutzung von KI-Anwendungen (wie auch bei anderen Technologien) dazu führen kann, dass menschliche Fähigkeiten nachlassen bzw. ganz verkümmern, können Abhängigkeiten von diesen

Technologien entstehen. Handelt es sich dabei um gesellschaftlich besonders bedeutsame oder kritische Einsatzbereiche, ist ein Verlust von menschlichen Kompetenzen und Fertigkeiten ein ernstzunehmendes Risiko.

- *Empfehlung Querschnittsthema 4:* Ob und inwiefern beim Einsatz von KI-Anwendungen Verluste menschlicher Kompetenz auftreten, die als unerwünscht eingestuft werden, muss sorgfältig beobachtet werden. Bei der Entwicklung und dem Einsatz neuer Technologien sind solch unerwünschte Kompetenzverluste durch eine sinnvolle Gestaltung des Zusammenspiels von Mensch und Technik, durch angemessene institutionelle und organisatorische Rahmenbedingungen sowie durch gezielte Gegenmaßnahmen wie etwa spezifische Trainingsprogramme zu minimieren bzw. zu kompensieren. Kompetenzverluste können sowohl individueller als auch kollektiver Natur sein. So gilt es zu verhindern, dass die Delegation von Aufgaben an Technologien dazu führt, dass Gesellschaften übermäßig anfällig werden, wenn diese Technologien (zeitweise) ausfallen. Jenseits dieser systemischen Aspekte müssen negative Auswirkungen solcher Delegation auf die individuelle Autonomie oder Selbstwahrnehmung mitigiert werden.

129. Das *fünfte Querschnittsthema* befasst sich mit dem Schutz von Privatsphäre und Autonomie versus Gefahren durch Überwachung und Chilling-Effekte. Die im Rahmen vieler KI-Anwendungen notwendige Erfassung großer Mengen an personenbezogenen Daten sowie die Möglichkeit auf ihrer Basis sensible Prognosen zu erstellen, beeinträchtigt nicht nur die Privatsphäre der Personen, von denen diese Daten stammen, sondern macht sie auch vulnerabel gegenüber möglichen Benachteiligungen oder Manipulation, welche aus der Verarbeitung der Daten resultieren können. Chilling-Effekte beschreiben in diesem Kontext Rückwirkungen auf das Verhalten von Menschen, die Sorge haben, dass ihr Verhalten beobachtet, aufgezeichnet oder ausgewertet wird.

- *Empfehlung Querschnittsthema 5:* Die beschriebenen Phänomene sollten in ihrer Entstehung, Ausprägung und Entwicklung umfassend empirisch untersucht werden. Um sowohl dem Problem der Überwachung sowie den parallelen Gefahren durch etwaige Chilling-Effekte Rechnung zu tragen, müssen angemessene und effektive rechtliche und technische (beispielsweise *privacy by design*) Vorkehrungen getroffen werden, die dem übermäßigen Tracking von Onlineverhalten und dem Handel mit personenbeziehbaren Daten Einhalt gebieten. Die Interessen der Datensubjekte müssen hierbei im Mittelpunkt stehen. Insbesondere ist dabei auf besonders vulnerable Gruppen zu

achten, da viele der Einsatzkontexte zudem von asymmetrischen Machtverhältnissen gekennzeichnet sind. Es muss Sorge getragen werden, dass die Erweiterung der Handlungsmöglichkeiten einiger nicht zulasten der Verminderung der Handlungsmöglichkeiten anderer, insbesondere benachteiligter Gruppen stattfindet.

130. Das *sechste Querschnittsthema* greift Konzepte von Datensouveränität und gemeinwohlorientierter Datennutzung auf, die der Deutsche Ethikrat bereist 2017 in seiner Stellungnahme zum Thema Big Data und Gesundheit entwickelt hat. Dabei geht es um die Suche nach Lösungen, wie im Kontext von KI-Anwendungen Daten sinnvoll für verschiedene wichtige Zwecke genutzt werden können, ohne zugleich den Schutz der Privatsphäre der Datengeber unzulässig zu beeinträchtigen. Hier stellt sich die Frage, ob das derzeitige Datenschutzrecht, bzw. die herrschende Datenschutzpraxis, diesen beiden Zielen gerecht wird. Während in manchen Handlungsfeldern berechtigte Sorgen vor unbemerkten und weitreichenden Verletzungen von Privatsphäre und informationeller Selbstbestimmung herrschen, werden in anderen Kontexten durch strenge Auslegungen von Datenschutzregeln wichtige soziale Güter, etwa mit Blick auf Patientenversorgung und wissenschaftlichen Erkenntnisgewinn, aber auch der kommunalen Daseinsvorsorge, nicht oder nur sehr schwer erreicht.

- *Empfehlung Querschnittsthema 6:* Mit Blick auf KI-Anwendungen müssen neue Wege gefunden werden, um innerhalb der jeweiligen Kontexte und mit Blick auf die jeweils spezifischen Herausforderungen und Nutzenpotenziale die gemeinwohlorientierte Daten(sekundär)nutzung zu vereinfachen bzw. zu ermöglichen und damit die Handlungsoptionen auf diesem Gebiet zu erweitern. Zugleich ist es essenziell, einen Bewusstseinswandel sowohl in der Öffentlichkeit als auch bei den praktisch tätigen Personen, die Datennutzung gestalten, herbeizuführen – weg von einer vornehmlich individualistisch geprägten und damit verkürzten Perspektive, hin zu einer Haltung, die auch systematische und gemeinwohlbasierte Überlegungen mit einbezieht und in einen Ausgleich bringt. Eine solche Haltung ist auch für die zukünftige Politikgestaltung und Regulierung deutlich stärker als bisher zugrunde zu legen. Nur so kann es gelingen, neben den Risiken, die sich aus breiterer KI-Anwendung ohne Zweifel ergeben, zugleich die wichtigen Chancen einer verantwortlichen Nutzung nicht aus dem Blick zu verlieren.

131. Das *siebte Querschnittsthema* betrachtet kritische Infrastrukturen, Abhängigkeiten und Resilienz. Im Zuge der Digitalisierung werden Infrastrukturen, wie beispielsweise Stromnetze, zunehmend digital überwacht und über das Internet gesteuert. Gleichzeitig werden digitale Technologien selbst zu Infrastrukturen. Am Vorhandensein und Funktionieren von Infrastrukturen richten Menschen ihr Handeln aus, und im Zuge dieser sozialen Aneignung entstehen Abhängigkeiten, die menschliche Autonomie gefährden können. Wenn KI-gestützte Systeme zusehends in die Steuerung von Infrastrukturen integriert werden, kommt hinzu, dass KI-Systeme nicht vollständig transparent und nachvollziehbar sind, und durch die fortwährende Komplexitätssteigerung von Infrastruktursystemen und ihrer Steuerung steigt die gesellschaftliche und institutionelle Vulnerabilität weiter an.

    - *Empfehlung Querschnittsthema 7:* Um die Autorschaft menschlicher Akteure und deren Handlungsmöglichkeiten zu erweitern, muss die Resilienz sozio-technischer Infrastrukturen gestärkt und die Abhängigkeit von individuellen Akteuren und Systemen minimiert werden. Dies umfasst zunächst die Notwendigkeit, die infrastrukturelle Bedeutung digitaler Technologien anzuerkennen und infolgedessen dem Schutz und der Resilienz kritischer digitaler Infrastrukturen mehr Aufmerksamkeit zuteilwerden zu lassen, auch im politischen Handeln. In allen Sektoren gilt es, einseitige Abhängigkeiten zu vermeiden, welche im Krisenfalle verletzlich und angreifbar machen.

      Für Nutzerinnen und Nutzer erfordert eine Verringerung der Abhängigkeit die Möglichkeit, zwischen Alternativen zu wählen, ohne große Teile der Funktionalität einzubüßen. Dies umfasst zum einen die Notwendigkeit von Interoperabilität, um einfach zwischen Systemen wechseln zu können. Hierfür ist auch der Auf- und Ausbau alternativer Infrastrukturen von besonderer Bedeutung. Im Kontext der öffentlichen Meinungsbildung erscheint die Etablierung unabhängiger, öffentlicher digital-kommunikativer Plattformen dringend geboten. Aber auch in anderen Sektoren wie der Verwaltung, der Bildung oder der Medizin vermindert eine zu große Abhängigkeit von wenigen Systemen oder Akteuren potenziell die individuelle wie kollektive Handlungsfähigkeit.

132. Das *achte Querschnittsthema* dreht sich um Pfadabhängigkeiten, Zweitverwertung und Missbrauchsgefahren. Pfadabhängigkeiten entstehen, wenn Entscheidungen, die zu Beginn einer bestimmten Entwicklung getroffen wurden, noch lange nachwirken und teils schwer wieder aufzuheben sind, auch wenn sich der Kontext der Nutzung möglicherweise

geändert hat. Sind Technologien einmal eingeführt, dürfte zudem eine Tendenz auszumachen zu sein, deren Möglichkeiten voll auszuschöpfen – auch über das ursprüngliche Anwendungsfeld hinaus. Solche Zweitverwertungen sind nicht prinzipiell problematisch, doch sobald eine Technologie etabliert ist, kann es schwer sein, weitere, auch missbräuchliche Nutzungsszenarien auszuschließen. Gerade digitale Technologien und insbesondere Grundlagentechnologien wie das maschinelle Lernen eröffnen oft sehr mannigfaltige Nutzungsmöglichkeiten, in denen die Frage der Abgrenzung von Ge- und Missbrauch zunehmend schwieriger wird.

- *Empfehlung Querschnittsthema 8:* Bei Technologien mit großen Auswirkungen oder hohem Verbreitungsgrad und vor allem dort, wo sich eine Nutzung von Technologien kaum oder gar nicht vermeiden lässt, müssen bereits zu Beginn der Entwicklungsplanung mögliche Langzeitfolgen wie Pfadabhängigkeiten im Allgemeinen sowie Dual-Use-Potenziale im Speziellen regelhaft und explizit mitgedacht und antizipiert werden. Dies gilt in besonderem Maße in der Anwendungsplanung. Dabei sind neben direkten, sektorspezifischen Schadenspotenzialen auch etwaige – natürlich deutlich schwieriger fass- und antizipierbare – sektorübergreifende Effekte zu bedenken. Hohe Standards für die Sicherheit und den Schutz der Privatsphäre (*security by design, privacy by design*) können ebenfalls dazu beitragen, spätere missbräuchliche Anwendungen einzuhegen bzw. möglichst zu verhindern.

  Bei besonders invasiven Technologien beispielsweise in der öffentlichen Verwaltung, die Bürgerinnen und Bürger gegebenenfalls verpflichtend nutzen müssen, sind besonders hohe Standards einzuhalten. Um dies sicherzustellen und überprüfen zu können, sind gegebenenfalls Open-Source-Ansätze angezeigt (vgl. Abschnitt 10. 10.).

133. Im *neunten Querschnittsthema* geht es um Bias und Diskriminierung. Datenbasierte KI-Systeme lernen auf Basis vorhandener Daten. Resultierende Prognosen und Empfehlungen schreiben somit die Vergangenheit in die Zukunft fort, wodurch Stereotypen, aber auch bestehende gesellschaftliche Ungleichheiten und Ungerechtigkeiten durch den Einbau in scheinbar neutrale Technologien reproduziert und sogar verstärkt werden können. Oft liegt bei der Entwicklung von KI-Systemen keine unmittelbare Diskriminierungsabsicht vor, sondern diskriminierende Effekte entstehen aus gesellschaftlichen Realitäten oder Stereotypen in Kombination mit technisch-methodischen Entscheidungen. Es ist allerdings zumindest denkbar, dass auch explizite Diskriminierungsabsichten in komplexen Systemen versteckt werden könnten.

- *Empfehlung Querschnittsthema 9:* Zum Schutz vor Diskriminierung in Anbetracht der zuvor dargelegten Herausforderungen bedarf es *angemessener Aufsicht und Kontrolle* von KI-Systemen. Besonders in sensiblen Bereichen erfordert dies den Auf- oder Ausbau gut ausgestatteter Institutionen. Hier gilt: je größer die Eingriffstiefe und je unumgänglicher die Systeme, desto höher die Anforderungen an Diskriminierungsminimierung.

  Auch bereits bei der Entwicklung von Technologien gilt es, Diskriminierung zu minimieren bzw. Fairness, Transparenz und Nachvollziehbarkeit herzustellen. Dies sollte sowohl durch Anreize – etwa Forschungsförderung – als auch durch entsprechende gesetzliche Anforderungen befördert werden, etwa hinsichtlich der Offenlegung, welche Maßnahmen zur Diskriminierungsminimierung bei der Softwareentwicklung ergriffen wurden.

  Allerdings haben technische wie regulatorische Maßnahmen zur Minimierung von Diskriminierung ihre Grenzen, unter anderem weil unterschiedliche Fairnessziele technisch nicht gleichzeitig erfüllt werden können. Es müssen also zugleich ethische und politische Entscheidungen getroffen werden, welche Kriterien für Gerechtigkeit in welchem Kontext zum Tragen kommen sollen. Diese Entscheidungen dürfen nicht den Personen, die Software entwickeln, und anderen direkt Beteiligten überlassen werden. Stattdessen bedarf es der Entwicklung geeigneter Verfahren und Institutionen, um diese Kriterien kontextspezifisch und demokratisch, gegebenenfalls immer wieder neu auszuhandeln. Je nach Anwendungskontext und Sensibilität des einzusetzenden Systems kann die Beteiligung der Öffentlichkeit erforderlich sein. Dabei sollte der Schutz der jeweils bedürftigsten bzw. von Entscheidungen besonders betroffenen Gruppen besonders berücksichtigt werden.

134. Das *zehnte Querschnittsthema* greift Fragen von Transparenz und Nachvollziehbarkeit sowie von Kontrolle und Verantwortung auf. Die häufige Undurchschaubarkeit von KI-Systemen hat verschiedene Ursachen, die vom Schutz geistigen Eigentums über die Komplexität und Nicht-Nachvollziehbarkeit der Verfahren bis hin zur mangelnden Durchsichtigkeit von Entscheidungsstrukturen, in die der Einsatz algorithmischer Systeme eingebettet ist, reichen. Die Transparenz und Nachvollziehbarkeit algorithmischer Systeme steht zwar in Zusammenhang mit deren Kontrolle und der Verantwortung für ihren Einsatz, ist für beides aber weder zwingend notwendig, noch hinreichend.

- *Empfehlung Querschnittsthema 10:* Es bedarf der Entwicklung ausgewogener aufgaben-, adressaten- und kontextspezifischer Standards für Transparenz, Erklärbarkeit und Nachvollziehbarkeit und ihrer Bedeutung für Kontrolle und Verantwortung sowie für deren Umsetzung durch verbindliche technische und organisatorische Vorgaben. Dabei muss den Anforderungen an Sicherheit und Schutz vor Missbrauch, Datenschutz sowie dem Schutz von intellektuellem Eigentum und Geschäftsgeheimnissen in angemessener Weise Rechnung getragen werden. Je nach Kontext sind hier unterschiedliche Zeitpunkte (ex-ante, ex-post, real-time) sowie unterschiedliche Verfahren und Grade der Offenlegung zu spezifizieren.

135. Zusammenfassend geht es in dieser Stellungnahme um die Auswirkungen einer zunehmenden Delegation menschlicher Tätigkeiten an digitale Technologien, insbesondere KI-basierte Softwaresysteme. In zahlreichen Beispielen aus den Bereichen der Medizin, der schulischen Bildung, der öffentlichen Kommunikation und Meinungsbildung sowie der öffentlichen Verwaltung zeigt sich, dass dieses Delegieren sowohl mit Erweiterungen als auch mit Verminderungen menschlicher Handlungsmöglichkeiten einhergeht und sich dadurch sowohl förderlich als auch hinderlich auf die Realisierung menschlicher Autorschaft auswirken kann.

136. Ziel und Richtschnur ethischer Bewertung muss dabei immer die Stärkung menschlicher Autorschaft sein. Dabei ist zu berücksichtigen, dass die Erweiterung von Handlungsmöglichkeiten für eine Personengruppe mit deren Verminderung für andere einhergehen kann. Diesen unterschiedlichen Effekten ist Rechnung zu tragen, insbesondere in Hinblick auf den Schutz und die Verbesserung der Lebensbedingungen vulnerabler oder benachteiligter Gruppen. Letztlich zeigt sich, dass die normativen Anforderungen an die Gestaltung und den Einsatz solcher Technologien, zum Beispiel in Bezug auf Anforderungen hinsichtlich Transparenz und Nachvollziehbarkeit, den Schutz der Privatsphäre sowie die Verhinderung von Diskriminierung, zwar in allen Bereichen und für alle Betroffenen von hoher Bedeutung sind, sie jedoch sektor-, kontext-, und adressatenspezifisch konkretisiert werden müssen, um angemessen zu sein und wirksam werden zu können.

+++ Vorabfassung! Wird durch die lektorierte Fassung ersetzt. +++

1 Einleitung

Digitale Technologien im Allgemeinen und Systeme sogenannter Künstlicher Intelligenz (KI) im Speziellen durchdringen zunehmend unsere Lebenswelt. Dies reicht von der suchmaschinenbasierten Sortierung von Ergebnissen und Empfehlungen für Filme und Musik über Navigationssoftware bis hin zur Nutzung von Risikoprofilen und Software zur Entscheidungsunterstützung im Sozial- und Justizwesen oder bei der Polizei. Wettervorhersagen und Klimamodellierung, Krebsdiagnostik in der Medizin und psychotherapeutische Erstversorgung mittels Chatbots, intelligente Tutorsysteme zum Vokabellernen oder Emotionserkennung in Videoanalysen – datenbasierte Analysen, Prognosen und das Delegieren von Entscheidungen an Softwaresysteme verdeutlichen, dass digitale Technologien und insbesondere auch KI in nahezu alle Bereiche des öffentlichen und privaten Lebens Einzug gehalten haben. Auch die Debatten um den im November 2022 vorgestellten Chatbot ChatGPT und andere Anwendungen sogenannter *generativer KI,* welche automatisiert neue Inhalte in einer Qualität produzieren, bei der oftmals nicht mehr erkennbar ist, dass diese rein maschinell erstellt wurden, zeigen, dass eine grundlegende Auseinandersetzung mit den Wechselwirkungen zwischen Mensch und Maschine erforderlich ist.

Für die ethische Bewertung solcher Technologien und ihres Einsatzes in verschiedenen Bereichen ist es nötig, nicht nur die Technologien selbst zu verstehen. Es gilt darüber hinaus, auch ihre Wechselwirkungen mit den Personen, die sie verwenden oder von ihrer Anwendung betroffen sind, zu analysieren und den Blick auf gesellschaftliche Effekte zu richten.

Hierbei stellen sich Fragen wie zum Beispiel: Welche – positiven wie auch negativen – Auswirkungen hat das Delegieren von Tätigkeiten, die zuvor Menschen vorbehalten waren, an Maschinen? Werden menschliche Autorschaft und die Bedingungen für verantwortliches Handeln erweitert oder vermindert? Wie wirkt sich der Einbezug digitaler Technologien und Künstlicher Intelligenz auf die Handlungsoptionen verschiedener Beteiligter und Betroffener aus? Wessen Möglichkeiten werden durch den Einsatz erweitert, wessen Möglichkeiten vermindert? Diesen Fragen geht der Deutsche Ethikrat in dieser Stellungnahme zum Thema Mensch und Maschine nach und schreibt damit Themen fort, die er bereits in seinen Stellungnahmen zu Big Data und Gesundheit (2017) sowie Robotik und Pflege (2019) angeschnitten hat. Mit den aktuellen Ausführungen reagiert der Deutsche Ethikrat zudem auf eine im Oktober 2020 vom Präsidenten des Deutschen Bundestages formulierte Bitte, die Arbeit der Enquete-Kommissionen zur Künstlichen Intelligenz und zur beruflichen Bildung in der digitalen Arbeitswelt um eine grund-

legende Einbettung des politischen und gesellschaftlichen Diskurses zum Thema KI im Rahmen einer multidisziplinären Stellungnahme zu den ethischen Fragen des Verhältnisses von Mensch und Maschine zu ergänzen.

Grundfrage und Maßstab der hier vorgelegten ethischen Bewertung technologischer Entwicklungen und ihres Einsatzes in verschiedenen Kontexten ist, ob durch die Delegation von Tätigkeiten an Maschinen – bis hin zu einer möglichen Ersetzung – die Bedingungen für verantwortliches Handeln und menschliche Autorschaft erweitert oder vermindert werden.

Die Stellungnahme gliedert sich in drei Teile:

Teil I dieser Stellungnahme liefert die technischen, theoretischen und methodischen Grundlagen. In Kapitel 2 werden in knapper Form die *technologischen Grundlagen* der in dieser Stellungnahme behandelten Entwicklungen dargelegt. Hierzu wird zunächst ein kurzer historischer Überblick über die Entstehungsgeschichte des Forschungsfeldes der Künstlichen Intelligenz seit den 1950er-Jahren nachgezeichnet. Gerade in Anbetracht der oft schillernden öffentlichen Debatte rund um sogenannte *starke* oder *generelle* Künstliche Intelligenz werden zentrale Begriffe und Debatten eingeordnet. Im nächsten Schritt werden bedeutsame Entwicklungen der letzten Jahre skizziert, in denen die Steigerung der Rechenleistung, die Miniaturisierung in der Computertechnik und die zunehmende Vernetzung von Geräten dazu geführt haben, dass unsere Lebenswelt immer stärker von digitalen Technologien durchdrungen ist, was wiederum eine wachsende Datenflut und immer neue Einsatzmöglichkeiten für algorithmische Systeme nach sich zieht. Das Kapitel endet mit einem knappen Verweis auf existierende Ethik-Leitlinien verschiedenster Akteursgruppen und den aktuellen Rechtsrahmen.

Kapitel 3 widmet sich den *zentralen Begriffen und philosophischen Grundlagen* dieser Stellungnahme. Hierfür werden zunächst zentrale Begriffe beleuchtet, die für das Verständnis sowohl der Unterschiede als auch der Wechselwirkungen zwischen Menschen und Maschinen bedeutsam sind. Dies umfasst zunächst den Begriff der *Intelligenz* selbst und sein Verhältnis zu Begriffen wie *praktischer* und *theoretischer Vernunft*. Von zentraler Bedeutung sind ferner die Begriffe der *Handlung* und der *Verantwortung*. Softwaresysteme werden zunehmend zur Unterstützung von Handlungen und Entscheidungen herangezogen oder diese werden gar vollständig an sie delegiert. In der Folge stellt sich auch die Frage, welche Auswirkungen diese Prozesse für Verantwortungsübernahme und -zuschreibung haben. Das Kapitel endet mit einem Blick auf die anthropologischen Aspekte des Mensch-Maschine-Verhältnisses und beleuchtet zentrale Differenzen zwischen Menschen und Maschinen.

In Kapitel 4 hingegen geht es weniger um die Unterschiede zwischen Menschen und Maschinen als vielmehr um deren vielfältige und mehrstufige Relationen und Wechselwirkungen. Nach einer kurzen Einordnung des im Deutschen Ethikrat verwendeten Verständnisses dieser Wechselwirkungen und dessen Einordnung zwischen den Polen des Technikdeterminismus und des Sozialkonstruktivismus, wird mit den Begriffen *Erweitern, Vermindern und Ersetzen* eine Matrix zur Beschreibung und Bewertung von Technikentwicklung und Technikeinsatz entwickelt. Diese Matrix bildet die Grundlage für die Analysen im folgenden zweiten Teil der Stellungnahme, in dem die Auswirkungen des Delegierens von menschlichen Handlungssegmenten bis hin zu einem vollständigen Ersetzen von Menschen durch Maschinen in verschiedenen Kontexten und Sektoren auf menschliche Autorschaft und die Bedingungen für verantwortliches Handeln untersucht werden.

Teil II dieser Stellungnahme werden die zuvor angestellten Überlegungen anhand von Analysen in vier ausgewählten Anwendungsfeldern exemplarisch konkretisiert: dem Bereich der *Medizin* (Kapitel 5), dem Bereich der *schulischen Bildung* (Kapitel 6), dem Bereich der *öffentlichen Kommunikation und Meinungsbildung* (Kapitel 7) sowie dem Bereich der *öffentlichen Verwaltung* (Kapitel 8). In allen vier Sektoren werden datenbasierte Software-Systeme eingesetzt. Allerdings unterscheidet sich die Tiefe und Breite der Durchdringung deutlich.

Während die für den Bereich der öffentlichen Kommunikation und Meinungsbildung so zentralen Sozialen Medien als Paradebeispiel einer sehr umfassenden Delegation vormals menschlicher Tätigkeiten an Algorithmen dienen können, beispielsweise hinsichtlich der Kuratierung und Moderation von Inhalten, zeigt sich in Deutschland im Bereich der schulischen Bildung noch eine vergleichsweise geringe Nutzung digitaler Technologien im Allgemeinen und KI-basierter Systeme im Speziellen. Eine vollständige Ersetzung menschlicher Tätigkeiten durch Maschinen scheint hier in weiter Ferne. In der Medizin stellt sich dies wiederum anders dar, werden doch zunehmend nicht nur einzelne Bearbeitungsschritte, sondern ganze Funktionen an KI-basierte Softwaresysteme delegiert. Dies reicht von der Nutzung KI-basierter Softwaresysteme zur Mustererkennung in der Krebsdiagnostik bis zu Chatbots als maschinellen Ersatz von therapeutischem Fachpersonal. Auch in der öffentlichen Verwaltung hat KI in Gestalt datenbasierter Software zur Erstellung von Risikoprofilen und zur Entscheidungsunterstützung Einzug gehalten.

Zentrale Einsicht dieser Analysen ist, dass Entscheidungen über die beste Form und das richtige Ausmaß der Delegation von Tätigkeiten und Funktionen an Softwaresysteme und KI nur kontext-, anwendungs- und personenbezogen spezifiziert werden können. Entsprechend enden alle

vier Kapitel mit sektorspezifischen Empfehlungen. Als Richtschnur der Bewertung gilt hierbei jedoch immer,

- ob die Delegation zu einer *Erweiterung* der Handlungsmöglichkeiten, zu einer *Erhöhung* der Möglichkeiten für verantwortliches Handeln und Autorschaft der verschiedenen Beteiligten und Betroffenen führt oder
- ob es möglicherweise zu einer *Verminderung* von Handlungsmöglichkeiten sowie negativen Auswirkungen auf Möglichkeiten der Autorschaft und Verantwortungsübernahme kommt.

Der Vergleich der Analysen in den vier ausgewählten Anwendungsfeldern erlaubt es, auch übergreifende Themen auszumachen, da es wiederkehrende Aspekte gibt, wenngleich sich diese in den vier Anwendungsfeldern sehr unterschiedlich darstellen. Diese *Querschnittsthemen* und mit ihnen verbundene übergreifende Empfehlungen werden in Teil III der Stellungnahme dargelegt. Dieser letzte Teil beginnt zunächst mit einer Rekapitulation der anthropologischen und ethischen Orientierung, welche dieser Stellungnahme zugrunde liegt, und fasst die Einsichten aus den vorausgehenden Kapiteln im zweiten Teil der Stellungnahme knapp zusammen. Sodann werden die zehn identifizierten Querschnittsthemen dargelegt, die jeweils auch übergreifende Empfehlungen enthalten.

Das erste Querschnittsthema beschäftigt sich mit der für diese Stellungnahme leitenden Grundfrage, welche Auswirkungen das Delegieren von Handlungen an Maschinen auf die Erweiterung oder Verminderung von Handlungsmöglichkeiten von Menschen hat. Hierbei wird deutlich, dass Vor- und Nachteile einer solchen Delegation sich nicht nur zwischen verschiedenen Sektoren, sondern auch für verschiedene Personengruppen stark unterscheiden. In der Folge muss ein verantwortungsvoller Einsatz von KI diese Nuancen reflektieren und berücksichtigen.

Querschnittsthema 2 adressiert die Auswirkungen von KI auf Wissenserzeugung und den Umgang mit KI-gestützten Voraussagen. Um eine Diffusion von Verantwortung zu verhindern, ist es hierbei von zentraler Bedeutung, KI-gestützte digitale Technologien zur Entscheidungsunterstützung und nicht zur Entscheidungsersetzung einzusetzen.

Querschnittsthema 3 untersucht, inwieweit das Individuum durch statistische Stratifizierung gefährdet ist vor dem Hintergrund, dass es bei datenbasierten Analysen und Prognosen oftmals als Teil eines statistischen Kollektivs behandelt und die Berücksichtigung individueller Aspekte dadurch vernachlässigt wird.

Querschnittsthema 4 beschäftigt sich mit der Frage, welche Auswirkungen das Delegieren von Handlungen an Maschinen auf menschliche Kompetenzen und Fertigkeiten hat und wie der Gefahr von Kompetenzverlusten und *Deskilling* entgegengewirkt werden kann.

Die Querschnittsthemen 5 und 6 adressieren Chancen und Risiken im Umgang mit Daten. Während es auf der einen Seite darum geht, die Privatsphäre und Autonomie vor übermäßigen Eingriffen und Gefahren durch Überwachung zu schützen, geht es andererseits auch darum, Daten bestmöglich für sinnvolle und gemeinwohlorientierte Nutzung zugänglich zu machen. Hier diskutieren wir, wie Datenschutzrecht und -praxis gestaltet sein müssten, um beiden Zielen optimal Rechnung zu tragen.

In Querschnittsthema 7 wird die Bedeutung digitaler Technologien als kritische Infrastrukturen betont und die Frage gestellt, wie solche kritischen Infrastrukturen sicher und resilient gestaltet und Abhängigkeiten reduziert werden können.

Querschnittsthema 8 knüpft an diese Fragen an und beleuchtet Pfadabhängigkeiten in der Technologieentwicklung einerseits und Fragen von Missbrauch und Dual-Use andererseits.

Die letzten beiden Querschnittsthemen spannen den Bogen noch einmal zurück zu zwei Problemfeldern, die sich in nahezu allen in Kapitel 2 erwähnten ethischen Leitlinien zu Künstlicher Intelligenz wiederfinden. Dies betrifft einerseits Fragen rund um systematische Verzerrungen und Diskriminierung (Querschnittsthema 9) sowie andererseits Fragen zur Transparenz, Nachvollziehbarkeit, Kontrolle und Verantwortung im Kontext von KI-Systemen (Querschnittsthema 10). Hier gilt es, kontext-, sektoren- und adressatenspezifische Standards für Transparenz, Nachvollziehbarkeit und die Vermeidung von Verzerrungen zu entwickeln, deren Umsetzung durch verbindliche technische und organisationale Vorgaben gesichert wird.

+++ Vorabfassung! Wird durch die lektorierte Fassung ersetzt. +++

TEIL I: TECHNISCHE UND PHILOSOPHISCHE GRUNDLEGUNGEN

+++ Vorabfassung! Wird durch die lektorierte Fassung ersetzt. +++

2 Zentrale Entwicklungen und technische Grundlagen Künstlicher Intelligenz

2.1 Historischer Kontext

Die Erfindung von Maschinen und deren Rückwirkung auf die menschliche Lebenswelt hat sich stets vor dem Hintergrund spezifischer gesellschaftlicher Wahrnehmungen und Erwartungen vollzogen, die Rückwirkungen auf die jeweilige technologische Dynamik und die von ihr ausgehenden Entwicklungen von einfachen Geräten über komplexe Maschinen bis zu großtechnischen Anlagen gehabt haben. Dies gilt auch für die Entstehung immer leistungsfähigerer Softwaresysteme, die in vielfältigen intelligent und autonom anmutenden Formen mit Menschen interagieren können. Auch die dadurch ausgelösten philosophischen und ethischen Fragen und Herausforderungen, mit denen sich der Deutsche Ethikrat in dieser Stellungnahme befasst, sind in diesem historischen Kontext zu betrachten.

Die Idee von Maschinen, deren Fähigkeiten denen des Menschen ähneln oder diese sogar übertreffen, lässt sich Jahrtausende vor Erfindung der ersten Softwaresysteme zurückverfolgen. Schon in den Geschichten der antiken griechischen Mythologie erschufen Götter und Helden animierte Maschinen wie die mechanischen Dienerinnen des Schmiedegottes Hephaistos oder die animierten Statuen des Erfinders Daedalus.[1]

Mit der Erfindung der ersten Computer, also von Maschinen, deren besondere Stärke darin bestand, Programme mit komplexen rechnerischen und logischen Operationen schnell und effizient auszuführen, rückte die Existenz maschineller Intelligenz erstmals in greifbare Nähe.[2] Konzeptionell schon Mitte des 19. Jahrhunderts in den Entwürfen einer „analytischen Maschine“ von Charles Babbage und Ada Lovelace erdacht, wurden die ersten programmierbaren digitalen Rechenmaschinen erst hundert Jahre später gebaut, beginnend mit der von Konrad Zuse und Helmut Schreyer 1941 in Berlin gebauten „Zuse Z3“. Parallel dazu entwickelte der englische Mathematiker Alan Turing 1936 das formelle mathematische Modell eines universell programmierbaren Computers, später Turingmaschine genannt, das mithilfe von Algorithmen – knapp umreißbar als eindeutig definierte Rechenvorschriften (vgl. Abschnitt 2.2) – aus einer codierten Eingabe[3] eine bestimmte Ausgabe ermittelt.

---

[1] Mayor, A. (2018): Gods and Robots: Myths, Machines, and Ancient Dreams of Technology. Princeton.
[2] Nilsson, N. J. (2010): The Quest for Artificial Intelligence: A History of Ideas and Achievements. Ebook: https://ai.stanford.edu/~nilsson/QAI/qai.pdf, S.55-59.
[3] Z. B. binär als Folge von 0 und 1.

Turing erkannte früh, dass die Potenziale der neuen Rechenmaschinen auch die Frage aufwarfen, ob und wie es eines Tages möglich sein sollte, dass solche Geräte eine dem Menschen teilweise oder vollständig gleichwertige oder sogar überlegene Intelligenz aufweisen könnten. 1950 formulierte er ein Kriterium für KI das später als Turing-Test bezeichnet wurde. Demnach sei dann von maschineller Intelligenz auszugehen, wenn das Verhalten einer Maschine für menschliche Beobachter nicht von dem eines Menschen unterscheidbar erscheint.[4] Der Turing-Test gilt als wichtige, wenngleich keineswegs unumstrittene Inspiration für das Forschungsfeld der KI. Der Begriff der Künstlichen Intelligenz selbst wurde in Vorbereitung der Dartmouth-Konferenz geprägt, die im Sommer 1956 über zwei Monate am Dartmouth College in New Hampshire stattfand. Er diente als Sammelbegriff für alle Verhaltensweisen von Maschinen, die man als intelligent bezeichnen würde, wenn Menschen sie zeigen. Dies basiert auf der Prämisse, man könne „jeden Aspekt des Lernens oder jedes andere Merkmal von Intelligenz im Prinzip so genau beschreiben, dass eine Maschine dazu gebracht werden kann, sie zu simulieren".[5] Im Rahmen der Dartmouth-Konferenz und nachfolgender Konferenzen in den 1950er-Jahren begann man, sich mit vielen Themen zu beschäftigen, die bis heute in der Forschung eine große Rolle spielen, darunter Mustererkennung, Sprachverarbeitung, Abstraktionsfähigkeit, Kreativität, flexibles Problemlösen, zum Beispiel in Strategiespielen wie Schach, und die Fähigkeit zu lernen und sich weiterzuentwickeln.[6]

Die Pionierarbeiten zur KI führten gemeinsam mit rasanten Fortschritten bei der Entwicklung von Computerhardware und Programmiersprachen bald zu großem Optimismus. Viele Forschende gingen davon aus, dass maschinelle Intelligenz in vielen oder sogar allen Facetten innerhalb weniger Jahrzehnte mit menschlicher Intelligenz gleichziehen oder diese sogar übertreffen würde. Zusätzliche Inspiration kam von parallelen Fortschritten in der biologischen Forschung und insbesondere den Neurowissenschaften, deren Ergebnisse zunehmende Einblicke in kognitive Prozesse ermöglichten. Schon zu Beginn der 1970er-Jahre gab es jedoch auch erste Enttäuschungen, weil die in Aussicht gestellten praktischen Erfolge scheinbar ausblieben und Fördermittel daher gekürzt wurden.[7]

---

[4] Turing, A. M. (1950): Computing Machinery and Intelligence. In: Mind, LIX (236), 433–460 (DOI: 10.1093/mind/LIX.236.433).
[5] McCarthy, J. et al. (2006): A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. In: AI Magazine, 27 (4), 12-14, 12.
[6] Nilsson, N. J. (2010): The Quest for Artificial Intelligence: A History of Ideas and Achievements. Stanford. https://ai.stanford.edu/~nilsson/QAI/qai.pdf [20.12.2022]. , 73-88.
[7] Lighthill, J. et al. (1972): Artificial Intelligence: A General Survey. London. http://www.chilton-computing.org.uk/inf/literature/reports/lighthill_report/p001.htm [20.12.2022].

Diese erste Skepsis führte dazu, dass der Schwerpunkt der Aufmerksamkeit in den folgenden Jahren stärker auf praktische Anwendungen gerichtet wurde, die zudem dank zwischenzeitlich erfolgten Fortschritten in der theoretischen Grundlagenforschung und Computerentwicklung erstmals in greifbare Nähe rückten. Neben Weiterentwicklungen einzelner Kerngebiete entstanden Bemühungen, diese in anspruchsvolleren Projekten zusammenzuführen. Dazu gehörten etwa erste Versuche, mit parallelen Datenverarbeitungsmethoden Computer deutlich leistungsfähiger zu machen und sich selbstständig fortbewegende Roboter und Fahrzeuge zu entwickeln. Die zunehmende Vernetzung von Forschenden in Fachgesellschaften und von Computern in Vorläufern des Internets sowie das wachsende Engagement von Forschungsorganisationen, Militär und Industrie trugen dazu bei, dass das Interesse an KI in den 1970er- und 1980er-Jahren erneut aufblühte.[8] Auch hier kam es allerdings zu überzogenen Erwartungen und Voraussagen, denen so viele Enttäuschungen und Mittelkürzungen folgten, dass die Phase ab den späten 1980er-Jahren auch als „KI-Winter“ bezeichnet wird.[9] Der anfänglichen Euphorie über viele Entwicklungen folgten Enttäuschungen, wo diese an Grenzen stießen, beispielsweise bei der Leistungs-, Parallelisierungs- und Vernetzungsfähigkeit von Computern sowie der Flexibilität und Lern- bzw. Entwicklungsfähigkeit ihrer Programme.[10]

Parallel zu dieser technischen Entwicklung entstand ein kritischer Diskurs, in dessen Zuge sich Computerethik zunehmend als eigene Disziplin etablierte.[11] Aufbauend auf schon in den 1940ern angestellten Pionierüberlegungen von Norbert Wiener zu gesellschaftlichen Auswirkungen von Interaktionen zwischen Menschen und intelligent agierenden Maschinen[12], hatte in den 1970er-Jahren unter anderen Joseph Weizenbaum[13] detailliertere ethische Analysen zur Einbeziehung von Softwaresystemen und KI in menschliche Entscheidungs- und Lebensprozesse entwickelt. 1985 veröffentlichte Deborah Johnson das erste Lehrbuch zum Thema Computerethik[14] und James Moor betonte die breite Relevanz des Themas. Nach Moor brachte gerade die Flexibilität von Computern als universelle Werkzeuge besondere und weitreichende

[8] Nilsson, N. J. (2010): The Quest for Artificial Intelligence: A History of Ideas and Achievements. Stanford. https://ai.stanford.edu/~nilsson/QAI/qai.pdf [20.12.2022]. , 343-377.
[9] Nilsson, N. J. (2010): The Quest for Artificial Intelligence: A History of Ideas and Achievements. Stanford. https://ai.stanford.edu/~nilsson/QAI/qai.pdf [20.12.2022]. , 381-429.
[10] Schwartz, J. (1986): Limits of Artificial Intelligence. Article for ‘Encyclopedia of Artificial Intelligence’, Technical Report 212, 1-39. New York.
[11] Bynum, T. (2001): Computer and Information Ethics. In: The Stanford Encyclopedia of Philosophy. Summer 2018 Edition. https://plato.stanford.edu/archives/sum2018/entries/ethics-computer/[20.12.2022].
[12] Wiener, N. (1950): The Human Use of Human Beings: Cybernetics and Society. Boston, New York.
[13] Weizenbaum, J. (1976): Computer Power and Human Reason: From Judgment to Calculation. San Francisco.
[14] Johnson, D. G. (1985): Computer Ethics. Englewood Cliffs (NJ).

ethische Herausforderungen mit sich, da sie Menschen ständig neue Handlungsmöglichkeiten erschlossen, für die es oft noch keine geeigneten Regeln oder ethischen Standards gab.[15]

Mit dem Rüstzeug zur systematischen Prüfung von Chancen, Risiken und fundamentalen ethischen Aspekten von KI kamen zunehmend auch philosophische Zweifel auf, ob insbesondere die von einigen Forschenden in Aussicht gestellten Visionen einer generellen oder starken KI jemals realisiert werden könnten – oder sollten.[16] Auch die unkritische Übertragung bzw. Verwendung normativ bedeutsamer Begriffe wie „Intelligenz“ im Zusammenhang mit Softwaresystemen geriet in Kritik. Viele Forschende konzentrierten sich fortan bewusst auf enger umrissene Teilbereiche des Felds wie beispielsweise das maschinelle Lernen und versuchten, diese auch sprachlich als eigenständig wahrgenommene Felder zu etablieren, um das mit dem Begriff der Künstlichen Intelligenz verbundene Stigma der Aussichtslosigkeit oder moralischen Verwerflichkeit zu vermeiden.

Ab den 1990er-Jahren nahmen drei Entwicklungen Fahrt auf, die der Entwicklung von KI zu neuer Dynamik verhalfen:

Es gab erstens große Fortschritte beim Bau immer leistungsfähigerer, erschwinglicherer und handlicherer Computer, was zu einer massiven Ausweitung ihrer Verbreitung und Nutzbarkeit auf allen Ebenen führte (vgl. Abschnitt 2.2.1). Parallele Datenverarbeitung wurde nach der Jahrtausendwende selbst für viele Privatgeräte Standard und umfasst bei Großrechnern inzwischen bis zu mehrere tausend Prozessoren. Die voranschreitende Miniaturisierung in der Computertechnik und die Entwicklung entsprechend robuster und kleiner leistungsstarker Sensoren erlaubt ihren Einsatz in immer mehr Geräten, von denen viele, wie zum Beispiel Smartphones, Tablets, Smartwatches und weitere „Wearables", mittlerweile auch mobil nutzbar sind. Viele solcher Geräte können zudem über vielfältige Sensoren Details über Umwelt und die sie nutzenden Personen wahrnehmen, etwa über Kameras, Mikrofone, Bewegungsdetektoren, Ortsdatenempfänger, Thermometer oder Pulsmesser.

Zweitens hat die dynamische Entwicklung des Internets und diverser, auch drahtloser Zugänge die Vernetzungsmöglichkeiten zwischen Geräten revolutioniert, sodass gerade dort, wo viele Alltagsgeräte am Austausch von Daten beteiligt sind, mitunter von einem *Internet der Dinge* die Rede ist.[17] Im Zuge erweiterter Vernetzungsmöglichkeiten sind neue Infrastrukturen zur

---

[15] Moor, J. (1985): What Is Computer Ethics? In: Metaphilosophy, 16(4), 266–275.
[16] Searle, J. R. (1980): Minds, Brains, and Programs. In: Behavioral and Brain Sciences 3 (3), 417-457. (DOI: 10.1017/S0140525X00005756).
[17] Mattern, F.; Flörkemeier, C. (2010): Vom Internet der Computer zum Internet der Dinge. In: Informatikspektrum, 33, 107-121 (DOI: 10.1007/s00287-010-0417-7).

parallelen und dezentralen Datenverarbeitung entstanden, darunter das sogenannte Cloud-Computing, bei dem Daten aus vielen Quellen zentral verarbeitet und gespeichert werden, oder Edge-Computing, bei dem die Datenverarbeitung zu großen Teilen dezentral auf den Endgeräten erfolgt. Gemeinsam erlaubten diese Entwicklungen immer schnellere, vielfältigere und umfangreichere Verknüpfungen verschiedener, von unterschiedlichen Geräten erfassten, produzierten und (vor)verarbeiteten Daten, und zwar sowohl zwischen dezentralen Geräten als auch zwischen diesen Geräten und zentralen Hochleistungsrechnern.

Dies hat zum dritten eine beispiellose und noch immer ansteigende Datenflut hervorgebracht. Der auch unter dem Stichwort *Big Data* zusammengefasste Trend, große Mengen vielfältiger Daten mit hoher Geschwindigkeit[18] zu generieren, zu erfassen, immer wieder neu zu verknüpfen und zu analysieren, hat viele der im Folgenden vorgestellten jüngeren Entwicklungen von Algorithmen vorangetrieben oder überhaupt erst ermöglicht, insbesondere neuere statistische Methoden zur Mustererkennung im Bereich des maschinellen Lernens und hier vor allem das sogenannte Deep Learning.[19]

Gemeinsam haben diese drei noch andauernden Trends der Leistungssteigerung und Miniaturisierung von Computern, der Vernetzung digitaler Systeme und der damit verbundenen neuen Möglichkeiten der Datenzusammenführung und -auswertung eine Reihe von nachfolgend näher vorgestellten Entwicklungen angestoßen, deren gesellschaftliche und ethische Analyse im Mittelpunkt dieser Stellungnahme steht.

2.2 Künstliche Intelligenz im 21. Jahrhundert: Big Data, Algorithmen und soziotechnische Ökosysteme

2.2.1 Digitale Durchdringung der menschlichen Lebenswelt

Die im vorigen Abschnitt beschriebene Dynamik der Entwicklungen seit der Jahrtausendwende und insbesondere in den letzten Jahren hat zu einer intensivierten Durchdringung der Alltagswelt mit Computern geführt, die dazu beigeträgt, dass unter dem Schlagwort Künstliche Intelligenz inzwischen eine Fülle unterschiedlicher, aber vielfach konvergierender Phänomene diskutiert wird. Lag der Fokus angesichts der rasant wachsenden Datenmengen und der damit

[18] Im Englischen werden die Kernmerkmale von Big Data auch als „die drei V" – Volume (Menge), Variety (Vielfalt) und Velocity (Geschwindigkeit) – zusammengefasst. Laney, D. (2001): 3-D Data Management: Controlling Data Volume, Velocity and Variety.In: Application Delivery Strategies by META Group, 949.
[19] Deutscher Ethikrat (2017): Big Data und Gesundheit – Datensouveränität als informationelle Freiheitsgestaltung. Berlin. 54f.

verbundenen neuen Auswertungsmöglichkeiten zunächst noch auf Big Data, stehen inzwischen vielfach die Leistungen allgegenwärtiger algorithmischer Systeme im Mittelpunkt, die auf breiten Datengrundlagen scheinbar selbst Entscheidungen treffen und so zumindest den Eindruck erwecken, sie würden eigenständig und intelligent agieren.

Vernetzte Computertechnik in Alltagsbegleitern wie Mobiltelefonen, Uhren und Haushaltgeräten wird nicht nur mit dem Attribut *smart* beworben, sondern wirkt tatsächlich „klug“, wenn sie über Sensoren erfasste lokale Gegebenheiten mit online verfügbaren Daten verknüpft, um sich an individuelle Besonderheiten und Bedürfnisse der Menschen, die sie nutzen, anzupassen. Das gilt erst recht, wenn eine personalisierte oder gar emotional wirkende Ansprache hinzukommt, zum Beispiel über die Stimmen virtueller Assistenzsysteme wie Alexa oder Siri. Den Empfehlungssystemen zahlreicher Internet-Angebote wie zum Beispiel Video-Streaming-Diensten oder Online-Shops gelingen derweil mittels Analysen von Klicks und Nutzungsprofilen immer treffsichere Vorhersagen über die Vorlieben von Einzelpersonen und auch Roboter und Fahrzeuge können sich dank zunehmend leistungsfähiger Sensorik und Computertechnik immer besser unabhängig von menschlicher Steuerung fortbewegen.

Durch diese vielfältige Verbreitung algorithmischer Technik entstehen soziotechnische Datenökosysteme, in denen über die von Menschen verwendeten Geräte und die sie verknüpfenden Datennetzwerke zunehmend akkurate und umfangreiche digitale Repräsentationen der Bewegungen, Handlungen, Eigenschaften und Präferenzen vieler Personen entstehen. Solche reichhaltigen digitalen Abbilder können nicht nur ausgewertet werden, sondern wirken auch unmittelbar oder mittelbar auf die analogen Prozesse der menschlichen Lebenswelt und menschliches Verhalten zurück, indem auf ihrer Grundlage Menschen Informationen oder Handlungsempfehlungen angeboten werden.

Hinter diesen und vielen weiteren Entwicklungen steht ein Portfolio technischer Entwicklungen, die im Folgenden kurz vorgestellt werden – wohl wissend, dass es sich dabei angesichts der schnellen Veränderungen in der Technikentwicklung und der Einsatzpraxis nur um eine Momentaufnahme handeln kann. Auch wenn in den folgenden Ausführungen Daten, technische Infrastrukturen und Algorithmen konsekutiv beschrieben werden, so ist zu berücksichtigen, dass diese in komplexen Systemen miteinander verbunden sind und ihre Leistungen erst im Zusammenwirken erbringen.

+++ Vorabfassung! Wird durch die lektorierte Fassung ersetzt. +++

2.2.2 Daten und digitale Infrastrukturen

Das Fundament digitaler Operationen und Interaktionen bilden Daten, die von höchst unterschiedlicher Natur wie Qualität sein können. Es gibt unterschiedliche Datentypen (z. B. binäre, ordinale, metrische oder textliche Daten, Bilder, Musik und Videos) und Datenstrukturen (z. B. Listen, Tabellen), die aus vielfältigen Quellen (z. B. Sensoren, Nutzerstatistiken oder Umfragen) in diversen thematischen Zusammenhängen (z. B. Gesundheitsbereich, Finanzsystem, Verkehr, sozialen Netzwerken) stammen können. Daten können auch synthetisch generiert sein, wenn nicht genügend „echte" Daten in der gewünschten Qualität zur Verfügung stehen. Dies ermöglicht es etwa, ein System im Umgang mit Daten zu trainieren, die in vorhandenen Datensätzen aufgrund von Verzerrungen unterrepräsentiert sind, so beispielsweise Daten von Personen bestimmter demografischer oder ethnischer Zugehörigkeit.

Die Qualität eines Datensatzes hängt zum einen davon ab, wie genau, vollständig, aktuell oder detailliert die Daten sind. Zum anderen wird sie von den begleitenden Metadaten beeinflusst, die als Leseanleitung Auskunft über den Kontext und die Semantik (Bedeutung) der Daten geben. Metadaten informieren zum Beispiel über die Herkunft der Daten (Erhebungszeitpunkt und -ort, Art oder genaue Identifikation der Quelle oder des Sensors), die Bedeutung konkreter Zahlencodes oder sonstiger Etiketten im konkreten Datensatz, oder auch über die Zuordnung zu bestimmten Personen oder Kategorien. Die Kombination und Verknüpfung verschiedener Daten kann nur dann sinnvoll gelingen, wenn Kontext und Semantik für den jeweiligen Zweck hinreichend klar und kompatibel und die Daten auf dieser Grundlage somit interoperabel sind. Besonders gute Interoperabilität kann durch eine klare Standardisierung von Datenformaten und zu erfassenden Metadaten erreicht werden. Dies erhöht die Chancen, dass Daten beispielsweise im Gesundheitssystem institutionenübergreifend analysiert werden können, selbst wenn sie in verschiedenen Praxen, Kliniken oder Forschungseinrichtungen erhoben wurden.

Die Datenqualität hängt dabei nicht nur von den zuvor genannten Faktoren ab, sondern auch vom Verhältnis zwischen den Erhebungs- und den Anwendungskontexten. Selbst wenn jedes einzelne Datum „korrekt" gemäß der im vorigen Absatz genannten Anforderungen erhoben wurde, haben die Umstände und Zwecke der Datenerhebung wie auch der Datenauswertung Einfluss darauf, wie sehr bestimmte Daten den konkreten Qualitätsanforderungen für einen bestimmten Auswertungszusammenhang genügen (Validität). So werden beispielsweise Daten dafür genutzt, um auf kognitive oder emotionale Zustände von Individuen zu schließen die nicht direkt beobachtbar sind, wie etwa Aufmerksamkeit, Stress, kognitive Belastung oder Angst.

Die Auswahl und Bewertung der Aussagekraft der zu verwendenden Daten ist daher alles andere als trivial und muss sorgfältig geprüft werden. Es geht also darum, dass Daten nicht nur prinzipiell von hoher oder niedriger Qualität sind, sondern auch für eine jeweilige Frage oder Aufgabenstellung mehr oder weniger angemessen sein können. Werden solche Fragen der Passung nicht rechtzeitig und angemessen berücksichtigt, sind Verzerrungen oder irreführende Analysen möglich.[20]

Entscheidend für die Leistungsfähigkeit datengetriebener Anwendungen ist auch die Hardware und Infrastruktur, die für die Handhabung und Nutzung von Daten zur Verfügung steht. Das Herzstück bildet hierbei die Rechenleistung der Prozessorkerne, die mit den Daten arbeiten. Gerade für besonders rechen- und datenintensive Anwendungen kommen inzwischen hochparallelisierte Rechnersysteme zum Einsatz, in denen viele Hunderte oder Tausende Prozessorkerne gleichzeitig arbeiten. Inzwischen gibt es auch Prozessoren, die auf schnelles maschinelles Lernen auf Grundlage großer Datenmengen spezialisiert sind, wie die von Google entwickelten Tensor-Prozessoren (TPU = Tensor Processing Units). Sie sind für das hochparallele Addieren und Multiplizieren von Matrizen in neuronalen Netzen optimiert und werden beispielsweise in KI-Systemen wie dem digitalen Brettspielmeister AlphaGO oder in den Algorithmen von Google Photos, Google Maps, der Datenanalyse-Plattform Kaggle und der Google Cloud Platform verwendet.

Ein Großteil datenintensiver Operationen findet mittlerweile in Großeinrichtungen wie Datenlagern (Data Warehouses, Data Lakes) und Serverfarmen statt, die auf Datenspeicherung und/oder Datenanalyse spezialisiert und durch das Internet untereinander und mit Endgeräten vielfältig und dynamisch vernetzt sind. Der Austausch zwischen den vielen verschiedenen und oft weltweit verteilten Geräten in einem Datennetzwerk funktioniert mithilfe standardisierter Protokolle und Schnittstellen zur Anwendungsprogrammierung (Application Programming Interface, API). Sind Netzwerkknoten erst einmal über solche Schnittstellen verknüpft, fließen Daten entsprechend den vorgenommenen Einstellungen automatisch und oft in Echtzeit. Dies ermöglicht es, Rechen- und Speicherressourcen schnell und flexibel an die Anforderungen bestimmter Projekte und Kunden angepasst zur Verfügung zu stellen. Als Sammelbegriff für solche über das Internet zugänglichen Datenspeicher und Datenanalysedienste hat sich der Begriff *Cloud Computing* etabliert. Der Markt wird im privaten wie im institutionellen Bereich von Angeboten großer Internetfirmen wie Amazon, Microsoft und Google dominiert.

---

[20] Vgl. Barocas, S.; Selbst, A. D. (2016): Big Data's Disparate Impact. In: California Law Review 104 (3), 671-732. (DOI: 10.15779/Z38BG31), 671.

+++ Vorabfassung! Wird durch die lektorierte Fassung ersetzt. +++

Mit der Leistungssteigerung der Prozessoren in Endgeräten wie Heimcomputern, Mobiltelefonen und Smartwatches wachsen auch die Möglichkeiten, Daten zumindest teilweise bereits lokal in den Geräten, die sie erheben, zu verarbeiten. Solche erhebungsnahen, dezentralen Ansätze zur Datenverarbeitung an den „Rändern" des Internets werden in Abgrenzung zum Cloud Computing auch als *Edge Computing* bezeichnet. Sie bieten neben Ressourcenschonung beim Datentransfer auch datenschutzfreundlichere Gestaltungsmöglichkeiten, da zum Beispiel bestimmte sensible Daten gar nicht mehr weitergegeben werden müssen, sondern nur die von ihnen abgeleiteten Ergebnisse der Vorverarbeitung.[21]

2.2.3 Algorithmen und Datenverarbeitung

Mit der Menge und Vielfalt von Daten und der sie miteinander vernetzenden Infrastrukturen wachsen sowohl die Ansprüche an als auch die Möglichkeiten zur Datenverarbeitung. Herzstück jeglicher Datenverarbeitung sind Algorithmen: Verarbeitungsanweisungen zur Lösung eines Problems. Algorithmen geben vor, wie eingegebene Daten meist schrittweise nach klar definierten Regeln umgeformt werden, bis der gesuchte Ausgabewert erreicht ist. In ihrer einfachsten Form können Algorithmen Anleitungen zur klar festgelegten Verarbeitung von Daten oder Informationen sein, zum Beispiel die Formel zur Berechnung des Body-Mass-Indexes[22] oder eine Regel zum Sortieren von Zahlen nach ihrer Größe. Die meisten Algorithmen enthalten konditionale Elemente wie Wenn-dann-Anweisungen oder ineinander verschachtelte Befehlsschleifen, die das Ausführen komplexer Operationen unter Berücksichtigung der jeweiligen Situation ermöglichen. In Computerprogrammen sind Algorithmen in Programmiersprachen codiert.

Im Mittelpunkt der Arbeit mit großen und vielfältigen Datenmengen, die kennzeichnend für moderne soziotechnische Ökosysteme ist, stehen statistische Analysen, mit denen Regelmäßigkeiten in Daten erkannt sowie Zusammenhänge und potenzielle Wirkmechanismen zwischen einzelnen Merkmalen identifiziert werden, zum Beispiel Korrelationen zwischen der Körpergröße eines Kindes und der Größe der Eltern, den finanziellen Ressourcen der Familie, oder der Ernährung. Ausgehend von solchen Korrelationen können Vorhersagen für ähnliche Datensätze

---

[21] Vgl. z. B. hierzu die Ansätze des *Federated Learning* (Kaissis, G. et al. (2021): End-to-end privacy preserving deep learning on multi-institutional medical learning. In: Nature Machine Intelligence, 3, 473-484 (DOI: 10.1038/s42256-021-00337-8) und *Swarm Learning* (Warnat-Herresthal, S. et al. (2021): Swarm Learning for decentralized and confidential clinical machine learning. In: Nature, 594, 265-270 (DOI: 10.1038/s41586-021-03583-3).

[22] Der Body-Mass-Index oder Körpermasseindex wird als Richtwert zur Beurteilung des Verhältnisses zwischen Köpergröße und Körpergewicht verwendet. Er wird berechnet, indem man die Körpergröße (in Metern) durch das Quadrat des Körpergewichts (in Kilogramm) teilt.

oder künftige Entwicklungen abgeleitet werden. Dabei kommen unterschiedliche statistische Methoden zum Einsatz, die von einfachen Regressionsverfahren bis zu Deep Learning reichen.

Geht es darum, kausale Mechanismen nachzuweisen, sind in der Regel weitere Überlegungen und Untersuchungen nötig, die eine plausible Erklärung für den vermuteten Wirkzusammenhang zwischen Merkmalen anbieten, welche sich auch empirisch – beispielsweise in Experimenten – überprüfen lässt. Eine plausible Hypothese und ein empirischer Nachweis für einen kausalen Zusammenhang zwischen der Anzahl von Störchen und Geburten wird sich wohl nicht etablieren lassen, der Wirkmechanismus eines potenziellen Arzneimittels hingegen schon. Ein solcher Nachweis ist beispielsweisebeispielsweise im medizinischen Bereich besonders wichtig, da klinische Studien nur sinnvoll und sicher durchgeführt werden können, wenn biologisch plausible Hypothesen und zumindest vorläufig gesicherte Erkenntnisse zur Wirkweise neuer Therapeutika vorliegen. In anderen Fällen ist die Frage nach Kausalität möglicherweise weniger bedeutsam und es genügt, wenn auf Grundlage vorhandener Daten hinreichend zuverlässige Modelle berechnet werden können. Wenn das Empfehlungssystem eines Onlineshops beispielsweise treffsicher vorhersagt, dass Personen mit bestimmten Merkmalskombinationen unterschiedliche Produkte mögen, reicht das vielleicht für den Einsatzzweck schon aus und die Frage, *warum* solche Zusammenhänge bestehen, muss nicht beantwortet werden.

Statistische Analysen enthalten Unsicherheiten und geben in der Regel an, wie groß die jeweils zu erwartenden Fehler und Unsicherheiten sind, um Ergebnisse angemessen interpretieren zu können. Mit der Menge und Qualität der Daten wächst dabei für gewöhnlich die Verlässlichkeit der Analyseergebnisse, da die Wahrscheinlichkeit sinkt, dass die beobachteten Muster rein zufallsbedingt sind. Ganz ausmerzen lassen sich Fehler und Unsicherheiten in der Regel allerdings nicht und mit der Minimierung bestimmter Fehlerquellen können andere Fehlerquellen verstärkt werden. Optimiert man beispielsweise einen Test auf eine Virusinfektion dahingehend, dass er durch Berücksichtigung möglichst vieler Merkmale wirklich keine infizierte Person übersieht, steigt das Risiko, auch Personen fälschlich als infiziert zu identifizieren, bei denen vielleicht nur wenige dieser Merkmale in schwacher Ausprägung vorliegen – es entstehen *falsch-positive* Ergebnisse. Optimiert man den Test hingegen so, dass solche Fehler vermieden werden, indem nur wenige und/oder starke Zusammenhänge zwischen Merkmalen berücksichtigt werden, steigt das Risiko, dass tatsächlich infizierte Personen übersehen werden – es entstehen *falsch-negative* Ergebnisse. Welche Fehler in statistischen Analysen am ehesten in Kauf zu nehmen sind, hängt daher auch immer von der konkreten Fragestellung und Zwecksetzung ab und ist in zahlreichen Bereichen nicht nur eine technisch-methodische, sondern eine ethische Frage. Dies zeigt sich beispielsweise bei Entscheidungsunterstützungssoftware im Bereich der

Prognose von Kindeswohlgefährdung. Schon im Rahmen der Softwareentwicklung muss entschieden werden, welcher Fehler schwerer wiegt: eine Kindeswohlgefährdung übersehen oder ein Kind unbegründeterweise aus einer Familie genommen zu haben.

Maschinelles Lernen

Für algorithmische Verfahren und Systeme, die ihre Mustersuche, Modellbildung und sonstige Funktionsweise datenbasiert optimieren können, hat sich der Begriff des maschinellen Lernens etabliert.[23] Maschinelles Lernen umfasst unterschiedliche Ansätze, die in verschiedenen Anwendungsfeldern zum Einsatz kommen und ständig weiterentwickelt werden. Gemeinsam ist diesen Ansätzen eine anfängliche Trainingsphase, in der ein Algorithmus sein Modell zur Mustererkennung durch wiederholte Analyse von Trainingsdaten aufbaut und verfeinert. In der öffentlichen Debatte wird der Begriff des maschinellen Lernens häufig synonym mit KI verwendet. Allerdings beschreibt er lediglich eine Reihe statistischer Verfahren zur Analyse großer Datenmengen, die von gut interpretierbaren Ansätzen wie Entscheidungsbaum-Algorithmen über statistische Optimierungsmethoden wie Support-Vector-Machines bis hin zu künstlichen neuronalen Netzen reichen.[24] KI hingegen beschreibt – wie in Kapitel 2 bereits dargelegt – ein breites Forschungsfeld, in dem bereits seit den 1950er-Jahren ganz unterschiedliche Methoden eingesetzt werden, um menschliche Kognition maschinell zu simulieren. Der aktuelle „KI-Sommer“ geht wesentlich auf rasante Weiterentwicklungen beim maschinellen Lernen zurück. Waren klassische Ansätze noch stark auf passend vorverarbeitete Daten angewiesen, können moderne Methoden insbesondere des Deep Learning (siehe unten) flexibel auf verschiedenste Daten wie Bilder, Videos und Texte angewendet werden.[25]

Im Kontext des maschinellen Lernens sind folgende Typen von Lernverfahren zu unterscheiden: überwachtes Lernen, unüberwachtes Lernen und Verstärkungslernen. Beim *überwachten Lernen* sind die Zuordnungen zwischen den Eingabe- und den gesuchten Ausgabedaten im Trainingsdatensatz bereits bekannt. In einem Trainingsdatensatz zur Erkennung von Hautkrebs wären dies zum Beispiel Bilder von gesunder Haut und Hautkrebs, deren gesicherte Zuordnung zu

[23] Zu Details zu den in diesem Abschnitt kurz vorgestellten Ansätzen maschinellen Lernens vgl. Bauckhage, C. et al (2021): Tiefe neuronale Netze. In: Schmid, U.; Görz, G.; Braun, T. (Hg.): Handbuch der Künstlichen Intelligenz. 6. Auflage. Berlin, Boston, 89-101. DOI: 10.1515/9783110218091-002.
[24] Schmid, U. (2022): Vertrauenswürdige Künstliche Intelligenz: Nachvollziehbar, Transparent, Korrigierbar. In: Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz; Rostalski, F. (Hg.): Künstliche Intelligenz: Wie gelingt eine vertrauenswürdige Verwendung in Deutschland und in Europa? Tübingen, 287-298.
[25] Dazu gehören beispielsweise convoluted neural Networks, Long Short-Term Memory Verfahren, und generative Ansätze, vgl. Kapitel 12 aus Schmid, U.; Görz, G.; Braun, T. (Hg.) (2021): Handbuch der Künstlichen Intelligenz. 6. Auflage. Berlin, Boston.

einer dieser beiden Kategorien in einem Etikett (Label) vermerkt ist. Der trainierende Algorithmus kann seinen Zuordnungserfolg anhand dieser Etiketten in jeder Trainingsrunde überprüfen und seine Rechenregeln anpassen, bis das System für die korrekte Erkennung von Hautkrebs optimiert ist. Überwachtes Lernen eignet sich besonders für Algorithmen, die bestimmte Muster in Daten zuverlässig erkennen und kategorisieren können sollen, also zum Beispiel für Diagnostikwerkzeuge, die wie im obigen Beispiel in der Bilderkennung eingesetzt werden, aber auch für Ansätze zur Sprach-, Text- oder Objekterkennung. Auch das Training von Algorithmen zur Vorhersage künftiger Entwicklungen, beispielsweise beim Wetter, im Finanzsektor oder beim Wasser- oder Stromverbrauch, geschieht häufig mit überwachtem Lernen, da Algorithmen hier Regressionsmodelle auf Grundlage bekannter und entsprechend etikettierter Zusammenhänge zwischen Merkmalen in Trainingsdaten aus der Vergangenheit erlernen.

Unüberwachtes Lernen hingegen funktioniert ohne vorherige Etikettierung der Trainingsdaten; stattdessen „sucht" der Algorithmus eigenständig nach Mustern. Auf diese Weise können neue Strukturen und Gruppierungen (Cluster) in Daten entdeckt und weiterverwendet werden, zum Beispiel zur Generierung neuer „kreativer" Inhalte durch den Algorithmus oder zur Analyse von Marktsegmenten oder Zielgruppen, um passgenaue Angebote zu liefern. Durch den Verzicht auf eine vorherige Etikettierung lassen sich beim unüberwachten Lernen zudem nicht nur die im Vorfeld notwendigen Vorverarbeitungsschritte, sondern auch bestimmte Verzerrungen (Bias) reduzieren , so zum Beispiel jene, die bei einer händischen Zuweisung von Merkmalen aufgrund von Vorurteilen oder bloßen Intuitionen seitens des Entwicklungsteams einfließen können. Andererseits sind unüberwachte Lernalgorithmen anfällig für jene Verzerrungen, die bereits in den Trainingsdaten enthalten sind. Ein Algorithmus, der Vorhersagen zur Medikamentenverträglichkeit mit Daten trainiert, die überwiegend von Männern stammen, ist so möglicherweise weniger präzise, wenn es um Vorhersagen für Frauen geht, da relevante Faktoren für diese Gruppe beim Training nicht berücksichtigt wurden. Außerdem steigt in dem Maße, in dem ein Algorithmus unabhängig von menschengemachten Etiketten und sonstigen Vorgaben seinen eigenen Weg zur Problemlösung „sucht"[26], auch die potenzielle Undurchdringbarkeit seiner Lösungsansätze.

Beim *Verstärkungslernen* optimiert der Algorithmus seine Operationen auf bestimmte Ziele hin und erhält dabei in der Trainingsphase für jeden Versuch eine Rückmeldung, ob dieser Schritt

[26] Wir verwenden gelegentlich mentale Prädikate, um Eigenschaften von Software zu beschreiben. Das entspricht einer etablierten Redeweise, die nicht dahingehend missverstanden werden darf, dass damit angenommen wird, Softwaresysteme hätten tatsächlich diese mentalen Eigenschaften, würden zum Beispiel etwas suchen.

das System dem Ziel nähergebracht oder es davon entfernt hat. Schritte in die richtige Richtung werden mit einer Bonusfunktion belohnt, solche in die falsche Richtung bestraft. Die Methoden sind in der Regel so aufgebaut, dass sich kurzfristig auch Schritte in die falsche Richtung lohnen können, wenn sie einer langfristig günstigen Strategie dienen. Verstärkungslernen lässt sich beispielsweise einsetzen, um Spielstrategien zu entwickeln, Verkehrsflüsse zu optimieren oder Klicks und Verweildauer auf einer Plattform zu maximieren.

Die algorithmischen Strategien, die im Laufe des Trainings zur Bewältigung der jeweiligen Aufgaben entwickelt werden, sind in der Regel selbst für geschultes Personal, das den Code vollständig einsehen kann, nicht auf Anhieb nachvollziehbar (Blackbox). Das gilt umso mehr, je dynamischer sich ein Algorithmus im Laufe seiner Arbeit weiterentwickelt, also insbesondere für Deep Learning. Es gibt verschiedene Lösungsansätze, um trotzdem eine für die jeweilige Zielgruppe angemessene Transparenz, Interpretierbarkeit oder Erklärbarkeit algorithmischer Prozesse zu erreichen (*explainable AI*); deren Auswahl und Anwendung ist jedoch technisch anspruchsvoll und stellt ein hochdynamisches Forschungsfeld dar.[27]

Der Begriff des maschinellen Lernens[28] kommt nicht von ungefähr, dienen menschliche Lernprozesse doch als Analogie – ähnlich wie beim Begriff der KI auch. So lernen auch Menschen „überwacht“, wenn sie beispielsweise unterrichtet werden oder sich von anderen etwas abschauen. Unüberwachtes maschinelles Lernen weist Parallelen zu Spiel und kreativen oder erkundenden Prozessen auf. Verstärkungslernen findet biologische Vorbilder in den gut erforschten Belohnungs- und Aversionssystemen von Gehirnen[29], die unmittelbar auf Lernprozesse zurückwirken, indem sie über die Ausschüttung diverser Neurotransmitter Verbindungen zwischen Nervenzellen stärken oder schwächen.

Biologisch inspiriert ist auch ein Teilbereich des maschinellen Lernens, der besonders für den Umgang mit großen Datenmengen geeignet ist und in den letzten Jahren zu einem wichtigen

---

[27] Samek, W. et al. (2021): Explaining Deep Neural Networks and Beyond: A Review of Methods and Applications. In: Proceedings of the IEEE 109 (3), 247-278 (DOI: 10.1109/JPROC.2021.3060483); Miller, T. (2019): Explanation in Artificial Intelligence: Insights from the Social Sciences. In: Artificial Intelligence, 267, 1-38 (DOI: 10.1016/j.artint.2018.07.007); und siehe https://facctconference.org/ [27.02.2023].

[28] Geprägt wurde der Begriff *Machine Learning* von KI-Forschenden, die die ersten Ansätze hierzu etabliert haben, siehe z. B. McCarthy, J.; Feigenbaum, E. A. (1990): In Memoriam. Arthur Samuel: Pioneer in Machine Learning. In: AI Magazine, 11 (3), 10-11 (DOI: 10.1609/aimag.v11i3.840); Carbonell, J. G.; Michalski, R. S.; Mitchell, T. M. (1983): Machine Learning: A Historical and Methodological Analysis. In: AI Magazine, 4 (3), 69-78 (DOI: 10.1609/aimag.v4i3.406). Parallel hat sich im Bereich Signalverarbeitung Mustererkennung (*Pattern Recognition*) als verwandtes Forschungsgebiet etabliert, siehe Shaw, S. W. (1990): Review of Pattern Recognizition. In: AI Magazine, 11 (2), 80-81 (DOI: 10.1609/aimag.v11i2.838). Beides wird heute als maschinelles Lernen zusammengefasst.

[29] Neftci, E. O.; Averbeck, B. B. (2019): Reinforcement learning in artificial and biological systems. In: Nature Machine Intelligence 1, 133–143 (DOI: 10.1038/s42256-019-0025-4).

Treiber für viele KI-Anwendungen geworden ist – Deep Learning. Hier kommen sogenannte neuronale Netze zum Einsatz, deren Funktionsweise entfernt an Netzwerkstrukturen im Gehirn angelehnt ist.

Infokasten 1: Neuronale Netze im Gehirn

Im Gehirn ermöglicht das Zusammenspiel von Nervenzellen in komplexen und hierarchischen Vernetzungen vielfältige und anpassungsfähige Funktionen. Eine einzelne Nervenzelle kann über Synapsen mit mehreren tausend anderen Nervenzellen verschaltet sein, und zwar sowohl in ihrem (postsynaptischen) Eingabebereich, in dem sie Signale empfängt, als auch in ihrem (präsynaptischen) Ausgabebereich, über den sie Informationen an „flussabwärts" liegende Zellen weitergibt. Bei der Verarbeitung der vielfältigen empfangenen Signale leistet schon die einzelne Zelle erstaunliche Integrationsarbeit, die zum Beispiel die Qualität und die raumzeitlichen Muster der eingehenden Informationen berücksichtigt. Wesentlicher Bestandteil von Lernprozessen sind plastische Veränderungen in der Funktion und Struktur von Synapsen, die in Reaktion auf neuronale Aktivität erfolgen. Eine große Rolle spielt dabei die Modulation des Verhältnisses von synaptischer Erregung und Hemmung innerhalb der neuronalen Netzwerke. Die Stärke der Synapsen wird durch die Anzahl verschiedener Rezeptoren bestimmt, die auf unterschiedliche chemische Botenstoffe (Neurotransmitter) reagieren. Lernergebnisse und neue Gedächtnisinhalte schlagen sich in der Stärkung einzelner Synapsen nieder, was wiederum Rückwirkungen auf die Funktion der beteiligten Netzwerke und sogar auf das Überleben einzelner Nervenzellen haben kann.

Die Bausteine eines maschinellen neuronalen Netzes sind künstliche „Nervenzellen" – Recheneinheiten, die auch als Perzeptrone bezeichnet werden. Jedes Perzeptron ist mit anpassbaren Gewichtungen und einem Schwellenwert ausgestattet, die mitbestimmen, wie das Perzeptron auf eingehende Signale reagiert und welchen Ausgabewert es produziert.

Die Perzeptrone sind miteinander vernetzt, wobei die Vernetzung meist als vollständige Verbindung aller Neuronen einer Schicht mit allen der nächsten Schicht realisiert wird. Jedes Neuronale Netz besteht aus einer Eingabeschicht, einer Ausgabeschicht und dazwischenliegenden „versteckten" Schichten, in denen die Datenverarbeitung stattfindet. Es wurden verschiedene Arten von Architekturen neuronaler Netze entwickelt, die für die Bearbeitung bestimmter Arten von Daten besonders gut geeignet sind. Eine der einflussreichsten Architekturen im Bereich Deep Learning sind Convolutional Neural Networks. Während klassische neuronale Netze wie auch die meisten anderen Ansätze des maschinellen Lernens Merkmale als Eingabe erwarten, könnten Convolutional Neural Networks Rohdaten wie Bilder direkt verarbeiten. Man spricht hier von Ende-zu-Ende-Lernen. Dies wird ermöglicht, indem spezielle, mit Filtern ausgestattete Schichten eingeführt werden, mittels derer eine bestimmte Information aus den Bildern herausgefiltert werden kann. Mathematisch entspricht dies der Operation der Faltung (*convolution*). Dabei werden nach und nach bessere Datenrepräsentationen aufgebaut. In der Bilderkennung

etwa entwickelt das Netzwerk zunächst Kontrastfilter, die helle und dunkle Bereiche voneinander unterscheiden, dann Filter, die beispielsweise Konturen oder Ecken erkennen, und schließlich Filter für komplexere visuelle Komponenten (ein Gebäude oder ein Rad) und die kompositionelle Repräsentation ganzer Gegenstände oder Szenen. Für verschiedene Arten von Daten (Text, Bild, Messreihen) und Problemen (Klassifikation, Segmentierung, Generierung) gibt es verschiedene Netzwerkarchitekturen, die sich insbesondere in den letzten Jahren entwickelt oder weiterentwickelt haben.

Im Unterschied zum menschlichen Gehirn sind künstliche Neuronale Netzwerke jedoch bislang trotz ihrer enormen Komplexität und ihrer beachtlichen Datenverarbeitungskapazitäten auf vergleichsweise enge Aufgabenbereiche ausgerichtet und auch nach wie vor deutlich schlichter. Der 2018 in Betrieb genommene Supercomputer SpiNNaker[30] hat über eine Million Rechenkerne, von denen jeder bis zu 256 Nervenzellen nachstellen soll. Im menschlichen Gehirn sind hingegen ca. 86 Milliarden Nervenzellen vernetzt.

Leistungsstarke, auf die Anforderungen neuronaler Netze spezialisierte Hardware zusammen mit neuen Methoden und Architekturen erlauben es, besser mit großen Datenmengen und mit unstrukturierten Daten umzugehen, da sie die damit einhergehenden Komplexitäten aufgrund ihrer modularen und hierarchischen Struktur besonders gut handhaben und abbilden können. Der aktuelle Boom bei vielen KI-Entwicklungen geht entscheidend auf die durch Deep Learning-Algorithmen eröffneten Möglichkeiten zurück und diese wiederum auf die enorm gestiegenen Datenmengen und Chipleistungen.

2.2.4 Einsatzbereiche algorithmischer Systeme und Künstlicher Intelligenz

Die oben beschriebenen kombinierten Entwicklungen von Hardware und Software, Vernetzung und Datenproduktion haben vielfältig einsetzbare Anwendungsmöglichkeiten von algorithmischen Systemen hervorgebracht, die auch immer wieder auf großes öffentliches Interesse treffen.

So gelingt es Computern inzwischen, Menschen in anspruchsvollen Strategiespielen wie Schach und Go zu schlagen. Während der Schachcomputer Deep Blue den damaligen Schachweltmeister Garri Kasparow 1997 noch mithilfe eines Algorithmus besiegte, der im Wesentlichen mittels schneller Suche in einer riesigen Datenbank möglicher Spielzüge triumphierte, meistert der 2019 entwickelte Algorithmus MuZero inzwischen Spiele wie Go, Schach und

[30] Furber, S.; Bogdan, P. (2020): SpiNNaker – A Spiking Neural Network Architecture. Hanover (MA) (DOI: 0.1561/9781680836530.ch4), 46.

Shogi ganz ohne die Berücksichtigung historischer Partien oder Kenntnis der Spielregeln. Dieser Algorithmus lernt vielmehr durch Versuch und Irrtum, indem er gegen sich selbst spielt und ausprobiert, was dabei am besten funktioniert.[31]

Im Bereich der Sprachverarbeitung erregen derweil regelmäßig Leistungssprünge in der Textproduktion Aufsehen. Das Neuronale Netz des Generative Pre-trained Transformer 3 (GPT-3) der amerikanischen Organisation OpenAI etwa wurde auf Basis einer ausnehmend umfassenden Menge an Texten – darunter die komplette englischsprachige Wikipedia – trainiert, bis es ab 2020 selbst Texte produzieren konnte, deren maschineller Ursprung oftmals nicht mehr zu erkennen war.[32] Das auf GPT-3 aufbauende, im November 2022 veröffentlichte Dialogsystem ChatGPT[33] kann auf unterschiedlichste Fragen mitunter so überzeugend und differenziert reagieren, dass sich selbst Antworten auf komplexe Aufgaben, wie die Erstellung wissenschaftlicher Hausarbeiten, nicht von qualitativ hochwertigen menschlich verfassten Eingaben unterscheiden lassen.[34]

Der Deutsche Ethikrat nimmt in dieser Stellungnahme vier Handlungsfelder in den Blick, in denen der Einsatz algorithmischer Systeme und Künstlicher Intelligenz entweder schon besonders weitreichende Veränderungen mit sich gebracht hat oder dies in näherer Zukunft bewirken könnte. Im *Gesundheitsbereich* (vgl. Kapitel 5) beispielsweise stellt die Auswertung großer Datenmengen durch maschinelles Lernen Fortschritte bei Diagnostik und individualisierten Präventions- und Therapieempfehlungen in Aussicht[35] und kommt KI-gestützte Software auch in Robotern zunehmend zur Anwendung, etwa in der Pflege[36], bei Operationen oder in der Psychotherapie. Im Bereich der *Bildung* (vgl. Kapitel 6) gibt es nicht erst seit der Corona-Pandemie vielfältige Ansätze, die Vermittlung von Wissen und Kompetenzen in der Schule mithilfe datenbasierter, KI-gestützter Lehr-Lern-Systeme und KI-gestütztem Unterrichtsmanagement effektiver zu gestalten und besser auf individuelle Belange von Lernenden einzugehen. Beson-

---

[31] Schrittwieser, J. et al. (2020): Mastering Atari, Go, chess and shogi by planning with a learned model. In: Nature 588, 604–660 (DOI: 10.1038/s41586-020-03051-4).
[32] Brown, T. B. et al. (2020): Language Models are Few-Shot Learners. Larochelle, H. (Hg.): NIPS'20: Proceedings of the 34th International Conference on Neural Information Processing Systems. Red Hook (NY).
[33] Schulman, J. et al. (2022): ChatGPT: Optimizing Language Models for Dialogue. In: OpenAI. https://openai.com/blog/chatgpt/ [10.02.2023].
[34] Pretschner, A. et al. (2023): Die mächtigen neuen Assistenzsysteme. Was aus der Künstlichen Intelligenz ChatGPT folgt, über die gerade alle sprechen. In: Frankfurter Allgemeine Zeitung. https://www.faz.net/aktuell/wirtschaft/digitec/chatgpt-und-ki-die-maechtigen-neuen-assistenzsysteme-18587321.html [18.01.2023].
[35] Deutscher Ethikrat (2017): Big Data und Gesundheit – Datensouveränität als informationelle Freiheitsgestaltung. Berlin.
[36] Deutscher Ethikrat (2020): Robotik für gute Pflege. Berlin.

ders weit in den Alltag vieler Menschen sind derweil Entwicklungen im Bereich der *öffentlichen Kommunikation und Meinungsbildung* (vgl. Kapitel 7) vorgedrungen, wo ein Großteil des Informationsaustauschs inzwischen über algorithmisch gestützte digitale Plattformen und Soziale Medien abläuft. Dabei können durch Verknüpfung vielfältiger Daten zunehmend präzise Profile von Einzelpersonen erstellt und zur Entwicklung maßgeschneiderter kommerziell, aber auch politisch motivierter Ansprachen und Angebote verwendet werden. Auch in der *öffentlichen Verwaltung* (vgl. Kapitel 8) kann der Einsatz algorithmischer Systeme das Leben vieler Menschen berühren, beispielsweise bei der Beurteilung oder Überwachung von Personen im Bereich des Sozial- oder Polizeiwesens.

2.2.5 Ethische Leitlinien und regulativer Rahmen für algorithmische Systeme und KI

Die vorstehend beschriebenen Entwicklungen von Datenanalysemethoden und Softwaresystemen sowie die damit verbundenen Veränderungen in vielen Bereichen menschlichen Handelns bringen auch Herausforderungen für die Ausgestaltung von Regeln für das menschliche Miteinander mit sich. In diesem Zusammenhang ist bereits eine Reihe von Regularien entstanden oder aktuell in der Entwicklung, die hier nur kurz benannt werden.

Ethische Leitlinien

Die potenziell weitreichenden Rückwirkungen datengestützter algorithmischer Anwendungen auf viele Bereiche der menschlichen Lebenswelt haben in der Industrie wie auch im akademischen und zivilgesellschaftlichen Diskurs Überlegungen ausgelöst, wie ein ethisch vertretbarer und gesellschaftlich verträglicher Einsatz solcher neuen Technologien gestaltet werden kann. Dieser Austausch hat eine Fülle an Leitlinien hervorgebracht – 84 allein nach einer Übersicht aus dem Jahr 2019.[37] Das Angebot reicht von Codizes einzelner Unternehmen wie beispielsweise der Deutschen Telekom[38], von SAP[39], Microsoft[40] oder Google[41] über Richtlinien von

[37] Jobin, A.; Ienca, M.; Vayena, E. (2019): The global landscape of AI ethics guidelines. In: Nature Machine Intelligence 1 (9), 389-399 (DOI: 10.1038/s42256-019-0088-2).
[38] Deutsche Telekom AG (2018): KI-Leitlinien. https://www.telekom.com/resource/blob/544508/ca70d6697d35ba60fbcb29aeef4529e8/dl-181008-digitale-ethik-data.pdf [18.01.2023].
[39] SAP (2021): SAP's Guiding Principles for Artificial Intelligence. https://www.sap.com/docs/download/2018/09/940c6047-1c7d-0010-87a3-c30de2ffd8ff.pdf [18.01.2023].
[40] https://www.microsoft.com/en-us/ai/responsible-ai; Microsoft AI (2020): Putting principles into practice: How we approach responsible AI at Microsoft. https://www.microsoft.com/cms/api/am/binary/RE4pKH5 [01.02.2023] und Amershi, S. et al (2019): Guidelines for Human-AI Interaction. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Paper No. 3, 1–13 (DOI: 10.1145/3290605.3300233).
[41] Google AI (2018): Artificial Intelligence at Google: Our Principles. https://ai.google/principles [18.01.2023].

Fachgesellschaften wie dem Institute of Electrical and Electronics Engineers[42] bis hin zu Werken auf nationaler oder internationaler Ebene. In Deutschland sind hier insbesondere die Stellungnahmen der Datenethikkommission[43] und der Enquete-Kommission Künstliche Intelligenz[44] zu nennen, auf internationaler Ebene die „Ethik-Leitlinien für eine Vertrauenswürdige KI“ der von der Europäischen Kommission eingesetzten Hochrangigen Expertengruppe für KI[45] und die „Recommendation on the Ethics of Artificial Intelligence“ der UNESCO[46]. Es gibt inzwischen mehrere Ansätze, solche Richtlinien zu kartieren und aus ihren überlappenden Inhalten sowohl Gemeinsamkeiten als auch Unterschiede herauszuarbeiten, die sich aus spezifischen Perspektiven und Interessen der beteiligten Personen und Institutionen ergeben.[47]

Dabei wurden eine Reihe von Themen identifiziert, die weitgehend anwendungsbereichsübergreifend bedeutsam erscheinen und daher in den meisten ethischen Leitlinien zu algorithmischen Systemen und KI eine Rolle spielen. Zu diesen gehören Bedenken hinsichtlich des Schutzes der Privatsphäre ebenso wie Fragen, die sich im Zusammenhang mit der Entwicklung und Funktion von Algorithmen ergeben, beispielsweise zu Voreingenommenheiten und Verzerrungen oder zu Transparenz und Verantwortung angesichts weitreichender Undurchschaubarkeiten von Ansätzen maschinellen Lernens. Versuche, die in diesen Kontexten aufgestellten normativen Forderungen übergeordneten ethischen Prinzipien zuzuordnen, ergeben teilweise eine enge

---

[42] Institute of Electrical and Electronics Engineers (2017): Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems. Version 2 – For Public Discussion. https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf [18.01.2023].

[43] Datenethikkommission (2019): Gutachten der Datenethikkommission. https://www.bmi.bund.de/SharedDocs/downloads/DE/publikationen/themen/it-digitalpolitik/gutachten-datenethikkommission.pdf;jsessionid=E5398475D4C9DA18058BF1CB957DB2D2.2_cid364?__blob=publicationFile&v=7 [18.01.2023].

[44] Deutscher Bundestag (2020): Bericht der Enquete-Kommission Künstliche Intelligenz – Gesellschaftliche Verantwortung und wirtschaftliche, soziale und ökologische Potenziale. Bundestagsdrucksache 19/23700. Berlin. https://dserver.bundestag.de/btd/19/237/1923700.pdf [18.01.2023].

[45] Hochrangige Expertengruppe für Künstliche Intelligenz (HEG-KI) (2019): Ethik-Leitlinien für eine Vertrauenswürdige KI. https://op.europa.eu/de/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1 [18.01.2023].

[46] UNESCO (2021): Recommendation on the Ethics of Artificial Intelligence. https://unesdoc.unesco.org/ark:/48223/pf0000380455.locale=en [18.01.2023].

[47] Fjeld, J. et al. (2020): Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI. Berkman Klein Center Research Publication No. 2020-1. (DOI: 10.2139/ssrn.3518482); Floridi, L. et al. (2018): AI4People – An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. In: Minds and Machines 28 (4), 689-707. (DOI: 10.1007/s11023-018-9482-5); Hagendorff, T. (2020): The Ethics of AI Ethics: An Evaluation of Guidelines. In: Minds and Machines 30 (1), 99-120. (DOI: 10.1007/s11023-020-09517-8); Jobin, A.; Ienca, M.; Vayena, E. (2019): The global landscape of AI ethics guidelines. In: Nature Machine Intelligence 1 (9), 389-399 (DOI:10.1038/s42256-019-0088-2); Rudschies, C.; Schneider, I.; Simon, J. (2021): Value Pluralism in the AI Ethics Debate – Different Actors, Different Priorities. In: The International Review of Information Ethics 29, 1-15 (DOI: 10.29173/irie419).

Anlehnung an die etablierten vier Prinzipien der Medizinethik[48] – Autonomie, Schadensvermeidung, Wohltätigkeit, Gerechtigkeit – ergänzen diese aber häufig um weitere Prinzipien wie Erklärbarkeit und menschliche Kontrolle, die sich auf spezifischere Aspekte von Digitalisierung und KI beziehen.

Regulativer Rahmen

Für das Verständnis des regulativen Rahmens ist hilfreich zu sehen, dass Regeln neue Entwicklungen – auch Bereich des Zusammenwirkens von Mensch und Maschine – auf unterschiedliche Weise erfassen können. Abstrakte, generelle Regeln können auf ein neues Phänomen angewandt werden, ohne dass es nötig ist, sie in ihrem Wortlaut zu verändern oder neue Regeln für das Phänomen zu schaffen. Bei gesetzlichen Regelungen ist es Aufgabe von Rechtswissenschaft und Rechtspraxis, die Regeln entsprechend zu interpretieren.

Vor diesem Hintergrund sind für die hier beschriebenen sozio-technischen Entwicklungen beispielsweise Rechtsnormen relevant, die gar keinen expliziten Technikbezug haben. So kann sich etwa aus den allgemeinen Regeln des Kartellrechts ergeben, ob es eine verbotene Preisabsprache darstellt, wenn von zwei Unternehmen eingesetzte Preisfindungsalgorithmen ohne menschliche Steuerung so miteinander interagieren, dass die Preise nicht mehr unabhängig festgelegt werden.[49] Aus generellen Vorschriften des Arbeitsrechts können sich Grenzen dafür ergeben, inwieweit der Arbeitgeber die Leistung von Mitarbeitenden mithilfe technischer Systeme erfassen darf.[50] Technikneutrale Arbeitsschutzbestimmungen sind auch auf den Einsatz von Robotern in der Produktion anwendbar.[51] Viele weitere Beispiel ließen sich hier anführen.[52]

---

[48] Beauchamp, T. L.; Childress, J. F. (2001): Principles of Biomedical Ethics. 5. Auflage. New York.
[49] Truby, J.; Brown, R. (2020): Human digital thought clones: the *Holy Grail* of artificial intelligence for big data. In: Information & Communications Technology Law 2021, 30 (2), 140-168 (DOI: 10.1080/13600834.2020.1850174).
[50] Holthausen, J. (2021): Big Data, People Analytics, KI und Gestaltung von Betriebsvereinbarungen – Grund-, arbeits- und datenschutzrechtliche An- und Herausforderungen. In: Recht der Arbeit 1, 19-32, 19; Waas, B. (2022): KI und Arbeitsrecht. In: Recht der Arbeit 3, 125-130, 125.
[51] Kollmer, N.; Klindt, T.; Schucht, C. (2021): Arbeitsschutzgesetz. 4. Auflage. München, Überblick vor § 1, Rn. 90 ff.; Günther, J.; Böglmüller, M. (2022): Arbeits- und Gesundheitsschutz/Haftung im Arbeitsverhältnis. In: Arnold, C.; Günther, J. (Hg.): Arbeitsrecht 4.0. 2. Auflage 2022. München, 183-224, § 4: Rn. 125-129.
[52] Im Überblick: Schulz, W.; Schmees, J. (2022): Möglichkeiten und Grenzen der Künstlichen Intelligenz in der Rechtsanwendung. In: Augsberg, I.; Schuppert, G. F. (Hg.): Wissen und Recht, Interdisziplinäre Studien zur Wissensgesellschaft. Baden-Baden, 561-594, 561 ff.

Zu diesen technikunspezifischen, aber hoch relevanten Normen gehören Haftungsregelungen. Sie prägen die Entwicklungen im Bereich Mensch-Maschine entscheidend.[53] Haftungsregeln legen fest, wer im Fall eines Schadens dafür (finanziell) einzustehen hat, unter welchen Voraussetzungen und in welchem Umfang. Sie finden sich als bereichsspezifische und bereichsübergreifende Haftungsregelungen. Auch wenn sie nur regeln, wer im Einzelfall nachträglich für Schäden aufkommen muss, wirken sie auf die Herstellung und auf die Anwendung technischer Systeme zurück. Wenn eine Person weiß, dass sie für die Schäden haftet, die das System anrichtet, so hat sie beispielsweise einen Anreiz, das System intensiver zu überwachen, seinen Aktionsradius einzuschränken oder eine menschliche Letztentscheidung vorzusehen. Knüpft die Haftung aber gerade an die menschliche Entscheidung an, kann der Anreiz andersherum auch gerade dahingehen, die Maschine selbstständig „laufen zu lassen". Es kommt also auf die Ausgestaltung der Haftungsregelungen an.

Eine – bislang theoretische – Diskussion zur Haftung betrifft die Frage, ob nicht eine Situation eintreten kann, in der Maschinen so unabhängig von Menschen agieren, dass es unangemessen erscheint, den Menschen, der die Maschine entwickelt oder sie eingesetzt hat, für die Schäden, die die Maschine anrichtet, haften zu lassen. Wenn die Maschine selbst nicht auf Schadenersatz in Anspruch genommen werden kann, entsteht eine Verantwortungs- und Haftungslücke (*responsibility gap*) – Geschädigte gingen also leer aus. Eine rechtliche Antwort darauf kann sein, den Maschinen selbst – oder durch eine Versicherung – Vermögen zuzuordnen, auf das dann im Schadensfall zurückgegriffen werden kann.[54] Dies verweist auf eine noch grundsätzliche rechtliche Diskussion, nämlich ob es irgendwann sinnvoll erscheint, technischen Systemen eine eigene Rechtspersönlichkeit zuzuweisen, wie man es seit Langem bei Organisationen tut (sogenannte *juristische Personen*).[55]

Neben oder anstelle der Anwendung abstrakter, genereller Regeln können sich im politischen Prozess für neue Phänomene spezielle Normen herausbilden. Als historisches Beispiel kann das Straßenverkehrsrecht dienen, das auf die Herausforderungen reagierte, die durch Mobilität mit Kraftfahrzeugen im Vergleich zu Pferdekutschen entstanden. Derartige Regeln unterscheiden

---

[53] Yeung, K. (2018): A Study of the Implications of Advanced Digital Technologies (Including AI systems) for the Concept of Responsibility within a Human Rights Framework. In: MSI-AUT 5, 1-94. https://ssrn.com/abstract=3286027 [10.02.2023].
[54] Denga, M. (2018): Deliktische Haftung für künstliche Intelligenz. In: Computer und Recht 34 (2), 69-78 (DOI: 10.9785/cr-2018-0203); Burchardi, S. (2022): Risikotragung für KI-Systeme. In: Europäische Zeitschrift für Wirtschaftsrecht 15, 685-692, 685.
[55] Riehm, T. (2020): Nein zur ePerson! In: Recht Digital 1, 42-48, 42.

sich im Hinblick darauf, wie explizit sie an der Technik selbst ansetzen und wie bereichsspezifisch sie sind. Die Regelungskonzepte sind vielfältig.

Nur verhältnismäßig wenige deutsche oder europäische Regelungen nehmen direkt auf Phänomene, wie sie in dieser Stellungnahme beschrieben werden, Bezug. Zu den Ausnahmen zählt die Spezialmaterie des deutschen Finanzmarktrechts, in der das selbstständige Bieten durch algorithmische Systeme geregelt und begrenzt wird, auch um zu verhindern, dass es zu einem von Computern ausgelösten Finanzcrash kommt.[56] Ohne das Wort „Algorithmen" zu erwähnen, wurden die Mediengesetze in Deutschland auf Vielfaltsrisiken bei der Auffindbarkeit von Medieninhalten erweitert, die durch Selektieren und Aggregieren durch sogenannte *Intermediäre* ausgehen – gemeint sind Plattformen wie Facebook oder Suchmaschinen wie Google (§ 2 Abs. 2 Nr. 16, §§ 91–96 MStV). Diese Intermediäre müssen unter anderem ihre zentralen Kriterien der Sammlung, Auswahl und Darstellung von Inhalten und deren Gewichtung offenlegen (Transparenzgebot); zusätzlich gilt ein Diskriminierungsverbot. Auch in anderen Bereichen wurde und wird auf technische Entwicklungen und Veränderungen durch Gesetzgebung reagiert.[57]

Da die Gefahren oft von Akteuren wie international agierenden IT-Unternehmen ausgehen, gewinnt die Regelung auf transnationaler und insbesondere europäischer Ebene zunehmend an Bedeutung. Relevanz hat vor allem die Datenschutzgrundverordnung (DSGVO) erlangt, die seit 2018 einen einheitlichen europäischen Rahmen für den Schutz personenbezogener Daten zur Verfügung stellt. Art. 22 DSGVO enthält das Recht, nicht einer Entscheidung unterworfen zu sein, die ausschließlich auf einer automatisierten Datenverarbeitung beruht; es gibt allerdings Ausnahmen.[58] Diese Regelungen der DSGVO sind zu einem Referenzpunkt der Diskussion über automatisiertes Entscheiden geworden, da auch grundsätzliche Fragen wie das Recht auf Eingreifen durch einen Menschen (*human in the loop*)[59] und auf aussagekräftige Informationen zur involvierten Logik[60] in der DSGVO vorgesehen sind.

---

[56] Kurth, M.-O. (2020): KI und Kapitalmarktrecht. In: Ebers, M. et al. (Hg.): Künstliche Intelligenz und Robotik. München, 484-511.
[57] Überblick bei Guckelberger, A. (2019): Öffentliche Verwaltung im Zeitalter der Digitalisierung. Baden-Baden; Hoffmann-Riem, W. (2022): Recht im Sog der digitalen Transformation. Tübingen.
[58] Laue, P. (2016): Öffnungsklauseln in der DS-GVO – Öffnung wohin? In: Zeitschrift für Datenschutz 10, 463-466, 463.
[59] Buckley, R. P. et al (2021): Regualting artificial intelligence in finance: Putting the human in the loop. In: Sydney Law Review 43 (1), 43-81. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3831758 [19.01.2023].
[60] Kumkar, L. K.; Roth-Isigkeit, D. (2020): Erklärungspflichten bei automatisierten Datenverarbeitungen nach der DSGVO. In: Juristen Zeitung 75 (6), 277-286. (DOI: 10.1628/jz-2020-0090).

Spezifische Gefahren Künstlicher Intelligenz will die Europäische Union mit einem neuen Rechtsakt regeln, dem Artificial Intelligence Act (AI Act).[61] Der Gesetzgebungsprozess zeigt allerdings auch, wie schwer es ist, Phänomene wie Künstliche Intelligenz für Rechtsanwender handhabbar zu definieren und adäquate Schutzkonzepte (etwa abgestuft nach Risiken) zu entwickeln. Rechtlich nicht bindende Empfehlungen gibt es in diesem Bereich bereits.[62] Ein weiteres Gesetzgebungspaket hat Daten als Referenzpunkt: Ein Data Governance Act trat im Juni 2022 in Kraft und schafft Prozesse, Strukturen und einen Rechtsrahmen für die gemeinsame Nutzung von personenbezogenen und nicht personenbezogenen Daten. Ein weiter gehender Data Act ist in Vorbereitung. Auch das Gesetzespaket für digitale Dienste, das die Europäische Union 2022 erlassen hat, bestehend aus dem Digital Markets Act und dem Digital Services Act, hat im vorliegenden Zusammenhang Bedeutung. Digital Markets Act und dem Digital Services Act stellen Regeln für die großen Plattformen und Gatekeeper der digitalen Ökonomie auf, die auch im Bereich der Entwicklung und Anwendung von technischen Systemen führend sind.[63]

Der Regelungsrahmen wird zentral durch rechtliche Normen wie die genannten (und zahlreiche weitere) bestimmt, aber nicht vollständig. Selbstregulierung vor allem durch die Industrie (etwa durch oben bereits erwähnte freiwillig erlassene Codizes) ist ebenso Teil des Regelungsrahmens, wie es beispielsweise Verträge etwa zwischen Plattformen und Nutzenden sind, in denen wiederum Verhaltensregeln festgelegt werden (etwa Community Standards).[64] Das normative System wird dadurch noch komplexer, dass etwa Grund- und Menschenrechte nicht nur staatliche Stellen binden, sondern – jedenfalls nach deutschem Rechtsverständnis – indirekt auch Unternehmen anhalten, etwa die Meinungsfreiheit der Personen, die ihr Angebot nutzen, zu

---

[61] Bomhard, D.; Merkle, M. (2021): Regulation of Artificial Intelligence. In: Journal of European Consumer and Market Law 10 (6), 257-261; Geminn, C. (2021): Die Regulierung Künstlicher Intelligenz. In: Zeitung für Datenschutz 7, 354-359, 354.

[62] Kettemann, M. C. (2022): UNESCO-Empfehlung zur Ethik Künstlicher Intelligenz: Bedingungen zur Implementierung in Deutschland. Herausgegeben von der Deutschen UNESCO-Kommission. Bonn. https://www.unesco.de/sites/default/files/2022-03/DUK_Broschuere_KI-Empfehlung_DS_web_final.pdf [19.01.2023]; Ad Hoc Committee on Artificial Intelligence (CAHAI) (2021):Possible elements of a legal framework on artificial intelligence, based on the Council of Europe's standards on human rights, democracy and the rule of law. Straßburg. https://rm.coe.int/cahai-2021-09rev-elements/1680a6d90d [19.01.2023]; OECD (2022): Recommendation of the Council on Artificial Intelligence. OECD/LEGAL/0449. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 [19.01.2023].

[63] Eifert, M. et al. (2021): Taming the giants: The DMA/DSA package. In: Common Market Law Review 58 (4), 987-1028 (DOI: 10.54648/cola2021065); Gielen, N.; Uphues, S. (2021): Digital Markets Act und Digital Services Act – Regulierung von Markt- und Meinungsmacht durch die Europäische Union. In: Europäische Zeitschrift für Wirtschaftsrecht 627.

[64] Wielsch, D. (2018): Die Ordnungen der Netzwerke. AGB – Code – Community Standards. In: Eifert, M.; Gostomzyk, T. (Hg.): Netzwerkrecht. Baden-Baden, 61-94; Kumkar, L. K. (2022): Plattform-Recht revisited: Umgang mit den Marktordnungen digitaler Plattformen de lege lata et ferenda. In: Zeitschrift für Europäisches Privatrecht 3, 530-564, 530; Mast, T. (2023): AGB-Recht als Regulierungsrecht. In: Juristenzeitung 2023 (im Erscheinen).

wahren. Zudem hat sich in der wissenschaftlichen Beobachtung die Erkenntnis durchgesetzt, dass die existierenden technischen Elemente ihrerseits eine normative Wirkung entfalten, die die Entwicklung prägt. Erst in der Zusammenschau all dieser Elemente zeigt sich der relevante Regelungsrahmen für das Verhältnis von Mensch und Maschine.

In dieser Stellungnahme befasst sich der Deutsche Ethikrat nicht im Detail mit dem hier kurz umrissenen reichhaltigen Fundus an ethischen Leitlinien und aktuellen wie noch in der Erarbeitung befindlichen rechtlichen Regulierungsansätzen, sondern knüpft auf einer anderen Ebene an diesen Rahmen an. Die in den folgenden zwei Kapiteln vorgelegte Analyse der Konsequenzen digitaler Entwicklungen für das menschliche Zusammenleben gründet auf einer philosophischen Auseinandersetzung mit den anthropologischen Grundbegriffen, die im Mittelpunkt des menschlichen Selbstverständnisses stehen (vgl. Kapitel 3), und entwickelt darauf aufbauend ein Verständnis von Mensch-Technik-Relationen, demzufolge es entscheidend darauf ankommt, wie die Delegation menschlicher Tätigkeiten an Maschinen und algorithmische Systeme auf zentrale anthropologische Konzepte zurückwirkt und dabei insbesondere menschliche Autorschaft erweitert oder vermindert (vgl. Kapitel 4). Im zweiten Teil der Stellungnahme wird dieses Konzept auf die vier ausgewählten Handlungsfelder Gesundheit, Bildung, Kommunikation und Verwaltung angewendet und mündet jeweils in bereichsspezifische Empfehlungen, bevor im dritten Teil die sektorübergreifend identifizierten normativen Querschnittsthemen noch einmal näher beleuchtet werden.

+++ Vorabfassung! Wird durch die lektorierte Fassung ersetzt. +++

3 Zentrale Begriffe und philosophische Grundlagen

3.1 Künstliche Intelligenz: Begriffliche Analyse

Der Begriff der Künstlichen Intelligenz wurde und wird keineswegs immer einheitlich verwendet: Seine Bedeutung hat sich im Laufe der Jahre verändert und unterscheidet sich sowohl innerhalb als auch zwischen verschiedenen Berufsgruppen und Disziplinen. Eine Unterscheidung, welche in der KI-Forschung selbst, mehr noch aber in der Reflexion sowie der medialen und öffentlichen Debatte über KI eine große Rolle spielt, ist die Unterscheidung zwischen sogenannter *schwacher* und *starker* KI.[65] Während sich insbesondere die aktuelle Forschung innerhalb der Informatik vorrangig mit Fragen der schwachen KI beschäftigt, sind öffentliche Debatten ebenso wie manche Fachdiskurse regelmäßig geprägt von Visionen einer starken – also menschenähnlichen oder gar menschliche Fähigkeiten übertreffenden – KI, die mit Sorgen ebenso wie mit Hoffnungen behaftet ist. Um die hinter den jeweiligen Begriffsverwendungen stehenden substanziellen Annahmen über die Grundlagen menschlicher wie Künstlicher Intelligenz herauszuarbeiten, ist eine Klärung der Begriffe notwendig. Dabei zeigt sich, dass unterschiedliche Einschätzungen hinsichtlich der Wahrscheinlichkeit und dem Erwünschtsein einer starken KI auch von unterschiedlichen Konzeptualisierungen des mit KI umschriebenen Phänomenbereichs abhängen.

Neben der heute häufig im Kontext der Reflexion auf KI vornehmlich verwendeten Unterscheidung von starker versus schwacher KI, sind in der Geschichte der KI und auch in der gegenwärtigen internationalen Debatte andere Unterscheidungen in Gebrauch, um unterschiedliche Formen oder Grade der Annäherung künstlicher an menschliche Intelligenz zu beschreiben. Dies sind insbesondere die nachfolgend noch näher betrachteten Unterscheidungen von *spezi-*

[65] Nida-Rümelin, J. (2022): Überblick über die Verwendung der Begriffe starke & schwache Künstliche Intelligenz. In: Chibanguza, K.; Kuß, C.; Steege, H. (Hg.): Handbuch Künstliche Intelligenz. Recht und Praxis automatisierter und autonomer Systeme. Baden-Baden, 75-90; Nida-Rümelin, J. (2022): Digitaler Humanismus – philosophische Aspekte Künstlicher Intelligenz. In: Chibanguza, K.; Kuß, C.; Steege, H. (Hg.): Handbuch Künstliche Intelligenz. Recht und Praxis automatisierter und autonomer Systeme. Baden-Baden, 29-40.

eller versus *allgemeiner/ genereller* KI sowie *enger* KI versus *breiter* KI Darüber hinaus werden in den letzten Jahren, vor allem im Kontext des Transhumanismus[66] sowie der Diskussionen um Singularität[67], die Begriffe *Artificial General Intelligence* oder *Super-Intelligence* verwendet, die im Folgenden allerdings keine Rolle spielen werden.

Den jeweiligen Unterscheidungen liegt trotz einzelner Bedeutungsunterschiede das Bemühen zugrunde, menschliche Intelligenz als Maßstab der Bestimmung Künstlicher Intelligenz heranzuziehen. Über diesen Maßstab glaubt man zu verfügen, weil Intelligenztests eine (wenn auch in Grenzen) erfolgreiche Operationalisierung des Begriffs der Intelligenz erlauben (vgl. Abschnitt 3.2.1). Auf dieser Grundlage erscheint es möglich, menschliche Intelligenz zu simulieren bzw. das Vorliegen von Intelligenz in Maschinen identifizieren und messen zu können. Eine Simulation menschlicher Intelligenz stand lange im Zentrum der Forschung zu KI und der Reflexion über ihre Möglichkeiten und Grenzen. Auch wenn Methoden der KI für ganz andere Zwecke verwendet werden, beispielsweise die Nutzung zur Auswahl passender Werbung, so gilt seit der Veröffentlichung von Turings „Can Machines Think?“ im Jahre 1950 und seiner Formulierung des berühmten Turing-Tests (vgl. Abschnitt 2.1) die äußere Ununterscheidbarkeit zwischen menschlichen und maschinellen kognitiven und operativen Leistungen weithin als „Lackmustest“ für das Vorliegen Künstlicher Intelligenz.

Die Unterscheidung zwischen *spezieller und allgemeiner/genereller KI* wurde in den 1950er-Jahren eingeführt, um die damals aktuelle Forschung zu KI von der Vision einer menschenähnlichen KI abzugrenzen. Während Maschinen zumeist für spezielle Tätigkeiten konstruiert sind, kennzeichnet menschliche Intelligenz eine Vielfalt von Kompetenzen und ein weites Spektrum unterschiedlicher Zwecksetzungen. Man erhoffte sich vom Computer von Anbeginn das Potenzial einer „universellen Maschine“, die alle Aufgaben lösen könne, die sich binär, zum Beispiel als eine Abfolge von Nullen und Einsen, darstellen ließen.[68] Die Hoffnung war, dass der Computer eine generelle Künstliche Intelligenz erreichen könnte, die über das Lösen spezifischer Aufgaben weit hinausgeht. Ein Ziel der frühen Forschung war die Entwicklung des Computers zum General Problem Solver (GPS). Der GPS sollte basierend auf menschlichen Heuristiken

[66] Unter Transhumanismus werden Positionen verstanden, wonach die digitalen Technologien es uns erlauben die Beschränktheiten der menschlichen Existenzform zu überwinden und wir diese Möglichkeiten nutzen sollten, um alte Menschheitsträume zu verwirklichen, wie die einer durch Verbindungen von Gehirn und Computer (brain computer interfaces) erzeugten Vervielfachung der Intelligenz oder der individuellen Fortexistenz in Form einer Software-Kopie des eigenen Gehirns.

[67] Der Begriff der Singularität verweist auf einen möglichen Umschlagpunkt in der Zukunft, ab dem Künstliche Intelligenz menschliche Intelligenz in jeder Hinsicht übertreffen und sich fortan unkontrolliert selbst weiterentwickeln könnte.

[68] Moor, J. H. (1985): What is Computer Ethics? In: Metaphilosophy 16 (4), 266-275.

Problemlösestrategien für ausreichend formalisierbare Probleme in verschiedenen Kontexten liefern. Er kann als Vorläufer heutiger Expertensysteme gelten. Die Komplexität der Aufgaben, die mit dem GPS gelöst werden konnten, blieb allerdings beschränkt. Die Vision einer allgemeinen KI jedoch hat bis heute Bestand.

Das Begriffspaar *speziell* und *allgemein* enthielt zumeist sowohl explizite Annahmen über die Breite des Fähigkeitenspektrum einer KI als auch Erwartungen zur grundlegenden Qualität der zukünftig daraus resultierenden Künstlichen Intelligenz (bis hin zum ontologischen Status). Entsprechend wurde im Verlauf weiter differenziert. Etabliert sind inzwischen die Unterscheidungen zwischen *enger* und *breiter* KI sowie *schwacher* und *starker* KI. Die Forschung zu KI ist heute zumeist auf einen klar umrissenen Anwendungsbereich begrenzt, etwa die Interpretation von Röntgenaufnahmen, und somit ein Fall von enger KI. Doch auch hier besteht die Vision einer breiten Ausweitung des Spektrums. Dabei ist zu betonen, dass der Unterscheidung zwischen engen und breiten Formen der KI, ebenso wie die Begriffe *speziell* und *allgemein*, nicht als Gegensätze zu verstehen sind, sondern jeweils Endpunkte eines Fähigkeitenspektrums beschreiben.

Ein aktueller Bereich, in dem die Frage nach breiter KI verhandelt wird, umfasst Sprachverarbeitungssysteme wie das im vorherigen Kapitel bereits erwähnte GPT-3 und ChatGPT. Solche Systeme produzieren Texte, bei denen oft schwer oder unmöglich zu erkennen ist, ob sie von einem Menschen oder einer Maschine verfasst wurden. Erste prägnante Beispiele entfachten eine intensive philosophische und auch öffentliche Debatte dazu, ob diese Form der Textproduktion nur eine Verbreiterung des Fähigkeitenspektrums darstellt, oder bereits als ein Übergang zur oder gar eine Manifestation einer generellen oder starken KI gelten könne – wenn auch nur mit Blick auf das Sprachvermögen.[69]

Die Verwendung der verschiedenen Begriffe zur Charakterisierung Künstlicher Intelligenz als spezifischer, enger, oder schwacher Intelligenz einerseits sowie allgemeiner, breiter oder starker Intelligenz andererseits, verweist nicht nur auf Differenzen zwischen den beiden jeweiligen

[69] Der Titel eines Artikels des MIT Technology Reviews (Heavon 2020) fasst diese Debatte mit Blick auf GPT-3 zusammen: Heavon, W. D. (2020): OpenAI's new language generator GPT-3 is shockingly good – and completely mindless. In: MIT Technology Review. https://www.technologyreview.com/2020/07/20/1005454/openai-machine-learning-language-generator-gpt-3-nlp/ [22.12.2022]. Der Titel spielt auf die Kritik von J. Searle an; vgl. dazu Abschnitte 3.1 und 3.4.2. Mit Blick auf jüngere Chatbots löste 2022 zudem der ehemalige Google-Mitarbeiter Blake Lemoine eine Debatte aus, als er postulierte, der von Google entwickelte Chatbot *LaMDA* sei bei Bewusstsein – ein Vorstoß, der später zu seiner Entlassung führte. Tiku, N. (2022): The Google engineer who thinks the company´s AI has come to life. In: Washington Post. https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/ [31.01.2023].

Arten bzw. Polen von Intelligenz. Dahinter verstecken sich, insbesondere beim Begriffspaar der schwachen und starken KI, vielmehr auch unterschiedliche Verständnisse von Intelligenz sowie unterschiedliche Positionen hinsichtlich der Kernfrage, ob es qualitative und kategorische oder nur quantitative und prinzipiell überwindbare Unterschiede zwischen menschlicher und Künstlicher Intelligenz gibt.

Wichtig für die jeweilige Beantwortung dieser Frage ist zum einen die Differenz hinsichtlich der Breite bzw. Enge des Fähigkeitenspektrums der Künstlichen Intelligenz. Die meisten Anwendungen Künstlicher Intelligenz entfalten ihre jeweilige Leistung auf klar umrissenen, engen Gebieten oder Domänen wie beispielsweise dem Spielen von Schach oder Go. Hier sind sie im direkten Vergleich Menschen inzwischen klar überlegen. Sprachproduktionssysteme wiederum sind zwar ebenfalls auf einen Kompetenzbereich beschränkt (sprachliche Ein- und Ausgabe), jedoch mittlerweile nicht mehr auf eine Domäne, *über* die gesprochen wird; hier erfolgt also eine jedenfalls funktionelle Verbreiterung des Fähigkeitenspektrums. Dennoch fehlt auch ihnen jedwedes sprachliche Verständnis über die Bedeutung der rezipierten oder produzierten Worte, operieren sie doch allein auf Basis der Wahrscheinlichkeit von Wortkombinationen.

Zum anderen hatte der Philosoph John Searle schon 1980 gegen den Turing-Test die These aufgestellt, dass die bloße Ununterscheidbarkeit von menschlicher und maschineller Sprachperformanz nicht ausreicht, um ein Text*verständnis* anzunehmen.[70] Hier geht es also zusätzlich darum, ob es jenseits einer quantitativen Differenz auch einen kategorialen Unterschied zwischen Mensch und Maschine gibt, bzw. darum, ob Intelligenz an bestimmte mentale Voraussetzungen geknüpft ist, welche über die bloße Simulation von Verständnis hinausgehen. Anders formuliert ergibt sich also die Frage, ob Intelligenz in allgemeiner oder starker Form jemals vollumfänglich Maschinen zukommen kann oder ob dafür spezifisch menschliche Eigenschaften Voraussetzung sind.[71]

Die Differenzen hinter den Begriffspaaren schwache versus starke, enge versus breite und spezielle versus allgemeine KI lassen sich vor diesem Hintergrund wie folgt zusammenfassen:

---

[70] Searle, J. (1980): Minds, Brains and Programs. In: Behavioral and Brain Sciences 3 (3), 417–457, (DOI: 10.1017/S0140525X00005756). Searles Gedankenexperiment und seine Kritik werden ausführlicher im Abschnitt 3.4 Anthropologie behandelt.

[71] Die Frage tierlicher Intelligenz wird im Rahmen dieser Stellungnahme ausgeklammert; vgl. dazu Huber, L. (2021): Das rationale Tier: Eine kognitionsbiologische Spurensuche. Berlin; sowie zur grundsätzlicheren Frage nach mentalen Zustände bei Tieren: Deutscher Ethikrat (2020): Tierwohlachtung – Zum verantwortlichen Umgang mit Nutztieren. Berlin.

Zum einen geht es um die *Breite der Fähigkeiten*, über die eine KI verfügt, sowohl graduell innerhalb von Domänen (z. B. Sprachverarbeitung) als auch bereichsübergreifend (z. B. Sprache und Motorik, Situationserfassung). Zum anderen geht es um die Antwort auf die Frage, ob die *Simulation von Intelligenz* mit *Intelligenz* gleichzusetzen ist, oder ob es einen kategorischen Unterschied zwischen der *Simulation von Verständnis* und *genuinem* (z. B. sprachlichem) *Verständnis* gibt, der für „echte" Intelligenz essenziell ist, über die jedenfalls die in dieser Stellungnahme diskutierten Systeme nicht verfügen.[72]

Am Beispiel der Sprachproduktionssysteme lassen sich die vorgestellten unterschiedlichen Verständnisse von KI noch einmal veranschaulichen. Bereits die Antwort auf die Frage, ob sie Beispiele für breite bzw. allgemeine KI sind, hängt von den jeweiligen Annahmen ab: Zwar sind einzelne Sprachproduktionssysteme nicht auf eine Domäne beschränkt und haben somit ein durchaus breites Funktionenspektrum. Um allerdings die Anforderungen an breite bzw. allgemeine KI zu erfüllen, würde gemeinhin verlangt, dass das System nicht nur im Bereich der Sprache menschliche Kompetenz (nahezu) perfekt simuliert, sondern dies eben auch zeitgleich in (allen) anderen Bereichen vermag, die gemeinhin im menschlichen Kontext als intelligentes Verhalten klassifiziert werden, wie beispielsweise koordinierte Bewegung im Raum und so weiter. Tatsächlich deuten bestimmte, auch bei sehr guten Sprachproduktionssystemen auftretende Fehlleistungen darauf hin, dass die hohe Leistungsfähigkeit nicht auf einem inhaltlichen Verständnis der Texte beruhen kann. Allgemein intelligentes Verhalten ist bei den gegenwärtigen Systemen schon funktional noch in weiterer Ferne und bedürfte zudem auch noch des Einbaus in einen physischen humanoiden Roboter. In einem humanoiden Roboter mit perfekten Bewegungsfähigkeiten und einer menschenähnlichen Mimik und Gestik würden manche ein Beispiel breiter oder gar starker KI sehen, wenn er in der Lage wäre, alle menschlichen kognitiven Fähigkeiten perfekt zu simulieren. Andere würden hingegen bestreiten, dass damit eine Form starker KI vorliegt, da auch eine perfekte Simulation nicht garantiere, dass ein solcher humanoider Roboter mentale Zustände aufweist, über Einsichts- und Urteilsfähigkeit sowie über emotive Einstellungen wie Hoffnungen und Ängste verfüge.

Die unterschiedlichen Konzeptionen hinter den diversen Begrifflichkeiten zur KI gehen auch auf verschiedene grundlegende anthropologische Theoriemodelle zurück (vgl. Abschnitt 3.4). Aus behavioristischer Sicht ist die Unterscheidung zwischen Simulans und Simulandum nicht

[72] Umstritten blieb, ob diese Aussage sich lediglich auf die aktuellen Systeme und die der absehbaren Zukunft bezieht, oder grundsätzlich zu verstehen ist.

sinnvoll, da diese epistemisch nicht unterscheidbar seien. In anderen Konzeptionen[73] jedoch werden mentale Zustände realistisch, das heißt als Merkmale der ontologischen Ausstattung der Welt interpretiert. In solchen Konzeptionen wird an einer kategorischen Unterscheidung zwischen Simulation und Realisierung festgehalten, auch wenn diese Differenz epistemisch nicht unmittelbar zugänglich ist. Entscheidend für den Unterschied zwischen menschlicher und Künstlicher Intelligenz ist demnach das Vorhandensein bestimmter mentaler Eigenschaften wie beispielsweise Verständnis oder Bewusstsein. In dieser Stellungnahme werden die Begriffe zur Charakterisierung der unterschiedlichen Formen Künstlicher Intelligenz in der unten stehenden Weise verwendet. Es wird hierbei vorausgesetzt, dass die Unterscheidung zwischen enger und breiter KI quantitativer bzw. gradueller Natur ist, die Entstehung einer starken KI jedoch einen qualitativen Sprung bedeuten würde:

Enge KI: KI-Anwendungen, die menschliche Fähigkeiten in einer Domäne simulieren bzw. Verfahren wie maschinelles Lernen verwenden, um spezifische Aufgaben zu erfüllen oder spezifische Probleme zu lösen. Nahezu alle derzeit verwendeten KI-Anwendungen fallen in diese Kategorie.

Breite KI: Breite KI-Anwendungen erweitern das Spektrum ihrer Anwendbarkeit über einzelne Domänen hinaus. Sprachproduktionssysteme wie etwa GPT-3 können als Beispiele für breiter werdende KI gelten, da sie zwar nicht domänenspezifisch, jedoch weiterhin auf sprachliche Ein- und Ausgabe beschränkt sind. Eine mögliche Zukunftsvision breiter KI wären Systeme, die solche Sprachkompetenzen mit weiteren kognitiven oder motorischen Kompetenzen zusammenführen, etwa durch Einbau in weitentwickelte Roboter.

Starke KI: Der Begriff der starken KI wird für die Vision einer Künstlichen Intelligenz verwendet, die jenseits der möglicherweise perfekten Simulation menschlicher Kognition auch über mentale Zustände, Einsichtsfähigkeit und Emotionen verfügen würde.

3.2 Intelligenz und Vernunft

3.2.1 Intelligenz

Die im vorigen Abschnitt vorgestellten unterschiedlichen Deutungen von Künstlicher Intelligenz werden seit mindestens den Siebzigerjahren des 20. Jahrhunderts von kontroversen Diskussionen der Frage begleitet, was Computer können und nicht können bzw. demnächst können

[73] Z. B. phänomenologische Positionen oder solche der intentionalistischen Semantik.

und nicht können werden.[74] Um diese Frage zu beantworten, müsste zunächst geklärt werden, von welchen Vorstellungen hinsichtlich der menschlichen Intelligenz dabei ausgegangen wird. Dazu wird jedoch in den sich damit beschäftigenden Wissenschaften, insbesondere der Psychologie, Philosophie und Informatik keine einheitliche Antwort angeboten.

Aus psychologischer Perspektive ist Intelligenz als ein hypothetisches Konstrukt aufzufassen, das als solches zwar verbal umschrieben werden kann, zum Beispiel im Sinne von Verstehen, Urteilen und Schlussfolgern[75] oder zielgerichtetem Handeln, rationalem Denken und effektiver Auseinandersetzung mit der Umwelt[76], aber nicht beobachtbar ist, sondern anhand von Indikatoren in relevanten Aspekten operationalisiert werden muss. In diesem Sinne sind Intelligenztests als Situationen aufzufassen, in denen Menschen Verhalten zeigen können, das vor dem Hintergrund eines theoretischen Vorverständnisses oder einer zugrunde gelegten Definition als mehr oder weniger „intelligent“ bezeichnet werden kann. Dabei können die gewählten Operationalisierungen – auf der Ebene von Teilkomponenten und Subskalen wie auf der Ebene von Testitems – sehr unterschiedlich sein. Sie bewähren sich im Kontext der Bestimmung von Gütekriterien, auf deren Grundlage abgeschätzt werden kann, inwieweit die Durchführung, Auswertung und Interpretation der Testung als objektiv, verlässlich (reliabel) und valide angesehen werden können.

Intelligenz wird hier im Sinne eines Abweichungsquotienten verstanden; der Mittelwert des (normalverteilten) Merkmals liegt in der Grundgesamtheit per definitionem bei 100, die Standardabweichung bei 15.[77] Dies hat Auswirkungen auf die Testkonstruktion, denn bei der Auswahl potenzieller Testitems muss berücksichtigt (geschätzt) werden, wie sich diese in der Grundgesamtheit verteilen, mit anderen Items korrelieren und zwischen verschiedenen Fähigkeitsniveaus unterscheiden. Bei der Validierung von Intelligenztests interessieren Korrelationen mit anderen Verfahren und Merkmalen, die aus theoretischer Perspektive als Indikatoren der Ausprägung verwandter und nicht verwandter Konstrukte betrachtet werden können und deshalb eine bestimmte Höhe aufweisen bzw. diese nicht überschreiten sollten. Darüber hinaus

[74] Dreyfus, H. L. (1992): What Computers still can’t do: a critique of artificial reason. 2., überar. Auflage. Cambridge (MA).
[75] Binet, A.; Simon, T. (1904): Methodes nouvelles pour le diagnostiqc du niveau intellectuel des anormaux. In: L'Année Psychologique 11, 191–244 (DOI:10.3406/psy.1904.3675).
[76] Wechsler, D. (1944). The Measurement of Adult Intelligence. Baltimore (DOI: 10.1037/11329-000).
[77] Das bedeutet, dass ca. 68% der Menschen einen IQ zwischen 85 und 115 sowie ca. 95% einen IQ zwischen 70 und 130 aufweisen.

ist die Möglichkeit, auf der Grundlage aktueller Messwerte zukünftige Leistungen bzw. interindividuelle Unterschiede in relevanten Merkmalen (z. B. Erfolg versus Misserfolg in Schule und Beruf) zu prognostizieren, von großem Interesse.

Der Wechsler-Intelligenztest (Wechsler Adult Intelligence Scale), auf dem die bekanntesten modernen Intelligenztests aufbauen, wurde 1955 von David Wechsler entwickelt und bereits 1956 in einer deutschsprachigen Version, dem Hamburg-Wechsler-Intelligenztest für Erwachsene verwendet. Die aktuelle Version des Tests besteht aus zehn Kerntests und fünf optionalen Untertests, welche unterschiedliche kognitive Fähigkeiten prüfen, die in vier Aufgabengruppen zusammengefasst werden: Sprachverständnis, wahrnehmungsgebundenes logisches Denken, Arbeitsgedächtnis und Verarbeitungsgeschwindigkeit. Hierbei ist wichtig festzuhalten, dass es keine Aufgaben gibt, die diese Dimensionen direkt messen. Vielmehr sind die verwendeten Skalen induktiv aus den im Kontext der Testkonstruktion und -normierung empirisch ermittelten Korrelationen zwischen Unterskalen und Außenkriterien abgeleitet. Das heißt, die Struktur der Intelligenz ergibt sich wesentlich induktiv aus der empirischen Erfassung und Validierung unterschiedlicher, aus dem theoretischen Vorverständnis der Testautorschaft abgeleiteten Aspekte bzw. Dimensionen kognitiver Leistungsfähigkeit (z. B. der Orientierung an einem Generalfaktormodell in der Tradition von Charles Spearman versus. einem Modell voneinander unabhängiger Primärfaktoren in der Tradition von Louis Leon Thurstone).[78]

Die Frage, ob Intelligenz eine einheitliche Fähigkeit ist oder viele Fähigkeiten umfasst, die gegebenenfalls auch voneinander unabhängig sein können, ist empirisch nicht eindeutig zu klären – die dimensionale Struktur wird auf der Grundlage von vorab festgelegten Modellannahmen bzw. Restriktionen ermittelt, die ihrerseits nicht Gegenstand einer empirischen Überprüfung werden können. Allerdings lässt sich festhalten, dass empirisch durchweg positive Korrelationen zwischen den verschiedenen Untertests bzw. Aufgabengruppen nachzuweisen sind. Im erwähnten Wechsler-Intelligenztest bedeutet dies etwa, dass höhere Werte in den Tests der Aufgabengruppe *Sprachverständnis* statistisch mit höheren Werten in den anderen drei Aufgabengruppen einhergeht, auch wenn das Ausmaß der Korrelationen zwischen den verschiedenen Tests je nach Nähe der Aufgaben variiert.[79] Diese durchweg positive Korrelation führte zur Annahme des sogenannten *Generalfaktors* der Intelligenz *g*, welcher den Anteil der

[78] Vgl. Sternberg, R. J. (Hg.) (2020): The Cambridge Handbook of Intelligence. 2. Auflage. Cambridge.
[79] Nach Deary liegt die durchschnittliche Korrelation zwischen allen Tests bei 0.45. Aufgrund ihres statistischen Ursprungs, ist Korrelation innerhalb der vier Skalen höher als zwischen den Skalen. Die höchste Korrelation von 0.74 besteht zwischen Wortschatz-Test und dem Test für Allgemeines Verständnis. Deary, I. J. (2020): Intelligence: A Very Short Introduction. Oxford.

allgemeinen Intelligenz bzw. der kognitiven Leistungen zugrunde liegenden allgemeinen geistigen Fähigkeit bezeichnet, deren Ausprägung sich – weil unidimensional – in einem einzigen Wert ausdrücken lässt, und ca. 40 Prozent der in Leistungsmessungen beobachteten Varianz erklärt. Die übrigen 60 Prozent lassen sich demnach auf unterschiedliche spezifische Fähigkeiten zurückführen.

Innerhalb der Allgemeinen Psychologie und insbesondere mit der Entwicklung der Kognitionspsychologie in den 1970er-Jahren rückte zunehmend die Analyse der Prozesse in den Fokus, die nötig sind, um die Aufgaben der Intelligenztests zu lösen. Umfangreiche Forschungen zur Bearbeitung informationsverarbeitender Aufgaben wie Informationskodierung und geteiltes Hören führten zu der Annahme, dass (verbale) Intelligenz durch die Fähigkeit zur Auswahl und Benutzung von Informationsverarbeitungsmethoden bestimmt wird.[80] Die Kernthemen der kognitionspsychologischen Forschung jedoch, nämlich Denken, Problemlösung und Entscheidungsfindung, die auch außerhalb der Disziplin der Psychologie häufig mit dem Begriff Intelligenz assoziiert werden und auf die häufig in der Entwicklung Künstlicher Intelligenz Bezug genommen wird, fallen in der Psychologie nicht unter den Intelligenzbegriff. Dies mag mit der zunehmenden Spezialisierung innerhalb der Psychologie zusammenhängen. So ist die Erfassung menschlicher Intelligenz ein Teilgebiet der Differenziellen Psychologie, die sich mit Unterschieden zwischen Menschen befasst, wohingegen die Allgemeine Psychologie sich mit den Grundlagen von Wahrnehmung, Lernen sowie insgesamt menschlicher Kognition und Emotion beschäftigt.

Verwiesen sei in diesem Zusammenhang auch auf die Unterscheidung zwischen Intelligenz und Kreativität[81], wobei Letztere in Anlehnung an Guilford, den Begründer der modernen Kreativitätsforschung, als flüssige, flexible und ursprüngliche Erzeugung von Konzepten von Lösungen für neuartige Probleme definiert werden kann.[82] Von Interesse ist hier seine Unterscheidung zwischen konvergentem und divergentem Denken.[83] Im Unterschied zum konvergenten Denken, das durch logische Schlussfolgerungen zu einer einzigen oder besten Lösung gelangt (wobei das Ergebnis mehr oder weniger vollständig durch die vorhandene Information determiniert ist), liefert das für Kreativität charakteristische divergente Denken mehrere alternative

[80] Hunt, E. et al. (1975): What does ist meant to be high verbal? In: Cognitive Psychology, 7 (2), 194-227 (DOI: 10.1016/0010-0285(75)90010-9).
[81] Kruse, A.; Schmitt, E. (2011): Die Ausbildung und Verwirklichung kreativer Potenziale im Alter. In dies. (Hg.): Kreativität im Alter. Heidelberg, 15-46; Lubart, T. I. (2018): The Creative Process: Perspectives from Multiple Domains. Paris.
[82] Guilford, J. P. (1950): Creativity. In: American Psychologist, 5(9), 444-454 (DOI: 10.1037/h0063487).
[83] De Vries, H. B.; Lubart, T. I. (2017): Scientific Creativity: Divergent and Convergent Thinking and the Impact of Culture. In: The Journal of Creative Behavior, 53 (2), 145-155 (DOI: 10.1002/jocb.184).

Lösungen, die jeweils den gegebenen Anforderungen entsprechen. Dabei gelten die Anzahl der generierten Lösungen und deren Qualität als Maß für die Ausprägung des divergenten Denkens. Neben dem divergenten Denken wurden und werden auch weitere kognitive Prozesse als zentrale Voraussetzungen für Kreativität diskutiert. Zu nennen sind hier insbesondere Fähigkeiten und Fertigkeiten im Bereich von Wahrnehmung, Problemdefinition, Einsicht, Induktion, Bildung von Analogien und ungewöhnlichen Assoziationen, die Bewertung von Ideen und die Organisation von Wissenssystemen.

Während der Begriff der Intelligenz ursprünglich auf ein Ensemble menschlicher Leistungen verweist, wie sie in klassischen Intelligenztests gemessen werden[84], hat sich der Blick auf Intelligenz in jüngerer Zeit sukzessive erweitert. So entstanden Konzepte wie die der sozialen bzw. emotionalen Intelligenz, welche einerseits den Begriff der Intelligenz auf weitere Fähigkeiten ausweiteten sowie andererseits die Wechselwirkungen zwischen Emotion und Kognition sowie den sozialen und kulturellen Aspekt von Intelligenz in den Blickpunkt rückten. Darüber hinaus entwickelte sich rund um die Stichwörter *embodied*, *embedded*, *enactive* und *extended cognition* ein Forschungsfeld, das in Philosophie, Psychologie und Robotik die Rolle des Körpers einerseits und der Umwelt andererseits für Intelligenz und kognitive Leistungen erforscht.[85]

Spätestens diese Erweiterungen werfen die Frage auf, wie die Übertragung des Intelligenzbegriffs auf technische Artefakte zu verstehen ist. Bei einigen Merkmalen wie beispielsweise der elementaren Rechenfähigkeit, dem logischen Schlussfolgern oder Gedächtnisleistungen entsteht der Eindruck, dass menschliche Fähigkeiten eindeutig auf technische Artefakte übertragen werden können. Auch diesbezüglich sind schon kritische Fragen zu stellen, zum Beispiel ob die menschliche Erinnerungsfähigkeit in gleicher Weise eine Gedächtnisleistung ist wie die Aktivierung eines technischen Speichers. Einerseits sind die quantitativen Leistungen technischer

[84] Verwiesen sei hier auf die von Thurstone faktorenanalytisch ermittelten sieben Primärfaktoren induktives Schließen, räumliches Vorstellungsvermögen, Wahrnehmungsgeschwindigkeit, Rechenfähigkeit, verbales Verständnis, assoziatives Gedächtnis und Wortflüssigkeit (vgl. Thurstone, L.L. (1938). Primary mental abilities. University of Chicago Press: Chicago. Thurstone, L.L. & Thurstone, G.W. (1941). Factorial Studies Of Intelligence. University of Chicago Press: Chicago.) sowie die Differenzierungen zwischen fluider vs. kristalliner Intelligenz bei John Horn und Raymond Catell (Horn, J.L. & Cattell, R.B. (1966). Refinement and test of the theory of fluid and crystallized intelligence. Journal of Educational Psychology, 57, 253 – 270.) und kognitiver Mechanik vs. Pragmatik bei Paul Baltes (Baltes, P. B. (1987). Theoretical Propositions of Lifespan Developmental Psychology: On the Dynamics between Growth and Decline. Developmental Psychology, 23, 611-626. http://dx.doi.org/10.1037/0012-1649.23.5.611).

[85] Clark, A (2012): Embodied, embedded, and extended cognition. In: Frankish, K.; Ramsey, W. (Hg.): The Cambridge Handbook of Cognitive Science. (DOI: 10.1017/CBO9781139033916.018).

Speicher der menschlichen Erinnerungsfähigkeit um Größenordnungen überlegen. Andererseits sortiert der Mensch seine Gedächtnisleistungen beispielsweise nach der jeweils kontextuell bestimmten Bedeutung, während ein technischer Speicher unterschiedslos je nach den technischen Vorgaben Daten aufnimmt oder nicht. Bei Intelligenzleistungen mit emotiven und kreativen Qualitäten verstärkt sich der Verdacht, dass es sich hierbei um anthropomorphe Übertragungen handelt. Die Klärung solcher Vergleichbarkeitsprobleme hängt somit wesentlich von den Kriterien ab, durch die man eine spezifisch menschliche Intelligenzleistung bestimmt sieht. Man sollte daher die Verwendung des Ausdrucks „Intelligenz" in der Wortverbindung „Künstliche Intelligenz" eher als eine Metapher einordnen, deren Beschreibungs- und Erklärungsfunktion genauerer Aufklärung bedarf.

3.2.2 Vernunft

Bereits lange vor der Einführung des Begriffs der Intelligenz wurde der Begriff der Vernunft verwendet, um die spezifische menschliche Fähigkeit zu kennzeichnen, sich in der Welt zu orientieren, selbstverantwortlich zu handeln und so der eigenen Lebenspraxis eine kohärente Struktur zu geben. Intelligenz ist für Vernunft eine wichtige Voraussetzung, aber keine hinreichende Bedingung. Der Begriff der Vernunft gehört zu den basalen Grundkategorien menschlicher Selbst- und Weltdeutung, die unsere Kultur seit der Antike maßgeblich geprägt haben. Schon das weite Wortfeld (griechisch: *logos, nous, dianoia, phronesis*; lateinisch: *ratio, mens, intellectus, prudentia*) deutet darauf hin, dass es sich um einen überaus komplexen Begriff handelt, der vielfältige Binnendifferenzierungen kennt und verschiedene (kognitive) Teilkompetenzen umfasst. Strukturell geht es um ein mehrdimensionales Beziehungsgefüge von Denk-, Reflexions- und Operationsformen, das in seiner Gesamtheit im Dienste einer möglichst adäquaten Wirklichkeitserschließung steht und in einen komplexen sozialen und kulturellen Kontext verwoben ist. Als Inbegriff bestimmter Ansprüche, denen wir uns im Denken, Sprechen, Erleben und Handeln unterstellen, umfasst der Vernunftbegriff unterschiedliche – propositionale und nichtpropositionale – Wissensformen und Rationalitätstypen, die von methodisch-prozeduralem Know-how über ästhetische Wahrnehmungsfähigkeit und Kreativität sowie verschiedene soziale Interaktionsfähigkeiten bis hin zu einer umfassenden Lebensführungskompetenz reichen. Von grundlegender Bedeutung für unsere Thematik ist dabei die Gegenüberstellung von *theoretischer Vernunft*, die sich auf den Erkenntnisgewinn richtet, um zu wahren empirischen oder apriorischen Urteilen zu gelangen, und *praktischer Vernunft*, die auf ein kohärentes, verantwortliches Handeln abzielt, um ein gutes Leben zu ermöglichen.

Vor allem im Blick auf den Gebrauch der *theoretischen* Vernunft, der primär auf Erkenntnisgewinn durch die Formulierung wahrer empirischer Urteile abzielt, scheinen sich zumindest prima facie einige Parallelen zur Arbeitsweise von KI-Systemen aufzudrängen. So spielen in beiden Bereichen Fähigkeiten der Informationsverarbeitung, des Lernens, des logischen Schlussfolgerns und konsistenten Regelfolgens sowie der sinnvollen Verknüpfung gespeicherter Daten eine zentrale Rolle. Bei näherer Betrachtung zeigen sich jedoch insofern gravierende Differenzen, als sich nicht nur die Arbeitsweise des menschlichen Gedächtnisses in mehrfacher Hinsicht vom technischen Speicher eines Computers unterscheidet, sondern auch die menschliche Urteilspraxis technisch nicht substituierbar ist. Auch wenn in diesem Zusammenhang die wahrheitstheoretischen Implikationen der Formulierung und Begründung deskriptiver Urteile nicht näher entfaltet werden können[86], ist doch darauf hinzuweisen, dass zumindest die bislang verfügbaren KI-Systeme die dafür relevanten Fähigkeiten des Sinnverstehens, der Intentionalität und der Referenz auf eine außersprachliche Wirklichkeit nicht besitzen.

Dieser Befund bestätigt sich auch bezüglich der uns hier besonders interessierenden *praktischen* Vernunft, die insofern noch weit komplexerer Natur ist, als ihr Ziel nicht nur in wohlbegründeten praktischen Einzelurteilen, sondern in einem möglichst richtigen und verantwortlichen Handeln besteht, das über einen langen Zeitraum aufrechterhalten wird, eine kohärente Ordnung der Praxis garantiert und damit ein insgesamt gutes Leben ermöglicht.[87] Dazu bedarf es mehrerer Einzelkompetenzen, deren Simulationsmöglichkeiten durch technische Artefakte gegenwärtig mit Blick auf die unterschiedlichen Relationen und Wechselwirkungen zwischen Mensch und Maschine (vgl. Abschnitt 4.3) kontrovers diskutiert werden. Ohne Anspruch auf Vollständigkeit seien dabei die folgenden acht Teilfähigkeiten exemplarisch besonders hervorgehoben:

Erstens braucht es ein *Verständnis* der für unsere Moralsprache konstitutiven evaluativen und deontischen Prädikatoren: Von einem vernünftigen Wesen erwarten wir, dass es über die Fähigkeit verfügt, die Bedeutung der verschiedenen, phänomenologisch gehaltvollen Ausdrücke zur Bezeichnung moralisch relevanter Güter, Werte und Haltungen sowie deontischer Prädikate zur Qualifizierung von Handlungen (wie richtig bzw. falsch) angemessen zu verstehen und situationsadäquat zu gebrauchen.

---

[86] Hier könnte man z. B. aufführen: Bovens, L.; Hartmann, S. (2004): Bayesian Epistemology. Oxford.
[87] Bormann, F.-J. (2021): Ist die praktische Vernunft des Menschen durch KI-Systeme ersetzbar? Zum unterschiedlichen Status von menschlichen Personen und (selbst-)lernenden Maschinen. In: Fritz, A. et al.(2021): Digitalisierung im Gesundheitswesen. Anthropologische und ethische Herausforderungen der Mensch-Maschine-Interaktion. Freiburg, 41-64. S. 48-51.

Zweitens wird ein *Unterscheidungs- und Einfühlungsvermögen* benötigt, um die moralisch relevanten Differenzen zwischen einzelnen moralischen Gütern, Werten, Handlungstypen und Lebensformen möglichst präzise und realitätsnah erfassen sowie anderen Menschen empathisch begegnen zu können.

Drittens muss die Fähigkeit zur *Abwägung* konfligierender Güter und Werte vorliegen: Die praktische Vernunft beinhaltet auch ein deliberatives Vermögen, das immer dann ins Spiel kommt, wenn komplexe Handlungsstrategien entwickelt werden müssen oder mehrere moralisch bedeutsame Gesichtspunkte aufgrund bestimmter ungünstiger Umstände in einer konflikthaften Beziehung zueinanderstehen. Mittels des Vermögens der Güterabwägung vermag die handelnde Person nicht nur zu erkennen, welche Güter in zeitlicher Hinsicht prioritär erstrebt oder gesichert werden müssen, um bestimmte Ziele zu erreichen, sondern welchen Gütern im Konfliktfall der Vorrang zuzuerkennen ist, um ein situativ richtiges Handeln zu ermöglichen.

Viertens bedarf es der Befähigung zum *reflektierten Umgang mit Regeln* unterschiedlicher Reichweite: Die praktische Vernunft schließt auch die Fähigkeit ein, moralische Regeln (wie z. B. Normen und Prinzipien) verstehen, korrekt anwenden und falls nötig auch weiterentwickeln zu können, um Probleme zu lösen und ein realitätsadäquates kohärentes Handeln über längere Zeiträume zu ermöglichen. Obwohl ein Großteil des menschlichen Handelns von Routinen und Konventionen bestimmt wird, gibt es auch vielfältige Herausforderungen und Konfliktsituationen, die durch ein konventionelles oder gar starr deterministisches Regelfolgen allein gerade nicht zu bewältigen sind, sondern Kreativität und einen flexibleren Umgang mit regulatorischen Vorgaben auf der Grundlage eines unvertretbaren Aktes der praktischen Urteilskraft erfordern.[88]

Fünftens wird die Fähigkeit zum *intuitiven Erfassen komplexer Handlungssituationen und Umstände* benötigt: Menschen müssen oft unter großem Zeitdruck weitreichende Entscheidungen treffen und dabei vielfältige Merkmale eines Handlungskontextes berücksichtigen.

[88] In diesem Zusammenhang ist auf einen grundsätzlichen Unterschied zwischen dem algorithmischen gegenüber einem heuristischen Regelverständnis hinzuweisen. In einem algorithmischen Verfahren bilden die Regeln den Auswahlfilter, durch den die Fälle als Kandidaten überprüft und verworfen oder zur Überprüfung im nächsten Schritt angenommen werden. In algorithmischen Verfahren steht damit das Verhältnis von Regel und Fall fest. In einem heuristischen Verfahren werden die Regeln durch die Subsumtion eines Falles pragmatisch und semantisch mit-konstituiert, sodass nicht prä-prozedural feststeht, unter welche Regel der Fall gehört.

Sechstens bedarf es eines *Urteilsvermögens,* mittels dessen Personen in der Lage sind, Entscheidungen zwischen verschiedenen Handlungsalternativen zu treffen und singuläre Handlungskonstellationen bestimmten generellen Handlungstypen zuzuordnen.[89]

Siebtens braucht es die Fähigkeit zur *Begründung* der eigenen moralischen Urteile und der ihnen korrespondierenden Praxis: Die Fähigkeit, Gründe zu geben und zu nehmen (*give and take reasons*) und sich im Urteilen und Handeln daran auszurichten, schließt neben der Bereitschaft zur kritischen Reflexion eigener partikularer Interessen auch die Fähigkeit ein, einen moralischen Standpunkt (*moral point of view*) einzunehmen, also die für die moralische Qualifikation einer Handlung relevanten Gründe aus der Dritte-Person-Perspektive zu beurteilen.

Achtens muss die Fähigkeit zur Affekt- und Impulskontrolle vorliegen, um die jeweils gefällten praktischen Urteile auch handlungswirksam werden zu lassen. Gerade bei der Verfolgung anspruchsvoller Ziele, die vielfältige Vorarbeiten und einen langen Atem verlangen, ist es wichtig, die erforderliche Willensstärke aufzubringen und zumindest solchen spontanen Affekten, Neigungen und Impulsen zu widerstehen, die den langfristigen Erfolg der jeweiligen Bemühungen gefährden oder sogar verunmöglichen können.

Die Unterscheidung der verschiedenen Teilkompetenzen der Vernunft ist für unsere Thematik aus zwei Gründen bedeutsam: Erstens ist es durchaus möglich, dass es partielle Überschneidungen des Kompetenzprofils moderner KI-Systeme mit dem komplexen Phänomen menschlicher Vernunft gibt, was insbesondere im Bereich des Regelfolgens und der Weiterentwicklung vorgegebener Algorithmen der Fall sein dürfte. Zweitens ist zu berücksichtigen, dass die hier genannten, für den Bündelbegriff der „praktischen Vernunft" konstitutiven Fähigkeiten nicht einfach im Sinne isolierter Einzelelemente beziehungslos nebeneinanderstehen. Vielmehr ist von vielfältigen Wechselwirkungen, Rückkopplungen und Bedingungsverhältnissen zwischen ihnen auszugehen. Sie bilden einen integralen Bestandteil einer komplexen menschlichen Natur, die im Sinne einer leib-seelischen Einheit zu verstehen ist. Menschliche Vernunft ist stets als *verleiblichte Vernunft* zu begreifen (vgl. Abschnitt 3.4.3). Nur so ist zu erklären, dass praktische Überlegungen überhaupt handlungswirksam werden können. Zurückzuweisen ist eine Deutung,

[89] In Anspielung auf die durch das richterliche Judiz zu leistende intellektuelle Aufgabe hat Kant für heuristische Verfahren dieser Art den Begriff der *Urteilskraft* geprägt. Das algorithmische Verfahren ordnet Kant der bestimmenden Urteilskraft zu, der Domäne des Verstandes (vgl. Kant, I. (1781): Kritik der reinen Vernunft- B 360f.). Das in den praktischen Disziplinen wie Pädagogik, Jurisprudenz oder Ökonomik zugrunde zulegende Verfahren der Suche nach der angemessenen Passung von Regel und Fall zeichnet Kant als reflektierende Urteilskraft aus, die zur Domäne der praktischen Vernunft gehört z. B. Kant, I. (1790): Kritik der Urteilskraft. AA 385.

+++ Vorabfassung! Wird durch die lektorierte Fassung ersetzt. +++

die versucht, vernünftige Vollzüge aus einer rein individualistischen Perspektive zu rekonstruieren. Da jeder Mensch Teil einer sozialen Mitwelt und kulturellen Umgebung ist, die sich nachhaltig auf seine Sozialisation auswirkt, müssen auch überindividuelle kulturelle Faktoren in die Deutung der praktischen Vernunft einbezogen werden. Ein angemessenes Verständnis insbesondere des praktischen Vernunftgebrauchs ist eng mit unserem basalen Selbstverständnis als handlungsfähige Personen verbunden. Da technische Artefakte in immer neuen Formen in die Handlungswelt der Menschen integriert werden, mit Menschen interagieren oder sogar Teilfunktionen menschlichen Handelns übernehmen, ist es wichtig, zunächst den Handlungs- und Verantwortungsbegriff zu klären.

3.3 Handlung und Verantwortung

3.3.1 Handlung

Auch wenn im Alltag gelegentlich alle möglichen Ereignisse als Handlungen bezeichnet werden, fassen die Normwissenschaften Ethik und Jurisprudenz und auch die Psychologie den Handlungsbegriff oft enger.[90] Dabei wird angenommen, dass Menschen in der Lage sind, aktiv, zweckgerichtet und kontrolliert auf die Umwelt einzuwirken und dadurch Veränderungen zu verursachen. Das bedeutet, dass nicht jedes menschliche Tun, das auf die Umwelt einwirkt, als Handlung zu verstehen ist, sondern nur solches, das zweckgerichtet, beabsichtigt und kontrolliert ist.[91] Unterstellt man, dass Maschinen nicht zweckgerichtet operieren, also keine Absichten haben, dann ist die Zuschreibung von Handlungen in Bezug auf Maschinen in diesem engen Sinne nicht möglich.

Für den Menschen, der im engen Sinn handelt, hat sich auf dem Hintergrund einer langen zurückreichenden Begriffsgeschichte[92] im Rahmen einer umfassenden Theorie praktischer Vernunft der Begriff der *Person* eingebürgert.[93] Personen sind Akteure, die Verantwortung für ihr

---

[90] In der Psychologie auch, vgl. z. B. die verschiedenen (etablierten) Handlungsphasenmodelle.
[91] Hier wird eine Form der Handlungserklärung unterstellt, die die Warumfrage (im Sinne von „Warum hast du das getan?“) durch Angabe der Absicht bzw. Zwecke der Handlung beantwortet sieht (intentionale, teleologische Handlungserklärung, „Intentionalismus“). Ihr steht eine Form der Handlungserklärung gegenüber, die die Angabe der die Handlung auslösenden Ursachen als Antwort vorsieht (kausale Handlungserklärung, „Kausalismus“). Ob eine Handlung intentionalistisch oder kausalistisch erklärt werden sollte, ist keine Frage der Wahrheit / Falschheit der Erklärung, sondern eine Frage der Kontextadäquatheit der Handlungsdeutung. Für die Normwissenschaften steht die Frage der Absicht bzw. der Zweck der Handlung im Vordergrund der Betrachtung, ohne dass die Möglichkeit einer kausalen Handlungserklärung in Abrede gestellt wird. Vgl. Horn, C.; Löhrer, G. (Hg.) (2010): Gründe und Zwecke. Texte zur aktuellen Handlungstheorie. Berlin.
[92] Fuhrmann, M. et al. (1989): Person. In: Ritter, J.; Gründer, K. (Hg.): Historisches Wörterbuch der Philosophie. Band 7: P-Q. Basel, Sp. 269-338 (DOI: 10.24894/HWPh.5339).
[93] Quante, M. (2007): Person. Berlin.

Verhalten tragen, die die kognitiven Bedingungen erfüllen, um Handlungsoptionen zu erkennen und die Gründe deliberieren können, also über ein hinreichendes Maß theoretischer und praktischer Vernunft verfügen.

Der Handlungsbegriff ist auch deswegen so bedeutsam in der Diskussion um maschinelle Fertigkeiten und Künstliche Intelligenz, da dieser Diskurs sich etwa seit der Jahrtausendwende von der Konzentration auf möglicherweise kognitive Kompetenzen abgewandt hat und inzwischen zunehmend auf praktische Kompetenzen konzentriert.[94]Nicht in welcher Weise, wenn überhaupt, Maschinen denken, sondern in welchem Sinne Maschinen handeln können, steht verstärkt im Vordergrund. Diese Verschiebung der Aufmerksamkeit geht Hand in Hand mit technischen Entwicklungen hin zu maschinellen Systemen, die nicht lediglich auf hohe Informationsverarbeitungskapazität setzen, um menschliches Handeln zu unterstützen, sondern jenes teilweise gar ersetzen können oder sollen.

Automatisierte oder algorithmische Entscheidungssysteme (ADM-Systeme, abgekürzt von *automated decision making* oder *algorithmic decision making*) zum Beispiel erstellen auf Basis von Berechnungen Prognosen darüber, wie geeignet eine sich bewerbende Person für eine Stelle ist oder mit welcher Wahrscheinlichkeit Menschen Kredite zurückzahlen oder straffällig werden (vgl. Kapitel 8). Auch wenn diese Systeme oftmals nur der Unterstützung der menschlichen Entscheidung dienen, so können Entscheidungen auch komplett an jene delegiert werden. Damit stellt sich die Frage, in welchem Sinne solche maschinellen Vollzüge außerhalb des obigen engen Handlungsbegriffs doch in bestimmten Kontexten als Handlungen in einem weiteren Sinne wahrgenommen werden können oder berücksichtigt werden müssen. Daran anknüpfend gibt es einen Diskurs, ob und inwieweit zunehmend eigenständige, das heißt ohne menschliches Zutun funktionierende maschinelle Systeme als „Agenten" in der Folge für ihr „Handeln" verantwortlich gemacht werden können, etwa mit Blick auf Fragen der Haftung (vgl. Abschnitt 2.2.5).[95]

Den folgenden Überlegungen liegt mit Blick auf die oft schillernde Verwendung zentraler ethisch relevanter Begriffe wie *Handlung* oder *Verantwortung* im Kontext der zeitgenössischen KI-Debatte die Annahme zugrunde, dass wenigstens drei Klassen von Entitäten terminologisch

[94] Dreyfus, H. L. (1992): What Computers still can't do. 2., überar. Auflage. Cambridge (MA), London.
[95] Vgl. dazu etwa Hilgendorf, E. (2014): Robotik im Kontext von Recht und Moral. Baden-Baden; Hilgendorf, E (2020): Digitalisierung, Virtualisierung und das Recht. In: Kasprowicz, D.; Rieger, S. (Hg.): Handbuch Virtualität. Wiesbaden, 405–424; Ebers, M. et al. (Hg.) (2020): Künstliche Intelligenz und Robotik. München.

klar gegeneinander abzugrenzen sind. Dies wären erstens Pflanzen und Tiere, die zwar in vielfältiger Weise auf ihre Umwelt reagieren können, in ihrem jeweiligen Repertoire aber doch so begrenzt sind, dass der ausgeführte enge Handlungsbegriff auf sie nicht anwendbar ist.[96] Zweitens sind Menschen in dem Maße im engen Sinn als handlungsfähig zu bezeichnen, als sie dazu imstande sind, absichtlich Veränderungen zu bewirken, wobei solche Handlungen nicht nur Taten, sondern auch deren bewusstes und absichtliches Unterlassen umfassen können.[97] Drittens gibt es technische Artefakte unterschiedlicher Komplexitätsgrade, deren jeweilige Vollzüge oder Operationen zwar Veränderungen in der Welt bewirken und flexibel mit anspruchsvollen Herausforderungen der menschlichen Lebenswelt umgehen können.[98] Da sie diese Veränderungen aber nicht absichtlich herbeiführen, haben sie selbige daher auch nicht in einem moralischen und rechtlichen Sinne zu verantworten (vgl. Abschnitt 3.3.2).

Auch wenn es zwischen diesen drei Klassen von Entitäten unterschiedliche Arten von Wechselwirkungen (vgl. Abschnitt 4.3) geben kann, scheint es sinnvoll, den Handlungsbegriff im engen Sinne Menschen vorzubehalten, um inflationären Ausweitungen des Akteur-Status zu vermeiden und konzeptionelle Grenzziehungen zu ermöglichen. Entscheidend ist demnach das Konzept der Handlungsurheberschaft bzw. Autorschaft, das auf die universelle menschliche Handlungserfahrung verweist, sich selbst und andere im Hinblick auf bestimmte Ereignisse und Zustände als Urheber anzusehen.[99] Die Fähigkeit zur Handlungsurheberschaft kann als Grundlage von Autonomie betrachtet werden, also dafür, dass handelnde Menschen ihre Handlungen nach Maximen ausrichten können, die sie sich selbst setzen. Diese Konzeption schließt jedoch nicht aus, dass Handlungen mitunter auch durch Befolgen von Autoritäten und Traditionen aus-

[96] Ausnahmen werden für einige hochentwickelte Tiere diskutiert, darunter z. B. Primaten und Rabenvögel; vgl. etwa Huber, L. (2021): Das rationale Tier: Eine kognitionsbiologische Spurensuche. Berlin.
[97] Für Unterlassungen, nicht aber für Nicht-Handeln, kann man getadelt oder verurteilt werden. Das Unterlassen bezieht sich demgemäß auf ein bestimmtes Handlungsschema (z. B. des Helfens) und ist nicht nur die unbestimmte Negation der Ausführung eines Handlungsschemas. Unterlassungshandlungen sind somit „Ereignisse in der Welt“ und als solche ohne Weiteres mögliche Ursachen von Wirkungen (Folgen) in der Welt. Siehe etwa Roxin, C.; Greco, L. (2020): Strafrecht, Allgemeiner Teil. Band I: Grundlagen. Der Aufbau der Verbrechenslehre. 5. Auflage. München, 335 ff.; Bottek, C. (2014): Unterlassungen und ihre Folgen. Handlungs- und kausaltheoretische Überlegungen. Tübingen.
[98] Rammert, W.; Schulz-Schaeffer, I. (2002): Technik und Handeln – wenn soziales Handeln sich auf menschliches Verhalten und technische Artefakte verteilt. Berlin. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-1107 [04.01.2023], 11-64.
[99] Deutscher Ethikrat (2017). Big Data und Gesundheit – Datensouveränität als informationelle Freiheitsgestaltung. Berlin, 175-178; Nida-Rümelin, J. (2020): Eine Theorie praktischer Vernunft. Berlin, 376-408.

geführt werden. Wo dies geschieht, setzt das Konzept der Handlungsurheberschaft jedoch voraus, dass Menschen ihr eigenes Dasein in ein Verhältnis zu solchen Bestimmungen setzen können, etwa durch Überwindung, Widerstand oder Nachgeben.

Aus einer Handlung können neben den beabsichtigten Folgen auch nicht beabsichtigte, aber der handelnden Person erkennbare Folgen erwachsen. Auf diese Weise erscheint es möglich, auch fahrlässiges Tun zu erfassen, das im Kontext von KI eine große Rolle spielt.[100] Des Weiteren finden Handlungen umgeben von anderen raumzeitlichen Ereignissen statt, die als Umstände der Handlung für deren moralische und rechtliche Bewertung von Bedeutung sein können. Auch werden Handlungen zwar methodisch primär individuellen Akteuren zugesprochen; das schließt aber kollektive Handlungen nicht aus, bei denen die einzelnen Personen von vornherein in einem Kontext der Koordination agieren, ihre Handlungen auf Kooperation bezogen sind und durch Kommunikation gestützt werden.

Der hier verwendete, eng gefasste Handlungsbegriff schließt nicht aus, dass Technologie erheblichen Einfluss auf menschliches Handeln oder die menschliche Handlungserfahrung haben kann. Technik beeinflusst und verändert Gesellschaft; und gleichzeitig beeinflusst Gesellschaft die Technikentwicklung und den Technikeinsatz. Gerade die in den letzten Jahren stark zunehmende Durchdringung der menschlichen Lebenswelt mit informationstechnisch immer leistungsfähigeren Maschinen, die mit anspruchsvoller Sensorik und Motorik sowie vernetzt arbeiten, führt zu hybriden, sozio-technischen Konstellationen, in denen Menschen und Maschinen eng verwoben sind und auf komplexe Weise interagieren. Dies kann das Verhalten und die Handlungen von Menschen stark beeinflussen und ihre individuellen Freiheitsspielräume und Kontrollmöglichkeiten einschränken. Zudem können fortgeschrittene und mit flexiblen, selbstlernenden Algorithmen arbeitende maschinelle Systeme menschliches Tun zum Teil so gut imitieren, dass sie wie intentionales menschliches Handeln erscheinen, was weitere ethische Fragen aufwirft (vgl. Kapitel 4).

Auch mit Blick auf diese sozio-technische Komplexität bis hin zu Unterscheidungsschwierigkeiten erscheint es sinnvoll, an einem engen Handlungsbegriff, der an das zentrale Kriterium der Intentionalität gebunden ist, festzuhalten. Dieses Intentionalitätskriterium ist zudem entscheidend für die Möglichkeit der Zuschreibung von Verantwortung im Kontext von Mensch-

[100] Ein Beispiel liefert etwa der Bereich des autonomen Fahrens: Unterläuft dem Programmierer ein Fehler, der später zur Schädigung von Verkehrsteilnehmern führt und war dies für ihn vorhersehbar und vermeidbar sowie aus Gründen des überwiegenden Schutzes der Interessen der anderen Personen auch von Rechts wegen zu vermeiden, liegt hierin ein fahrlässiges Fehlverhalten, für das er zur Verantwortung zu ziehen ist. Hevelke, A.; Nida-Rümelin, J. (2015): Responsibility for Crashes of Autonomous Vehicles: An Ethical Analysis. In: Science and Engineering Ethics 21, 619-630.

Maschine-Interaktionen in zunehmend komplexer sozio-technischer Vernetzung. Im Hinblick auf digitale Techniken handeln Menschen nicht jederzeit in vollem Maße autonom, sondern verstehen oft weder die technischen Zusammenhänge, noch haben sie immer umfassende Informationen oder hinreichende Wahlfreiheit, um bewusste Entscheidungen im Umgang mit der Technik treffen zu können. Solche Konstellationen können vielfältig ethische Bedeutung entfalten, wenn es um Fragen der Verantwortung geht.

3.3.2 Verantwortung

Fragen der Verantwortung und Verantwortlichkeiten knüpfen an den Handlungsbegriff an. Die Rolle menschlicher Verantwortung für die kausale Verursachung von Handlungsfolgen kann in vielfacher Weise thematisiert, reflektiert, diskutiert und modifiziert werden. Bei der Verantwortung für die Resultate von Zuschreibungshandlungen in sozialen Kontexten geht es vor allem darum zu klären, welche Verantwortung von welchen Akteuren übernommen werden *soll*, um einer zunehmenden Verantwortungsdiffusion entgegenzuwirken.[101] Verantwortungszuschreibung und -verteilung erfolgen zu dem Zweck, Praxisfelder wie den Straßenverkehr, den Schulbetrieb oder den Umgang mit KI in verschiedenen Anwendungsbereichen so zu strukturieren und gegebenenfalls rechtlich zu regulieren, dass sich dadurch eine möglichst „gute Praxis“ entfalten kann.

In den Zuschreibungen wird der Kreis der verantwortungsfähigen Individuen abgegrenzt, der Stellenwert der kausalen Verursachung geregelt und es werden Kriterien festgelegt, welche Voraussetzungen Menschen erfüllen müssen, um ihnen Verantwortung zuschreiben zu können. Somit stellt sich die Frage, wer für was direkt oder indirekt Verantwortung übernehmen kann oder soll, wenn Individuen, Gruppen und Institutionen aus und in verschiedenen Bereichen wie im Privatleben, in der Forschung, Wirtschaft und Politik sowohl miteinander als auch mit maschinellen Systemen und insbesondere KI-Systemen zusammenwirken.

Verantwortung kann als Konzept einer vielfachen Relation rekonstruiert werden. Im Kontext dieser Stellungnahme erscheint eine fünfstellige Relation angemessen:[102] Wer ist für was, gegenüber wem, vor wem und unter welcher Norm verantwortlich? Allerdings sprechen wir auch

---

[101] Vgl. Grunwald, A. (2021): Der homo responsibilis. Nachdenklicher Gang durch den Garten aktueller Erzählungen. In: ders. (Hg.): Wer bist du, Mensch? Transformationen menschlicher Selbstverständnisse im wissenschaftlich-technischen Fortschritt. Freiburg, 216-239.

[102] Loh, J. (2016): Strukturen und Relata der Verantwortung. In: Heidbrink, L. et al.(Hg.): Handbuch Verantwortung. Kiel, Berlin, Wien.

von einer verantwortlichen Person; in diesem Falle ist der Begriff einstellig, und der Zusammenhang zwischen dem einstelligen moralischen Begriff der verantwortlichen Person und den unterschiedlichen, meist mehrstelligen Kriterien der Verantwortungszuschreibung ist eine eigene philosophische und rechtstheoretische Thematik. Ohne die genuine individuelle moralische Verantwortung wären auch die weitergehenden Ausdifferenzierungen gegenstandslos.[103]

Demnach kann man erstens ganz grundsätzlich beim Verantwortungssubjekt (*wer*) ansetzen, das Verantwortung übernehmen kann. Verantwortungssubjekte tragen Verantwortung, als Einzelperson oder Mitglieder eines Kollektivs, beispielsweise einer Institution. Davon zu unterscheiden ist zweitens das Verantwortungsobjekt (*was*), für das Verantwortung übernommen wird, zum Beispiel Handlungen sowie deren Gründe und Folgen.[104] Als drittes Relationselement werden die vom Handeln des Verantwortungssubjektes (direkt oder indirekt) Betroffenen benannt (*gegenüber wem*). Das vierte Relationselement bildet die Instanz, vor der die Verantwortung übernommen wird (*vor wem).* Das Gewissen als Inbegriff der praktischen Vernunft, andere Personen oder auch eine staatliche Rechtsgemeinschaft sind hier denkbar. Für eine normative Stellungnahme ist zudem ein fünftes Relationselement bedeutsam, nämlich die Regel oder das Prinzip, dem eine verantwortliche Praxis gerecht werden sollte, zum Beispiel das Prinzip, andere nicht zu schädigen (*unter welcher Norm).*[105]

Vor allem das dritte Relationselement – die Betroffenen – ist nicht immer leicht einzugrenzen. Ungeachtet dessen ist deren Berücksichtigung und Einbezug gerade in Anbetracht des steigenden Einsatzes von KI-Systemen in vielen Gesellschafts- und Lebensbereichen von zentraler Bedeutung. An dieses Desiderat knüpfen sich auch Forderungen nach Transparenz und Nachvollziehbarkeit, welche eine Voraussetzung für die Beteiligung und Berücksichtigung von Betroffenen darstellen.

---

[103] Nida-Rümelin, J. (2011): Verantwortung. Stuttgart.

[104] Personen können auch für ein Unterlassen verantwortlich sein. Auch durch ein Unterlassen wird ein Ereignis in der Welt verursacht. Die Verantwortung kann aus einer Sonderbeziehung des Unterlassenden im Hinblick auf das dadurch beeinträchtigte Rechtsgut erwachsen („Garantenstellung"). Siehe Freund, G. (1992): Erfolgsdelikt und Unterlassen. Zu den Legitimationsbedingungen von Schuldspruch und Strafe. Köln, München. Ein „Jedermanns-Unterlassen", für das es auf keine solche Sonderbeziehung ankommt, lässt sich darüber hinaus auf allgemeine Solidaritätspflichten stützen, siehe Frisch, W. (2016): Strafrecht und Solidarität – Zugleich zu Notstand und unterlassener Hilfeleistung. In: Goltdammer's Archiv für Strafrecht 163 (3), 121-137.

[105] Damit bildet der Verstoß gegen eine Norm den Anknüpfungspunkt von Verantwortung. Normen schränken die Freiheit des Einzelnen ein und bedürfen daher der Legitimation. Im Recht kann insoweit auf den Grundsatz der Verhältnismäßigkeit (i.w.S.) zurückgegriffen werden. Zu diesem Grundsatz siehe Dechsling, R. (1989): Das Verhältnismäßigkeitsgebot: Eine Bestandsaufnahme der Literatur zur Verhältnismäßigkeit staatlichen Handelns. München.

Über diese Konstellation hinaus ist in der Verantwortungsdiskussion zum wissenschaftlich-technischen Fortschritt die epistemologische Dimension zu bedenken. Denn Handlungsfolgen sind oft nur unter hohen und nicht eliminierbaren Unsicherheiten des Wissens antizipierbar.[106] Verantwortungszuschreibung muss daher die Dimension des Handelns unter Unsicherheit und damit die Risikothematik[107] berücksichtigen. Dies lässt auch noch einmal zwischen retrospektiver und prospektiver Verantwortung unterscheiden. Beim Blick auf vergangene Handlungen kann eine nachträgliche Verantwortungsübernahme unter Umständen angezeigt sein. Vor allem die prospektive Verantwortung unterliegt den gerade konstatierten Unsicherheiten.

Um Verantwortung im Zusammenspiel von Menschen und maschinellen Systemen näher zu betrachten, kann das fünfstellige Verantwortungskonzept technikspezifisch gerahmt werden. Ausgangspunkt ist hier, dass eine Verantwortungsübernahme (als *Verantwortungssubjekt*) nur Personen als verantwortlichen Wesen möglich ist, beispielsweise den Individuen, die Technik entwickeln und herstellen, die ihren Einsatz etwa in der Politik oder Unternehmen ermöglichen und fördern, oder denjenigen, die Technologien einsetzen. Das *Verantwortungsobjekt* ist dann je nach Rolle der Verantwortungssubjekte und ihrer Handlungen zu beschreiben: zum Beispiel Planen, Erfinden, Entwickeln oder Anwenden. Zu den *Betroffenen* können sowohl die von dem technischen Angebot direkt angesprochenen Personen(gruppen) gehören, zum Beispiel Angestellte in einem Krankenhaus, die mithilfe KI-gestützter Software Entscheidungen treffen, als auch weitere Personen wie zum Beispiel diejenigen, die auf Grundlage solcher Entscheidungen Diagnosen, Therapieempfehlungen oder sonstigen medizinischen Rat erhalten. Relevante *Instanzen* und relevante *Normen* sind hierbei verknüpft. Rechtliche Verantwortung besteht in letzter Instanz gegenüber der staatlich verfassten Gemeinschaft.

Moralische Verantwortung können nur natürliche Personen übernehmen, insofern sie über Handlungsfähigkeit verfügen, das heißt in der Lage sind, aktiv, zweckgerichtet und kontrolliert auf die Umwelt einzuwirken und dadurch Veränderungen zu verursachen. Träfe dies auch auf Maschinen zu, wären auch diese verantwortungsfähig. Dann müsste Maschinen der Personenstatus zugeschrieben werden, was jedoch weder aktuell noch angesichts der in absehbarer Zukunft erwartbaren qualitativen Entwicklungen maschineller Systeme angemessen wäre. Verantwortung kann daher nicht direkt von maschinellen Systemen übernommen werden, sondern nur von den Menschen, die in je unterschiedlichen Funktionen hinter diesen Systemen stehen,

---

[106] Grunwald, A. (2013): Modes of orientation provided by futures studies: making sense of diversity and divergence. In: European Journal of Futures Research 15:30 (DOI 10.1007/s40309-013-0030-5).
[107] Nida-Rümelin, J. et al. (2012): Risikoethik. Berlin.

gegebenenfalls im Rahmen institutioneller Verantwortung. Auch wenn ein technisches System eingesetzt wird, um im Rahmen einer automatisierten Datenauswertung Schlussfolgerungen wie die Gewährung eines Kredites anzuwenden, ist es die Verantwortung des Menschen, dieses System in einer ethisch vertretbaren Weise zu entwickeln und einzusetzen.[108]

Wer nun konkret als Verantwortungsträger fungiert, kann mit dem Konzept der *Multiakteurs-verantwortung* umrissen werden.[109] Kommt es bereits zu facettenreichen Verantwortungsgefügen, wenn man von nur drei prinzipiellen Ebenen möglicher Verantwortungszuschreibung ausgeht – Individuen, Organisationen und Staat –, so entsteht ein noch komplexeres Bild, wenn man Wechselwirkungen zwischen verschiedenen Akteuren aus diesen drei Ebenen berücksichtigt. Dies gilt erst recht, wenn diese Interaktionen zumindest teilweise von algorithmischen Systemen gestützt oder vermittelt werden, die mitunter für andere Beteiligte selbst autonom und kaum durchschaubar zu agieren scheinen.

Hier stellt sich die Frage, wie Verantwortung sinnvoll zwischen unterschiedlichen Beteiligten geteilt werden kann, zum Beispiel zwischen denjenigen, die maschinelle Systeme konzipieren und entwickeln, die ihre Nutzung beauftragen oder vorantreiben, die in Nutzungsprozesse oder ihre Überwachung direkt eingebunden sind, die Ergebnisse solcher Prozesse verwenden oder von ihnen direkt oder indirekt betroffen sind oder die ihre Auswirkungen auf unterschiedlichen gesellschaftlichen Ebenen beobachten und eventuell regulierend eingreifen können. In Anlehnung an das Konzept der Datensouveränität[110] ist die geeignete Gestaltung von Multiakteursverantwortung demnach zentral für eine angemessene informationelle Freiheitsgestaltung, die den Chancen und Risiken einer zunehmend digital vernetzten und algorithmisch gestützten Welt gerecht wird. Eine solche Freiheitsgestaltung kann nur dann verantwortlich sein, wenn sie sich gleichzeitig an den gesellschaftlichen Anforderungen von Solidarität und Gerechtigkeit orientiert.

Auch für die Diskursführung über die zahlreichen Wechselwirkungen von Menschen und Maschinen und die gesellschaftlichen Auswirkungen einer zunehmenden Durchdringung der menschlichen Gesellschaft mit algorithmischen Systemen muss Verantwortung übernommen

---

[108] Datenethikkommission der Bundesregierung (2019): Gutachten der Datenethikkommission. Berlin. https://www.bmi.bund.de/SharedDocs/downloads/DE/publikationen/themen/it-digitalpolitik/gutachten-datenethikkommission.pdf;jsessionid=4A012DD4E717D4E3FA62DD51238229C3.1_cid295?__blob=publicationFile&v=7 [10.02.2023].
[109] Deutscher Ethikrat (2017): Big Data und Gesundheit – Datensouveränität als informationelle Freiheitsgestaltung. Berlin, 249 f.
[110] Deutscher Ethikrat (2017): Big Data und Gesundheit – Datensouveränität als informationelle Freiheitsgestaltung. Berlin, 252 f.

werden. Warnungen vor unkritischem Vertrauen in maschinelle Systeme, insbesondere im Falle Künstlicher Intelligenz, sollten einen Platz haben und sind Ausdruck wahrgenommener Verantwortung. Ebenso sind die Auswahl und Gewichtung bestimmter normativer Kriterien und Prinzipien im Diskurs zur Ethik von maschinellen Systemen, Algorithmen und KI selbst Gegenstand von Kontroversen.[111] Wer hat die Deutungshoheit, Werte und Normen, die im Umgang mit KI relevant sind, zu bestimmen? Wer prüft, welche Betroffenen vornehmlich in den Blick genommen werden oder wie die Lasten und Nutzen bestimmter Anwendungen verteilt sind? Es stellt sich also allgemein die Frage, wer für das fünfte Relationselement, die Bestimmung normativer präskriptiver Prinzipien, zuständig ist.

3.4 Anthropologische Aspekte des Mensch-Maschine-Verhältnisses

3.4.1 Philosophische Grundbestimmung des Menschseins

Handlung, Vernunft und Verantwortung stehen im Zentrum humanistischer[112] Philosophie. Menschen sind befähigt zur Handlungsurheberschaft und somit zur Autorschaft ihres Lebens. Sie sind frei und tragen daher Verantwortung für die Gestaltung ihres Handelns. Freiheit und Verantwortung sind zwei einander wechselseitig bedingende Aspekte menschlicher Autorschaft. Autorschaft ist wiederum an Vernunftfähigkeit gebunden. Die strafrechtlichen Kriterien für Schuldfähigkeit konvergieren mit der lebensweltlichen Praxis moralischer Zuschreibungen. Personen sind jedenfalls in wichtigen sozialen Kontexten moralisch verantwortlich. Das heißt, man erwartet von ihnen, dass sie zurechnungsfähig handeln und urteilen.[113]

Diese Trias aus Vernunft, Freiheit und Verantwortung prägt heute sowohl die lebensweltliche Moral als auch die Rechtsordnung in hohem Maße. Im Mittelpunkt steht dabei das Phänomen

---

[111] Jobin, A. et al. (2019): The global landscape of AI ethics guidelines. In Nature Machine Intelligence 1, 389-399 (DOI: 10.1038/s42256-019-0088-2); Rudschies, C. et al. (2021): Value Pluralism in the AI Ethics Debate – Different Actors, Different Priorities. In: International Review of Information Ethics 29. https://informationethics.ca/index.php/irie/article/view/419/396 [04.01.2023].

[112] Die unterschiedlichen Verwendungsweisen des Humanismusbegriffs in der Philosophie stimmen in einigen normativen Kernelementen überein, die zur Klärung anthropologischer Aspekte des Mensch-Maschine-Verhältnisses wichtig sind und nachfolgend näher entfaltet werden.

[113] In der Technikanthropologie, die sich mit anthropologischen Aspekten des Technischen und der Technik befasst, vor allem mit Mensch/Technik- oder Mensch/Maschine-Konstellationen, wird eine Vielzahl auch anderer Perspektiven verfolgt (Heßler, M.; Liggieri, K. (2020): Technikanthropologie. Handbuch für Wissenschaft und Studium. Baden-Baden). Hierzu gehören etwa homo faber und homo creator, trans- und posthumanistische Positionen sowie die Akteur-Netzwerk-Theorie. Die enge Verbindung ethischer Fragen zu den Konzepten von Freiheit und Verantwortung impliziert, diese in den Betrachtungen nicht eigens zu berücksichtigen, sondern die humanistische Perspektive in die Mitte zu stellen.

der Affektion durch Gründe. Praktische Gründe sprechen *für* Handlungen, sie sind per se normativ, nicht erst über den Umweg eigener Wünsche. Ein Grund spricht dafür, das zu tun, was diesen Grund erfüllt, wenn nicht andere Gründe dem entgegenstehen.[114] Theoretische Gründe sprechen *für* Überzeugungen; auch diese sind normativ. In der Regel gibt es Gründe das eine zu tun und das andere zu lassen, die gegeneinander abgewogen werden müssen. Der Konflikt von Gründen zwingt dann zur Abwägung und zur Systematisierung dieser Abwägung in Gestalt ethischer Theoriebildung.

Die menschliche Lebensform ist von reaktiven Einstellungen und moralischen Gefühlen geprägt, die von normativen Gründen begleitet sind. Wir vergeben einer Person, die uns Unrecht getan hat, wenn wir den Eindruck haben, sie habe das Unrechte ihres Tuns eingesehen und werde diese Praxis nicht fortsetzen. Wir sind dankbar, wenn wir meinen, dass eine Person etwas Gutes getan hat, ohne daraus Vorteile zu ziehen, wir nehmen etwas übel, nur dann, wenn wir die betreffende Person für voll zurechnungsfähig und in ihrem Handeln frei halten.[115] Die verobjektivierende Einstellung gegenüber anderen Menschen, die diese zum bloßen Gegenstand der Beeinflussung macht, sie gewissermaßen zu einem Teil der Umwelt degradiert, lässt sich nur für ganz spezifische Situationen – wenn überhaupt – durchhalten. Aber wenn diese verobjektivierende Einstellung ohne moralische Empfindungen zur allgemeinen Praxis würde, gäbe es die menschliche Lebensform nicht mehr. Diese ist gerade dadurch geprägt, dass wir ein Verhalten übelnehmen, wenn es uns inakzeptabel erscheint, dass wir zum Beispiel in der Lage sind zu verzeihen, wenn wir dafür Gründe haben, oder, dass wir Dankbarkeit empfinden.

Freiheit kommt insofern ins Spiel, als wir rationaliter bestimmte moralische Gefühle und reaktive Einstellungen zurückstellen, wenn wir erfahren, dass die betreffende Person in ihrem Handeln nicht frei war, was immer die Ursachen dieser Unfreiheit sind, wie beispielsweise äußerer Zwang, psychische Erkrankung oder überwältigende Angst. Diese Praxis der Zuschreibung von Freiheit und Verantwortung ist essenziell für die Grundlegung moralischer Beurteilung wie auch für moralische Gefühle und reaktive Einstellungen, und daher ist es ausgeschlossen, diese aufzugeben, wie überzeugend auch immer wissenschaftliche Theorien, die prima facie dagegen sprechen, sein mögen.[116] Es macht unsere Zugehörigkeit zur menschlichen Lebensform aus,

---

[114] Scanlon, T.M (1998): What We Owe to Each Other. Cambridge. Scanlon, T.M (2014): Being Realistic about Reasons, Oxford. Halbig, C. (2007): Praktische Gründe und die Realität der Moral. Frankfurt a. M.
[115] Vgl. den einflussreichen Ansatz von Strawson, P. F. (1964): Freedom and Resentment and other Essays. London.
[116] Siehe beispielsweise die Debatte rund um die Experimente von Libet. Libet, B. (2004): Mind Time. The Temporal Factor in Human Consciousness. Cambridge (MA), London.

dass solche moralischen Beurteilungen, Gefühle und Einstellungen unsere soziale Praxis prägen. Die Normen von Moral und Recht sind ohne die Annahme menschlicher Verantwortlichkeit und damit Freiheitsfähigkeit und Vernunftfähigkeit unbegründet. Sie würden in bloße Instrumente der Verhaltenssteuerung transformiert[117] und paradoxerweise wäre es gerade diese Transformation, die ihre Wirksamkeit für die Verhaltenssteuerung zugleich gefährden würde.

Wenn menschliche Freiheit im Sinne des Anderskönnens bestritten wird, kann an der Verantwortlichkeit menschlicher Personen nicht festgehalten werden.[118] In § 20 StGB werden die praktischen und epistemischen Bedingungen von Schuldfähigkeit dargestellt, die ohne diese anthropologischen Voraussetzungen von Freiheit und Vernunft nicht aufrechtzuerhalten wären. Ohne die Fähigkeit, sich anders entscheiden zu können, gibt es keine Handlungsurheberschaft und keine Verantwortung.

Obwohl die humanistische Perspektive nicht nur die lebensweltliche Normativität, sondern auch die juridische Deliberation von den Menschenrechten bis zum Strafrecht imprägniert und das kulturelle Fundament demokratischer Ordnungen ausmacht, wird sie doch immer wieder infrage gestellt. Zwei jüngere Formen der Kritik, die sich teilweise überlagern, stützen sich einerseits auf neurowissenschaftliche Begriffe und Paradigmen sowie andererseits auf solche aus Debatten um das Thema Künstliche Intelligenz. In den Neurowissenschaften wurden bestimmte empirische Studien, nach denen zum Beispiel das motorische Zentrum des Gehirns schon mit der Vorbereitung einer Bewegung beginnt, bevor man sich bewusst für die Ausführung der Bewegung entschieden hat[119], als Beleg dafür interpretiert, dass es Freiheit und damit menschliche Verantwortlichkeit nicht gebe. Stattdessen handele es sich dabei lediglich um – möglicherweise sozial nützliche – Illusionen. Tatsächlich lassen solche Befunde jedoch unterschiedliche Interpretationen zu und eignen sich nicht als Widerlegung menschlicher Freiheit und Verantwortlichkeit. Auch wenn alle mentalen Prozesse neurophysiologisch realisiert sind, sprechen die empirischen Befunde aus den Neurowissenschaften nicht gegen das Phänomen

---

[117] Vgl. dazu Schlick, M. (1930): Fragen der Ethik. Wien. der exemplarisch für diese anti-humanistische Auffassung steht.

[118] Es gibt allerdings in der zeitgenössischen Philosophie auch die Auffassung, Verantwortlichkeit könnte auch ohne die Bedingung der Freiheit postuliert werden, besonders prominent bei Harry Frankfurt. Kane, R. (Hg.) (2011): The Oxford Handbook of Free Will. 2. Auflage. Oxford (DOI: 10.1093/oxfordhb/9780195399691.001.0001), hierbei das Kapitel 5 „Moral Responsibility, Alternative Possibilities, And Frankfurt-Type Examples“. Siehe auch Frankfurt, H. (1971): Freedom of the Will and the Concept of a Person. In: The Journal of Philosophy 68 (1), 5-20 (DOI: 10.2307/2024717).

[119] Libet, B. (2004): Mind Time. The Temporal Factor in Human Consciousness. Cambridge (MA), London.

normativer Gründe und ihrer Rolle für menschliche Handlungsmotivation, da wir es hier mit zwei Sprachebenen zu tun haben, die sich wechselseitig nicht in die Quere kommen können.[120]

Die zweite, von der KI-Debatte inspirierte Kritik der humanistischen Anthropologie changiert zwischen einer Überwindung des Menschen in Gestalt des Transhumanismus, der mit neuen Mensch-Maschinen-Symbiosen die Reichweite menschlichen Wirkens in neue Dimensionen heben möchte und einem Maschinenparadigma, das den menschlichen Geist auf das Modell eines algorithmischen Systems reduziert. Gerade Letzteres entfaltet besondere Relevanz im Kontext dieser Stellungnahme, da es großen Einfluss auf die Interpretation der Wechselwirkungen zwischen Mensch und Maschine und deren Rückwirkungen auf das menschliche Selbstverständnis hat.

3.4.2 Der Mensch als Maschine – die Maschine als Mensch?

Der Mensch als Maschine ist eine alte Metapher, deren Ursprünge bis in die Frühe Neuzeit zurückreichen. Der mechanistische Materialismus des rationalistischen Zeitalters lässt die Welt als Uhrwerk erscheinen und den Menschen als Rädchen im Getriebe. Der große Uhrmacher ist dann der Schöpfer, der dafür gesorgt hat, dass nichts dem Zufall überlassen ist und ein Rädchen ins andere greift. Für menschliche Freiheit, Verantwortung und Vernunft ist in diesem Bild kein Platz.

Die vielleicht aktuell größte Herausforderung für das humanistische Menschenbild stellt das digital erneuerte Maschinenparadigma des Menschen dar. Das digitale Weltmodell, die Welt als umfassender Algorithmus[121], scheint als zeitgenössische Variante des Maschinenparadigmas einer Deutung der Welt als Maschine eine attraktive Interpretation anzubieten. Diese beruht auf der Unterscheidung zwischen Software und Hardware. Es handelt sich dabei um zwei Beschreibungsebenen: die der Hardware, die lediglich auf physikalische und technische Begriffe zurückgreifen muss, und die der Software, die sich wiederum in eine syntaktische und eine semantische Ebene aufteilen lässt. Die syntaktische Beschreibung beruht auf der Zeichenverarbeitung, genauer dem Vokabular und den Regeln der Zeichenverarbeitung. Die Semantik unterlegt den Zeichen und den Regeln, nach denen sie verarbeitet werden, eine Bedeutung. Im Falle von Behauptungssätzen führt diese Unterlegung zu einer Wahrheitswertverteilung; die

[120] Strawson, P.F. (1974): Freedom and Resentment and Other Essays, London. Nida-Rümelin, J. (2005): Über menschliche Freiheit. Stuttgart. Korsgaard, C.M. (1992): The Sources of Normativity. Cambridge.
[121] Nida-Rümelin, J.; Weidenfeld, N. (2018): Digitaler Humanismus. München.

Sätze werden dann als wahr oder falsch markiert; im Falle einer arithmetischen Semantik folgt die Wahrheitswertverteilung mathematischen Regeln.

Die Beschreibung (und Erklärung) von Softwaresystemen als Hardware ist geschlossen: Jeder Vorgang (Ereignis, Prozess, Zustand) lässt sich als kausal determiniert durch den vorausgegangenen Zustand der Hardware eindeutig bestimmen. Als Modell auf den Menschen übertragen heißt dies, dass die physikalisch-physiologische „Hardware" wie ein algorithmisches System mit einer durch Genetik, Epigenetik und sensorische Stimuli eindeutig festgelegten zeitlichen Zustandsfolge von Eigenschaften funktioniert, die durch mentale Termini beschrieben wird und damit bedeutungsvolles Reden und Handeln ermöglicht. Das humanistische Menschenbild und damit die normativen Grundlagen von Moral und Recht erweist sich dann als pure Illusion bzw. kollektive menschliche Selbsttäuschung.[122]

Schon in der ersten Digitalisierungswelle nach dem Zweiten Weltkrieg erwies sich interessanterweise nicht das eben geschilderte materialistische Maschinenparadigma, sondern eine animistische Variante als wirkungsmächtiger. Das animistische Paradigma geht gewissermaßen den umgekehrten Weg der Interpretation: Anstatt den menschlichen Geist und mentale Zustände als Epiphänomene materieller Prozesse in einer physikalisch geschlossenen Welt zu interpretieren und das Materielle mechanistisch zu beschreiben, wird nun im Kontext des Turing-Tests (vgl. Abschnitt 2.1) das algorithmische System mit mentalen Eigenschaften ausgestattet, sofern es in seinem äußeren (Ausgabe-) Verhalten demjenigen von Menschen hinreichend (das heißt verwechselbar) ähnelt.

Entsprechend war in der ersten Phase der Diskussion um Künstliche Intelligenz ab den Siebzigerjahren des 20. Jahrhunderts die Frage, ob Computer denken können, leitend. Für die Fragerichtung ist die kontroverse Diskussion um die Interpretation des von Turing vorgeschlagenen Kriteriums paradigmatisch.[123] Wie zuvor bereits dargelegt, kann nach Turing die Frage, ob technische Artefakte „denken" können, dadurch entschieden werden, dass eine Person (für sie verdeckten) Menschen und Geräten beliebige Fragen stellt. Wenn in einer größeren Zahl von Durchgängen mit wechselnden Fragenden und wechselnden Menschen/Geräten die Antworten zu einem hinreichend großen Anteil (z. B. 50%) nicht eindeutig Mensch oder Gerät zugeordnet

---

[122] Singer, W. (2004): Verschaltungen legen uns fest. Wir sollten aufhören, von Freiheit zu sprechen. In: Geyer, C. (Hg.): Hirnforschung und Willensfreiheit. Zur Deutung der neusten Experimente. Berlin, 30-65. Dennett, D. et al. (2007): Neuroscience and Philosophy: Brain, Mind, and Language. New York.

[123] Turing, A. M. (1950): Computing Machinery and Intelligence. In: Mind LIX (236), 433-460 (DOI: 10.1093/mind/LIX.236.433); vgl. die kritische Darstellung bei Mainzer, K. (1995): Computer – Neue Flügel des Geistes? Berlin, 113 f.

werden können, gibt es nach Turing keinen Grund, technischen Artefakten weniger Denkvermögen zuzuschreiben als Menschen.

Die in Diskursen rund um KI teilweise verbreitete Tendenz, im Anschluss an den Turing-Test eine äußerliche Ununterscheidbarkeit von menschlicher und maschineller Performanz pauschal mit der Annahme von Intelligenz und Denkvermögen solcher Maschinen gleichzusetzen, auf diese Weise die Differenz zwischen dem Simulierten und dem Simulierenden einzuebnen und die menschliche Vernunft damit tendenziell für maschinell ersetzbar zu halten, ist kein Zufall, sondern das Ergebnis bestimmter theoretischer Vorannahmen insbesondere *behavioristischer* und *funktionalistischer* Art.[124] Schon der klassische Behaviorismus[125] hatte sich zu Beginn des 20. Jahrhunderts in dem Bemühen, das menschliche Verhalten auf der Grundlage präzise beschreibbarer Reiz-Reaktion-Schemata zu erklären und die Psychologie damit in eine exakte Wissenschaft zu verwandeln, im Grunde einer Black-Box-Methode bedient, die das Innenleben derart beschriebener Organismen komplett ausblendet.

Der Text von Alan Turing ist zweifellos vom Logischen Behaviorismus inspiriert, der in den Nachkriegsjahren die zeitgenössischen Debatten insbesondere in der britischen Philosophie zunehmend prägte[126]und nach dem sich mentale Zustände ontologisch auf Verhaltensdispositionen reduzieren lassen, also auf die Neigung eines Organismus, sich unter bestimmten Bedingungen auf eine bestimmte Weise zu verhalten. Ein mentaler Zustand wie Schmerz ist demnach lediglich ein Verhaltensmuster, etwa die Veranlagung zu schreien oder zu weinen, wenn man sich verletzt hat. Auch Turings Text identifiziert den Sinn eines sprachlichen Ausdrucks nicht etwa mit der Intention der sprechenden Person, sondern mit den empirischen Verhaltensmustern, die mit einer Äußerung üblicherweise einhergehen. Die Paradoxa, die den Logischen Behaviorismus unglaubwürdig machen, gelten auch für die Turing'sche Variante: Auch wenn wir die Bedeutung eines Satzes, wie „Diese Person hat Schmerzen" lernen, indem wir darauf achten, welches Verhalten jeweils darauf hinweist, dass sie Schmerzen hat, so kann schon deshalb die

[124] Bormann, F.-J. (2021): Ist die praktische Vernunft des Menschen durch KI-Systeme ersetzbar? Zum unterschiedlichen Status von menschlichen Personen und (selbst-)lernenden Maschinen. In: Fritz, A. et al. (Hg.): Digitalisierung im Gesundheitswesen. Anthropologische und ethische Herausforderungen er Mensch-Maschine-Interaktion. Freiburg, 41-64, 51 ff.
[125] Watson, J. B. (1925): Der Behaviorismus. New York.
[126] Wittgenstein, L. (1953): Philosophische Untersuchungen. Oxford, Malden. Ryle, G. (1949)*:* The Concept of Mind. London. Einen guten Zugang zur damaligen *ordinary language philosophy* vermittelt Savigny, E (1973): Zur Philosophie der normalen Sprache. Frankfurt a. M.

+++ Vorabfassung! Wird durch die lektorierte Fassung ersetzt. +++

Bedeutung von „Schmerzen haben“ nicht lediglich ein Verhaltensmuster sein, weil „Superspartaner“, die keine Schmerzen zeigen, dann auch keine Schmerzen haben könnten.[127] Obwohl sich der behavioristische Theorieansatz schon bald als zu eng erweisen sollte und in den folgenden Jahrzehnten verschiedene Transformationen erfuhr, fand er mit dem sich seit den 1950er-Jahren ausbreitenden Funktionalismus, nach dem mentale Zustände funktional vollständig erfasst werden, eine Nachfolgetheorie, die der aufkeimenden Kognitionspsychologie und Computerwissenschaft noch besser entsprach, weil Computer in dieser Lesart ein geeignetes Modell für mentale Prozesse sein können.[128]

Der Reiz funktionalistischer Ansätze besteht zunächst darin, dass sie sich gegenüber den ontologischen Implikationen des Leib-Seele-Problems neutral verhalten, das heißt sie umgehen die Frage nach der Beziehung zwischen Körper und Geist sowie die Frage, wo und wie sich in diesen das denkende und fühlende Subjekt verorten lässt. Der Funktionalismus plädiert dafür, die Frage nach der Seinsart mentaler Zustände zugunsten der genauen Beschreibung ihrer Funktion aufzuheben. Durch die These der *multiplen Realisierung*[129], nach der bestimmte mentale Ereignisse, Eigenschaften oder Zustände durch ganz unterschiedliche physikalische Ereignisse, Eigenschaften oder Zustände realisiert werden können, schien es zudem möglich, auch Computern mentale Zustände zuzuschreiben, obwohl sie keine biologischen Strukturen besitzen.[130] Indem der Funktionalismus durch die ausdrückliche Anerkennung theorierelevanter innerer Zustände eines Systems nicht nur das Blackbox-Prinzip des Behaviorismus überwand, sondern mit der funktionalen Interpretation solcher Zustände auch die Integration biologischer und maschineller Entitäten in eine umfassende einheitliche Theorie zu ermöglichen schien, bahnte er insofern auch einer animistischen Deutung von KI-Systemen den Weg, als diese nun umso

[127] Putnam, H. (1965): Brains and Behaviour. In: Butler, R. J. (Hg.): Analytical Philosophy, Band 2. Oxford, 24-36.
[129] Vgl. Putnam, H. (1980): Philosophy and Our Mental Life. In: Block, N. (Hg.): Readings in Philosophy of Psychology, Vol. I § 7. Cambridge (MA), London, 134-143 (DOI: 10.4159/harvard.9780674594623.c11). Von dierser Position distanziert sich Putnam zu späterer Zeit, siehe dazu: Putnam, H. (1992): Renewing Philosophy. Cambridge (MA).
[129] Für die unterschiedlichen Entwicklungsstufen dieser These vgl. Bickle, J. (1998):Multiple Realizability. In: Stanford Encyclopedia of Philosophy, Summer 2020 Edition.
https://plato.stanford.edu/archives/sum2020/entries/multiple-realizability/ [04.01.2023]. Sowie Polger, T.; Shapiro, L. (2016): The Multiple Realization Book. New York.
[130] Putnam, H. (1967): Psychological Predicates. In: Capitan, W. H.; Merrill, D. D. (1967): Art, Mind, and Religion. Pittsburgh, 37–48. (dt. Übersetzung: Die Natur mentaler Zustände, In: Metzinger, T. (2007): Grundkurs Philosophie des Geistes 2: Das Leib-Seele-Problem. Paderborn, 372–385) sowie Fodor, J. (1974): Special Sciences (or: The disunity of science as a working hypothesis). In: Synthese 28 (2), 97–115 (DOI: 10.1007/BF00485230).

leichter als handlungsfähige Agenten mit einem mentalen Innenleben vorgestellt werden konnten, denen man zutraute, die menschliche Vernunft irgendwann einmal ersetzen zu können.

Obwohl der Funktionalismus aufgrund dieser Vorteile zunächst viel Zuspruch sowohl in der analytischen Philosophie des Geistes als auch in der KI-Forschung fand, wurden schon bald Einwände gegen dieses Theoriemodell vorgetragen. Eine erste gewichtige Kritik einer funktionalistischen Betrachtung des Mentalen, die den menschlichen Geist als Rechenmaschine[131] fasst, deren innere Zustände allein von ihrer Funktion im Sinne einer kausalen Verknüpfung von Eingabe und Ausgabe bestimmt werden, stammt von Thomas Nagel. Seines Erachtens lässt unsere gewöhnliche Auffassung mentaler Phänomene eine solche reduktionistische Sichtweise allein schon deswegen nicht zu, weil das Mentale neben seiner bloßen Funktion auch durch ein bestimmtes *phänomenales Bewusstsein*[132] geprägt sei, das in dieser Beschreibung verloren gehe. Nagel leugnet weder, „daß bewußte mentale Zustände und Ereignisse Verhalten verursachen, noch, daß man sie funktional charakterisieren könnte", sondern bestreitet lediglich, „daß derartiges eine vollständige Analyse ergibt".[133]

Ein Wesen kann nur *als* dieses Wesen mentale Zustände haben. Daher kann von den *Empfindungsqualitäten* des phänomenalen Bewusstseins nicht abgesehen werden, die allein aufgrund *äußeren* Verhaltens nicht zugänglich sind. Es fühlt sich für uns immer auf eine ganz bestimmte Art und Weise an, ein erlebendes, denkendes und handelndes Subjekt zu sein. Diese besonderen Qualitäten sind kein flüchtiges Beiwerk mentaler Zustände, sie gehören vielmehr insofern konstitutiv zu allen unseren Erfahrungen, als wir die uns umgebende Welt prinzipiell gar nicht anders erleben können als aus der Perspektive eines solchen phänomenalen Bewusstseins.

Dieses phänomenale Bewusstsein setzt dem Vermögen, die Qualität des Erlebens oder die mentalen Zustände anderer Lebewesen zu beurteilen, gewisse Grenzen. Dies illustriert Nagel am Beispiel des Empfindens einer Fledermaus, deren Orientierung aufgrund spezifischer sensorischer Besonderheiten (wie Radar oder Echolotortung) ganz anders strukturiert ist als beim Menschen. Als Menschen können wir zwar versuchen, uns vorstellen, wie es ist, sich auf gänzlich andere Weise im Raum zu orientieren, wir bleiben dabei aber immer unserer eigenen, spezifisch

---

[131] oder – um eine Kritik von Ned Block aufzugreifen – wie ein Getränkeautomat: Block, N. (1979): Troubles with Functionalism. In: Minnesota Studies in the Philosophy of Science 9, 261–325.
[132] Zur Debatte um die Bedeutung des phänomenalen Bewusstseins vergleiche auch die Textsammlungen von Heckmann, H.-D.; Walter, S. (2006): Qualia. Ausgewählte Beiträge. Paderborn und Metzinger, T. (2009): Grundkurs Philosophie des Geistes 1: Phänomenales Bewusstsein. Paderborn.
[133] Nagel, T. (1981): Wie ist es, eine Fledermaus zu sein? In: Bieri, P. (Hg.): Analytische Philosophie des Geistes. Königstein im Taunus, 261–275, 262.

menschlichen Weise des Erlebens verhaftet, ohne jemals Zugang zu den besonderen Qualitäten der mentalen Zustände einer Fledermaus zu erhalten. Unsere subjektive Perspektive bleibt unüberwindlich.[134] Vor diesem Hintergrund dieses auch als Qualia-Argument[135] bezeichneten Gedankengangs basiert die funktionalistisch inspirierte Mensch-Computer-Analogie auf einer fragwürdigen Reduktion, die nur gelingen kann, „wenn die artspezifische Betrachtungsweise von dem, was reduziert werden soll, ausgeklammert wird".[136]

Gegen eine funktionalistische Interpretation Künstlicher Intelligenz hat der Philosoph John Searle ein Argument entwickelt, das als das meistdiskutierte in der zeitgenössischen Philosophie gilt: *The Chinese room*. Es verweist auf ein Gedankenexperiment, in dem eine Person in einem Zimmer sitzt, in das durch einen Schlitz jeweils Fragen in chinesischer Schrift gereicht werden. Die Person reicht Antworten auf diese Fragen ebenfalls durch den Schlitz heraus. Wenn die Antworten hinreichend plausibel erscheinen, mag man vermuten, dass die Person im chinesischen Zimmer des Chinesischen mächtig ist. Nun stellt sich aber heraus, dass jede der eingereichten Fragen eine Ziffer trägt und die Person über vorgefertigte Antworten und eine Tabelle mit Zuordnungen verfügt, sodass sie lediglich eine Antwort heraussuchen muss, die eine Ziffer trägt, die der Ziffer der Fragestellung zugeordnet ist.

Die Person beherrscht die chinesische Sprache nicht, und auch das Zimmer als Ganzes mit einer Eingabe- und Ausgabefunktion beherrscht diese Sprache nicht. Aber zweifellos muss es irgendjemanden geben, der des Chinesischen mächtig ist und daher in der Lage war, den Fragen mithilfe der Ziffern Antworten so zuzuordnen, sodass der Eindruck entsteht, dass die Person im chinesischen Zimmer Chinesisch versteht. Die Analogien zu Softwaresystemen liegen auf der Hand. Es handelt sich um Zuordnungsregeln (genauer um Algorithmen), die lediglich für die Programmierung und den Gebrauch der digitalen Maschine Bedeutung haben, aber nicht das Softwaresystem selbst zu einer semantischen Maschine machen. Dieses verfügt nicht über Bedeutungen, es versteht nichts, es entscheidet nichts. Softwaresysteme verfügen nicht über eine Semantik, es handelt sich nicht um semantische Maschinen.[137]

---

[134] Nagel, T. (1981): Wie ist es, eine Fledermaus zu sein? In: Bieri, P.(Hg.): Analytische Philosophie des Geistes. Königstein im Taunus, 261–275, 262.
[135] von lat. *qualis* „wie beschaffen"
[136] Nagel, T. (1981): Wie ist es, eine Fledermaus zu sein? In: Bieri, P. (Hg.): Analytische Philosophie des Geistes. Königstein im Taunus, 261–275, 269.
[137] Searle, J. R. (1980): Minds, Brains and Programs. In: Behavioral and Brain Sciences 3 (3), 417-457. Eine frühere Zurückweisung des sogenannten Funktionalismus stammt von Block, N. (1978): Troubles with Functionalism. In Savage, C. W. (Hg.): Perception and Cognition. Minneapolis, 9-261.

+++ Vorabfassung! Wird durch die lektorierte Fassung ersetzt. +++

3.4.3 Verleiblichte Vernunft

Die Zurückweisung funktionalistischer Maschinenparadigmen lenkt den Blick auf die bereits im Qualia-Argument angedeutete besondere Qualität menschlicher Vernunft und deren Bedeutung für das menschliche Selbstverständnis. Menschliche Vernunft ist leibliche Vernunft. Diese Einsicht wendet sich gegen den in der abendländischen Tradition lange herrschenden Gedanken eines Dualismus zwischen Vernunft und Natur, Körper und Geist, der den Menschen als Naturwesen auf der einen und als Vernunftwesen auf der anderen Seite begreift. Solch dualistische Vorstellungen wurden sowohl durch Erkenntnisse in der Evolutionsbiologie als auch in der Hirnforschung und in den Kognitionswissenschaften infrage gestellt, die stattdessen auf die Relevanz des Leiblichen für die Bestimmung menschlicher Intelligenz und auf die Bedeutung unbewusster Prozesse für die Entwicklung höherer geistiger Leistungen verweisen.

Mit Maurice Merleau-Pontys Unterscheidung von zweierlei Arten des Körpers kann dies veranschaulicht werden.[138] Er unterscheidet den lebendigen handelnden *Leib* vom rein physikalischen *Körper*. Die Fähigkeit, soziale Bindungen einzugehen, sich in andere hineinzuversetzen, wird ermöglicht dadurch, dass der Mensch Leib ist und nicht nur einen Körper hat. Mit dem Leib ist der empfindende Organismus gemeint, mit seinem Vermögen, zu fühlen und sich zu bewegen. Wir sind an diesen Leib gebunden, in allem, was wir denken und tun. Er ist daher Ausgangspunkt und Bestandteil jeder Wahrnehmung und Empfindung. Als solcher ist er Voraussetzung für unser In-der-Welt-Sein und zugleich dafür, eine Welt zu haben und eine Beziehung zu anderen herzustellen.

Kognitive Fähigkeiten sind in ihrem Entstehungs- und Vollzugsprozess also an Sinnlichkeit und Leiblichkeit gebunden. Dies hat Konsequenzen für unser Verständnis vom Gehirn als Erkenntnisorgan und von der Vernunft als Erkenntnisvermögen – und damit auch für das Verständnis unseres Zugangs zur Realität. Wesentlich ist dabei, dass das in der Verkörperung des Gehirns eingeschlossene Naturverhältnis einer leiblichen Vernunft des Menschen seine Sozialität impliziert und seine Kulturalität bestimmt. Im menschlichen Leib sind Sozialität und Kulturalität von Anfang an angelegt, vor aller Entwicklung eines reflexiven und sprachlich vermittelten Bewusstseins.[139] Denn leibliche Vernunft vollzieht nicht nur einen kognitiven

---

[138] Vgl. dazu Merleau-Ponty, M. (1966): Phänomenologie der Wahrnehmung. Übersetzt und mit einem Vorwort von Boehm, R. Berlin.

[139] Vgl. hierzu Fuchs, T. (2013): Verkörperung, Sozialität und Kultur. In: Breyer, T. et al. (Hg.): Interdisziplinäre Anthropologie, Leib – Geist – Kultur. Heidelberg, 9-261, 11 ff. Vier Erscheinungsformen des Leibes können diese Anlage der Vermittlung zwischen der Natur- und der Kulturseite des Menschen plausibel machen: (i) ein mit der Umwelt vertrauter Leib, der sich vor allem im Umgang mit kulturellen Gegenständen entwickelt, (ii) ein

Informationsaustausch, sondern mit ihr spielen auch Kommunikation und Kooperation eine Rolle.[140] Beides sind Faktoren, die von Kindheit an entscheidend sind für jene bewussten Prozesse, in denen sich die Kulturfähigkeit bildet, die dem Menschen als sozusagen „zweite Natur“[141] zuwächst. Für das menschliche Gehirn bedeutet das, dass es mit all seinen sich entfaltenden Fähigkeiten von Anfang an in biologisch-organische wie in sozial-kulturelle Entwicklungsprozesse eingebunden ist.

Ein solches Verständnis des Menschen geht nicht von einer blutleeren „reinen“ Vernunft aus, sondern begreift auch die Vernunft als immer schon leiblich eingebunden und sozial wirksam. Damit ist die Frage nach dem Praktischwerden der Vernunft, das heißt nach den normativen Orientierungen und nach der Motivation moralischen Handelns, keine zweite Frage, sondern sie begleitet alles Denken, das als solches Lebensgestaltung nicht nur ermöglicht, sondern immer bereits vollzieht.

Bereits die praktische Einsicht, dass bestimmte normative Gründe für eine Handlung sprechen, wird damit *handlungswirksam*. Damit wäre der entscheidende Aspekt einer umfassenden Theorie praktischer Vernunft, nämlich die Frage, wie sich moralische Überzeugungen in Handlungen überführen lassen, berührt. Bezüglich dieser Frage plädiert der Philosoph John McDowell dafür, in diesem Kontext nicht entweder die Vernunft oder aber die subjektiven Einstellungen und Strebungen zum alles beherrschenden Faktor der Handlungsverursachung zu stilisieren, sondern zu akzeptieren, dass in der menschlichen moralischen Erfahrung diese beiden (das heißt die kognitive und die appetitive Dimension) immer schon unauflöslich miteinander verschränkt seien.[142]

Entscheidend ist dabei, dass leibliche Erfahrungsstrukturen einhergehen mit der Fähigkeit, sich in andere hineinversetzen und sich mitteilen zu können, das heißt mit einer Prosozialität, auf deren Grundlage sich die Fähigkeit zu geteilter Intention entwickeln und Empathie und Motivation initiieren kann.

Insofern nun das Gehirn kein isolierter Gegenstand ist, sondern eingelassen ist in Erfahrungen gemeinsamer Praxis, in der sich körperlich-biologisches und kulturell-soziologisches Erleben

---

„passiv-affizierbarer“ Leib der affektiv mit anderen verbunden ist, (iii) ein „mimetisch-resonanter“ Leib, der durch Nachahmung in grundlegende Kommunikation mit anderen eingebunden ist, bis er so als (iv) kooperativ kultivierter Leib zum Körper für andere wird, indem er Haltungen und Rollen übernimmt, die ihm somit zur „zweiten Natur“ werden. Vgl. ebd. 26f.

[140] Schmitz, H. (1990): Der unerschöpfliche Gegenstand. Grundzüge der Philosophie. Bonn.

[141] McDowell, J. (2001): Geist und Welt. Frankfurt a. M., 109. McDowell, J. (2009): Zwei Arten von Naturalismus. In: ebd.: Wert und Wirklichkeit. Aufsätze zur Moralphilosophie. Frankfurt a. M., 30-73.

[142] McDowell, J (2002): Interne und externe Gründe. In: ebd.: Wert und Wirklichkeit. Aufsätze zur Moralphilosophie, Frankfurt a. M., 156-178. 177.

verschränken, entwickeln sich in solch sozialer Praxis ein Bedeutungswissen und ein Wissen um die Perspektivität von Erkenntnis, die reflexiv auf die Relationalität von Wissen und Erkenntnis verweist. Denn diese Perspektivität des Wissens erschließt sich insbesondere durch dessen In-Relation-Stehen zum eigenen Leib, durch die sogenannte „Eigenleiberfahrung". So wird mit dem Rekurs auf die verleiblichte Vernunft deutlich, dass zur menschlichen Intelligenz unabdingbar Reflexivität hinzugehört. Diese setzt menschliche Erkenntnis instand, zu unterschiedlichen Perspektiven Stellung nehmen und urteilen zu können.

Grenzen der Formalisierbarkeit und Simulierbarkeit menschlicher Vernunft

Mit der Reflexivität des Bewusstseins ist das Verstehen- und Vermittelnkönnen angesprochen, mit anderen Worten die hermeneutische Dimension, die sich auch in der Unterscheidung und Anwendung verschiedener Wissensformen darstellt und die ein besonderes Charakteristikum menschlicher Intelligenz bildet. Diese hermeneutische Dimension von Wissen ist aber nur begrenzt formalisierbar oder simulierbar und bezieht sich auf den Sinn und die Bedeutung menschlichen Erkennens und Handelns. Die Aneignung menschlicher Erfahrung ist immer mit Deutungsprozessen verbunden und setzt immer ein Beteiligtsein, ein Engagement voraus.[143] Die Art und Weise, *wie* wir wissen (*knowing how*), ist eine eigene Kompetenz, die sich nicht durch bloßes Sachwissen (*knowing that*) abbilden lässt.

Auch hier spielt der Leib eine wichtige Rolle, denn er ermöglicht ein Handeln, das allein mittels bewusster Planung und Berechnung so nicht möglich wäre. In der leiblichen Verfasstheit gründet daher auch die Nichtsimulierbarkeit des Denkens. Mit der leiblichen Verankerung des Bewusstseins, die eine Komplexität von Hintergrunderfahrungen mit sich führt, die Voraussetzung für alle bewussten Prozesse der Planung und Entscheidung sowie deren Begründung bilden, stößt die Entwicklung von Künstlicher Intelligenz an ihre Grenzen.

Der Sachverhalt solch leiblich verfassten Hintergrundwissens bedeutet mithin eine Grenze des rationalistischen Versuchs, Wissen vollständig in formalisierte Regeln zu überführen und künstlich nachzubilden. Es zeigt sich in der leiblichen Verschränkung von kognitiven und emotional appetitiven Momenten im Vernunftvollzug dann vielmehr die Relevanz des Nichtformalisierbaren. Es geht im Begreifen nicht mehr nur darum, das *Was* – die Fakten – zu begreifen,

[143] Meyer-Drawe, K. (2001): Leiblichkeit und Sozialität. 3., unveränderte Auflage. München.

sondern darum, *wie* wir verstehen. Und dies wird entscheidend durch unsere leiblichen Vollzüge und Fähigkeiten bestimmt, die vorbegrifflich und unausgesprochenen unser Verhalten mitbestimmen.

Menschliches Wissen ist insofern eingebettet in einen Horizont des Nichtwissens. Im Raum individueller und sozio-kultureller Erfahrung wird deutlich, dass nicht nur die Klarheit logischen Schließens, sondern auch die Vagheit und Offenheit menschliches Denken auszeichnet. Gerade Vagheit und Unbestimmtheit des Wissens sind Voraussetzung für Kreativität und Intuition, die ein Handeln unter der Bedingung von Ungewissheit ermöglichen, mit der Menschen situativ auf konkrete Herausforderungen reagieren und Verantwortung übernehmen können. Für menschliche Intelligenz ist kennzeichnend, dass sie sich auf plötzliche Situationen einstellen kann, um in Erlebnisgegenwart Entscheidungen für oder gegen Zukunftsszenarien zu treffen.

Wesentlich ist daher für menschliche Intelligenz und deren Verantwortungsfähigkeit auch das Erleben von und der Umgang mit Zeit. Entscheidungen werden in Gegenwart erlebt und in solchem Gegenwartserleben bewusst gehalten. Auch dieses Gegenwartserleben ist in der Leiblichkeit der Vernunft verankert. Veranschaulicht werden kann dies an der Bedeutung des menschlichen Gedächtnisses. Das menschliche Gedächtnis funktioniert nicht wie ein Speicher, der einen gedanklichen Bestand bildet und abrufbar wäre. Vielmehr ist es durch Prozesse des Erinnerns und Vergessens ausgezeichnet. Was jeweils im Moment erinnert – oder vergessen – wird, hängt von der je konkreten leiblichen Verfasstheit und den sozialen Bezügen, in denen der Mensch steht, ab. Erinnern ist damit nicht gleichbedeutend mit dem Abrufen einer Information. Es ist vielmehr ein hermeneutischer Akt, mit dem sich ein Erfahrungsraum[144] vergegenwärtigt.

3.5 Fazit

Aus den vorherigen Überlegungen lassen sich einige entscheidende Aspekte kognitiver Leistungen und Operationen von Menschen und Maschinen zusammenfassen. Das Kognitive ist im Falle menschlicher Intelligenz unauflöslich mit den kognitiven und emotiven, ästhetischen und ethischen, technischen und gestalterischen, sozialen und individuellen sowie zeitlichen Dimensionen der menschlichen Lebenswelt verbunden. Menschliche Intelligenz zeigt sich nicht nur in kognitiv kohärentem Urteil, sondern auch in einer kohärenten Praxis. Diese ist gründegeleitet und Ausdruck von akzeptierten Werten und Normen, die nicht beliebig zur Disposition stehen.

[144] Meyer-Drawe, K. (2001): Leiblichkeit und Sozialität. 3., unveränderte Auflage. München.

Der Mensch ist durch die Fähigkeit, Gründe zu geben und zu nehmen und sich im Urteil und im Handeln an diesen zu orientieren, als Spezies charakterisiert. Veränderungen im normativen Gefüge der eigenen Praxis bedürfen der Begründung und bedrohen im Grenzfall die persönliche Integrität und Identität. Ein hinreichend entwickeltes lebensweltliches Orientierungswissen ist Voraussetzung für eine intelligente Praxis. Damit sich dieses aufbaut, muss die betreffende Person die Fähigkeit haben, Wichtiges von Unwichtigem zu unterscheiden und normative Grenzen zu akzeptieren.

Menschliche Intelligenz ergibt sich zudem nicht allein aus dem Orientierungsbedarf des Individuums in einer natürlichen Welt, sondern ist das Ergebnis sozialer Interaktion. Von Geburt an hängt das Wohlbefinden menschlicher Wesen und hängen deren Entwicklungschancen vom Austausch mit anderen Menschen ab. Der intelligente Umgang mit den Herausforderungen der Welt ist nicht das Ergebnis eines fortgesetzten Puzzlespiels, sondern im Wesentlichen Folge der Einbettung der eigenen individuellen Praxis in den größeren sozialen und kulturellen Zusammenhang. Mit dem Erwerb der Sprache können Kinder auf Gründe reagieren, sich von Gründen affizieren lassen und diese selbst auf ihr eigenes Handeln applizieren. Das kulturelle Wissen wird über diese Praxis von einer Generation auf die nächste übertragen, immer wieder veränderten Bedingungen angepasst und bettet das einzelne Individuum in die menschliche Lebensform ein. Im Ergebnis verweben sich dann die Einheit der Person mit der Einheit des Wissens und der Einheit der menschlichen Lebensform. Die einzelne Person zerfällt nicht in Funktionalitäten, sondern wird zusammengehalten durch Gründe, die ihre theoretische und praktische Lebensorientierung bestimmen. Das Individuum wird zur Person und zum Handelnden. Die Identität der Person äußert sich in einer kohärenten Praxis, die von stabilen Gründen geleitet ist. Diese integriert unterschiedliche Aspekte menschlicher Existenz – kognitive, emotive, soziale, ethische, ästhetische, technische und gestalterische.

Es ist fraglich, ob eine derart gründegeleitete, multidimensional bestimmte und soziokulturell eingebettete kohärente Praxis selbst für komplexe maschinelle Systeme jemals plausibel sein könnte. Softwaresysteme leisten Beachtliches. In vielen Bereichen sind sie menschlichen Fähigkeiten bei Weitem überlegen. Aber sie verfügen nicht über ein Analogon zu menschlicher Intelligenz. Es wird der Softwareentwicklung der Zukunft vermutlich in wachsendem Umfang gelingen, menschliche Fähigkeiten zu simulieren und in vielen Fällen zu übertreffen. Das sollte uns aber nicht dazu verführen, ihnen personale Eigenschaften zuzuschreiben, die für genuine menschliche Existenz essenziell sind.

Trotz dieser kategorialen Unterschiede von Mensch und Maschine beeinflussen Mensch und Maschine einander fortwährend. Menschen entwickeln zu bestimmten Zwecken Technologien, die auf die Handlungsmöglichkeiten von Menschen zurückwirken, indem sie jene verändern, erweitern oder vermindern. Diese Mensch-Technik-Relationen und ihre ethische Relevanz genauer zu bestimmen, ist Gegenstand des folgenden Kapitels.

+++ Vorabfassung! Wird durch die lektorierte Fassung ersetzt. +++

4 Mensch-Technik-Relationen

4.1 Einleitung

Das Verhältnis zwischen menschlichem Handeln, der Verfügbarkeit von Technik und technischen Prozessen ist für die Ethik hoch relevant, denn in diesem Verhältnis können sich Einfluss- und Randbedingungen für Autonomie und Freiheit des Menschen und damit die Möglichkeit der Zuschreibung von Verantwortung auf durchaus komplexe Weise ändern. Dies gilt vor allem und auf spezifische Weise bei KI-Systemen. Es ist daher im Rahmen einer ethischen Analyse und Beurteilung relevant, das Zusammenspiel von Mensch und Technik bzw. von menschlichem Handeln und technischen Prozessen differenziert zu erfassen. Menschen entwickeln und gestalten Technik und nutzen technische Produkte und Systeme oder darauf aufbauende Dienstleistungen als Mittel zum Zweck. Gleichzeitig wirken diese häufig zurück und beeinflussen menschliche Handlungsmöglichkeiten, von der Eröffnung neuer Optionen und der Vergrößerung von Freiheitsgraden bis hin zur Anpassungserzwingung. In diesem Kapitel soll es darum gehen, in welcher Art und Weise verschiedene Mensch-Technik-Relationen die Handlungsmöglichkeiten des Menschen erweitern oder vermindern können, bis hin zur Ersetzung menschlicher Handlungen durch maschinelle Vollzüge. Damit verbunden ist die Frage, wie sich die Spielräume für die Entfaltung menschlicher Autorschaft und die Übernahme von Verantwortung jeweils verändern.

Zum einen geht es darum, dass Tätigkeiten, die vormals (allein) von Menschen durchgeführt wurden, graduell an technische Systeme delegiert werden. Dies reicht vom Delegieren einfacher Tätigkeiten über das Automatisieren komplexer Tätigkeiten oder ganzer Funktionen bis hin zur vollständigen Ersetzung des Menschen durch Technik. Der Begriff „*Ersetzen*“ beschreibt hier also den Endpunkt einer vollständigen Delegation. Zum anderen geht es um Rückwirkungen dieses mehr oder minder umfassenden Delegierens auf menschliche Akteure, das heißt um Fragen, inwiefern jenes Delegieren Handlungsfähigkeit, Möglichkeiten, Fertigkeiten und Kompetenzen von Menschen *erweitert* oder *vermindert*.

Die drei Begriffe des Erweiterns, Verminderns und Ersetzens dienen in diesem Kapitel als analytische Matrix. Sie werden in den folgenden Kapiteln auf ausgewählte Sektoren bezogen, um ein differenziertes Bild der Veränderungen durch KI und ihrer ethisch relevanten Aspekte zu gewinnen.

+++ Vorabfassung! Wird durch die lektorierte Fassung ersetzt. +++

4.2 Technikdeterminismus versus Sozialkonstruktivismus

Mensch-Technik-Relationen sind Gegenstand vieler Disziplinen. Verortet zwischen Informatik und Psychologie, beschäftigt sich insbesondere das Feld der Human-Computer-Interaction bzw. Computer-Human-Interaction mit dem Verständnis und der Gestaltung von Mensch-Maschine-Schnittstellen. In den Geistes- und Sozialwissenschaften haben insbesondere die Wissenschafts- und Technikforschung, die Science and Technology Studies (STS), die Techniksoziologie und die Technikphilosophie Konzepte und Theorien zur begrifflichen Analyse von Mensch-Technik-Relationen bereitgestellt. Das Verhältnis von Menschen und Gesellschaft zur Technik wurde vielfach entlang der Deutungslinie zwischen sozialem Konstruktivismus und technologischem Determinismus beschrieben.[145] Dahinter steht die Frage nach dem letztlich treibenden Faktor: Folgen Technikgestaltung im Einzelnen und damit auch der technische Fortschritt als Prozess eher menschlich gesetzten Zwecken oder eher einer Eigendynamik, der sich Mensch und Gesellschaft letztlich unterordnen und anpassen müssen. Auch wenn es keine einheitliche Verwendungsweise dieser beiden Deutungen gibt und auch wenn viele Ansätze keine der Extrempositionen vertreten, sondern sich an unterschiedlichen Stellen zwischen Sozial- und Technikdeterminismus verorten, ist eine kurze Erläuterung illustrativ und inhaltlich für diese Stellungnahme wichtig.

In der bereits seit den 1920er-Jahren vertretenen technikdeterministischen Sichtweise wird eine Eigenlogik im technischen Wandel vermutet, die Mensch und Gesellschaft zur Anpassung nötigt. Während sich einzelne Techniken auf menschliche Zwecke zurückführen lassen, folge die gesamte Technologieentwicklung einer inneren und damit nicht oder kaum beeinflussbaren Dynamik. Als Treiber hinter dieser vermuteten Eigendynamik wird immer wieder auf ökonomische Verhältnisse und insbesondere den wirtschaftlichen Wettbewerb zwischen Unternehmen, aber auch den Wettbewerb zwischen Staaten und Weltregionen um vordere Plätze in der technologischen Forschung und Entwicklung hingewiesen. Der auf diese Weise zustande kommende, sozusagen blinde technische Wandel wirke sodann mit seinen Produkten auf die Gesellschaft ein und führe zu Anpassungsnotwendigkeiten, die von konkreter Akzeptanz einzelner Techniken bis hin zur Adaptation an letztlich technologisches Denken reichen.[146]

---

[145] Grunwald, A. (2007): Technikdeterminismus oder Sozialdeterminismus: Zeitbezüge und Kausalverhältnisse. In: Dolata, U.; Werle, R. (Hg.): Gesellschaft und die Macht der Technik. Sozioökonomischer und institutioneller Wandel durch Technisierung. Frankfurt a.M., New York, 63–82.

[146] Rapp, F. (1978): Analytische Technikphilosophie. Freiburg; Ropohl, G. (1982): Kritik des technologischen Determinismus. In: Rapp, F.; Durbin, P. T. (Hg.): Technikphilosophie in der Diskussion. Braunschweig, 3–18. Grunwald, A. (2019): Technology Assessment in Practice and Theory. Abingdon.

In der sozialkonstruktivistischen Sichtweise dagegen treten Technologien nicht eigendynamisch oder zwangsläufig auf den Plan, sondern sind das Ergebnis komplexer und sozial situierter Entwicklungs- und Gestaltungspraktiken bzw. von Ko-Konstruktions-Prozessen unter Mitwirkung zahlreicher Akteure. Die Technikgeneseforschung[147] hat herausgearbeitet, nach welchen Mechanismen Technik durch Entscheidungsprozesse aus ersten Ideen über Entwicklungsprogramme, Experimente und Prototypen bis zum letztendlichen Ergebnis entsteht. Beispielsweise wurde die Rolle von gesellschaftlichen Leitbildern in diesen Prozessen untersucht. [148] Sozialkonstruktivistisch gesehen werden Algorithmen, Roboter, digitale Dienstleistungen oder Geschäftsmodelle für digitale Plattformen von Menschen in möglicherweise langwierigen Entscheidungsprozessen und Handlungssträngen erfunden, entworfen, hergestellt und eingesetzt sowie weiterentwickelt und an neue Umgebungen angepasst. Die „Macher" der Digitalisierung arbeiten in der Regel in Unternehmen, Forschungsinstitutionen oder Geheimdiensten mit bestimmten Agenden, Interessen und Geschäftsmodellen. Wenn *andere* Personen und Institutionen mit *anderen* Werten, Perspektiven und Interessen mitgestalten könnten, würden beispielsweise KI-Systeme mit anderen Eigenschaften entstehen, als wenn man diese den einschlägigen Konzernen mit deren Interessen und Geschäftsmodellen überlässt. Diese Sicht eröffnete Gestaltungsmöglichkeiten und motivierte partizipative Ansätze in der Technikgestaltung wie beispielsweise das Constructive Technology Assessment.[149]

Die grobe Skizze dieser beiden Positionen macht die jeweiligen blinden Flecken deutlich. Weder darf die Bedeutung gesellschaftlicher Hintergründe oder spezifischer Entscheidungen bei der Entwicklung von Technik ignoriert werden, so etwa in Unternehmen oder der öffentlich geförderten Forschung, noch kann abgestritten werden, dass Technologien auf gesellschaftliche Realitäten und menschliche Handlungsmöglichkeiten zurückwirken. Daher erscheint es ratsam, Technikdeterminismus und Sozialkonstruktivismus als Pole eines empirisch vielfältigen und

[147] Bijker, W. E. et al. (1987): The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology. Cambridge (MA), London; Weyer, J. et al. (1997): Technik, die Gesellschaft schafft. Soziale Netzwerke als Ort der Technikgenese. Berlin. Weingart, P. (1989): Technik als sozialer Prozess. Frankfurt a.M.
[148] Dierkes, M. et al. (1992): Leitbild und Technik. Zur Entstehung und Steuerung technischer Innovationen. Berlin, Frankfurt a. M., New York.
[149] Rip, A. et al. (1995): *Managing Technology in Society. The Approach of Constructive Technology Assessment.* London.

differenzierten Spektrums zu begreifen, die den Blick auf unterschiedliche Aspekte von Technikentwicklung und Mensch-Technik-Relationen werfen.[150] Entsprechend sind weder die technikdeterministische noch die konstruktivistische Betrachtung der Mensch-Technik-Relation falsch, beide sind jedoch unterkomplex. Sie werden empirisch unzutreffend, wenn sie in ihrer jeweiligen Perspektive verabsolutiert werden. Die Mensch-Technik-Relation unterliegt vielmehr von Grund auf einem Ko-Konstruktions-Verhältnis und kann als Ko-Evolution beschrieben werden.[151] Soziale Kontexte und normative Kriterien auf der einen und Technologien auf der anderen Seite entwickeln sich weiter in gegenseitiger Wechselwirkung. Die Verfügbarkeit von Technik beeinflusst Handlungsmöglichkeiten und deren Realisierung, aber auch die Bedingungen und Möglichkeiten menschlicher Weltwahrnehmung, wodurch sich Lebensstile und Einstellungen verändern können. Umgekehrt entstehen, wie dies die Technikgeneseforschung in vielen empirischen Studien belegt hat, neue Techniken vor dem Hintergrund von sozialen Befindlichkeiten, normativen Kriterien und Zukunftsentwürfen.

Gerade im Kontext der Künstlichen Intelligenz sind die Arbeiten der Anthropologin Lucy Suchman von großer Bedeutung. Ihr Buch „Human-Machine Reconfigurations“[152] liefert eine Reflexion und Kontextualisierung ihrer Studien der KI-Forschung in den 1980ern, welche 1987 unter dem Titel „Plans and Situated Action“ veröffentlicht wurden.[153] Suchman kritisiert das Planungsmodell von Interaktion, das einem Großteil der damaligen Forschung zugrunde liegt, und schlägt einen Perspektivenwechsel in der Betrachtung der Mensch-Maschine-Relation vor, der Einsichten aus den Sozialwissenschaften Rechnung trägt. Danach ist menschliches Handeln auf vielfältige Weise sozial situiert und beeinflusst, ohne vollständig determiniert zu sein. Sie argumentiert, dass Menschen sinnvoll handeln, indem sie auf der Grundlage ihrer sozialen und ökologischen Ressourcen häufig weniger planen als improvisieren. Sie kritisiert also die theoretisch-konzeptionellen Grundlagen des Designs interaktiver technischer Systeme als aus anthropologischer Sicht unangemessen, weil menschliches Handeln ständig aus dynamischen Interaktionen mit der materiellen, insbesondere technischen, und der sozialen Welt konstruiert und rekonstruiert werde.

---

[150] Dolata, U.; Werle, R. (2007): Gesellschaft und die Macht der Technik. Sozioökonomischer und institutioneller Wandel durch Technisierung. Frankfurt a. M., New York.
[151] Rip, A. (2007): Die Verzahnung von technologischen und sozialen Determinismen und die Ambivalenzen von Handlungsträgerschaft im „Constructive Technology Assessment“. In: Dolata, U.; Werle, R. (Hg.): Gesellschaft und die Macht der Technik. Sozioökonomischer und institutioneller Wandel durch Technisierung. Frankfurt a. M., New York, 83–106.
[152] Suchman, L. A. (2007): Human-Machine Reconfigurations. Plans and Situated Actions, 2. Auflage. Cambridge.
[153]Suchman, L. A. (1987). Plans and situated actions: The problem of human-machine communication. Cambridge.

Technikphilosophie und -soziologie haben die zunehmende Komplexität der Mensch-Maschine-Relation in unterschiedlichen Theorien gedeutet und zugespitzt. In der Technikphilosophie wird Technik häufig nicht mehr als Ensemble technischer Objekte verstanden, sondern als Medium, mit dem sich menschliches Handeln und Verhalten vollzieht. Während die einzelnen Elemente dieses Mediums instrumentellem Zweck-Mittel-Denken entstammen, stelle ihre Gesamtheit eine *Zweite Natur* dar, die Randbedingungen und Erfolgsbedingungen für weiteres menschliches Leben setzt und auch Weltsicht und das Problemlösen beeinflusst.[154] Als die bereits technologisch orientierten Menschen, zu denen sie im Rahmen vieler Technisierungsprozesse geworden sind, werden sie zum Beispiel dazu neigen, Herausforderungen von Kommunikation oder Sicherheit als Probleme anzusehen, die primär technologisch zu lösen sind. Somit ist neue Technologie oft bereits das Ergebnis einer technologischen Art und Weise, wie Menschen die Welt sehen und sich zu ihr in Beziehung setzen.

Techniksoziologisch sind hier vor allem Ansätze der Ko-Evolution von Technik und Gesellschaft zu nennen.[155] In ihnen sind die sozialkonstruktivistischen Motive der Gestaltung aufgenommen, jedoch wird ihnen die vielfältige Rückwirkung einmal entwickelter und verfügbarer Technik auf Mensch und Gesellschaft zur Seite gestellt, zu denen beispielsweise die großen Infrastruktursysteme – wie jene der Mobilität – geeignetes Illustrationsmaterial liefern. Zunächst gestaltet nach Zweck-Mittel-Erwägungen unter Berücksichtigung vielfältiger gesellschaftlicher Belange beispielsweise aus Wirtschaft, Bürgerschaft und Umweltschutz, werden sie nach Fertigstellung zu Randbedingungen menschlicher Entscheidungen, zum Beispiel in Bezug auf die Wahl des Wohnortes oder die Ansiedlung von Betrieben. Die Akteur-Netzwerk-Theorie[156] sowie unterschiedliche Sichtweisen innerhalb der Technikphilosophie[157] haben die Gedanken von Ko-Konstruktion und Ko-Evolution erweitert und teils die technischen Objekte aufgrund ihres Einflusses auf den Menschen als Ko-Akteure (Aktanten) definiert (siehe unten).

---

[154] Hubig, C. (2006): Die Kunst des Möglichen I. Technikphilosophie als Reflexion der Medialität. Bielefeld.
[155] Rip, A. (2007): Die Verzahnung von technologischen und sozialen Determinismen und die Ambivalenzen von Handlungsträgerschaft im „Constructive Technology Assessment". In: Dolata, U.; Werle, R. (Hg.): Gesellschaft und die Macht der Technik. Sozioökonomischer und institutioneller Wandel durch Technisierung. Frankfurt a. M., New York, 83–106.
[156] Latour, B. (2007): Eine neue Soziologie für eine neue Gesellschaft. Einführung in die Akteur-Netzwerk-Theorie. Frankfurt a. M.; bzw. die englische Version: Latour, B. (2005): Reassembling the Social. An Introduction to Actor-Network-Theory. Oxford.
[157] Hubig, C. (2006): Die Kunst des Möglichen I. Technikphilosophie als Reflexion der Medialität. Bielefeld; Ihde, D. (1990): Technology and the lifeworld. Bloomington.

4.3 Mehrstufige Mensch-Technik-Wechselwirkungen

Die zunehmende Komplexität der Mensch-Technik- bzw. Mensch-Maschine-Relation verändert auch deren Wahrnehmung. Digitale Technik, insbesondere KI-gesteuerte Systeme wie Produktionsroboter, „autonome" Fahrzeuge, Therapieprogramme oder Schachcomputer sind Beispiele, in denen die vormals klaren Unterscheidungen von Mensch und Technik weniger eindeutig zu werden scheinen. Androide Roboter erscheinen menschenähnlich, Hilfesuchende interagieren mit Therapieprogrammen, als ob es sich um menschliche Therapiekräfte handeln würde, und der Schachcomputer scheint die Partie gewinnen „zu wollen". Die Anthropomorphisierung digitaler Technik ist in der Umgangssprache weit fortgeschritten. Sie zeigt sich darin, dass KI und Robotern Fähigkeiten wie Denken, Lernen, Entscheiden oder Emotionalität zugeschrieben werden, wodurch sie scheinbar in die Gemeinschaft der denkenden, lernenden, entscheidenden und fühlenden Menschen aufgenommen werden.

Phänomenologisch geht damit einher, dass sich durch „autonom" werdende KI-gestützte Technik Subjekt-Objekt-Verhältnisse zwischen Mensch und Technik verändern. Im traditionellen Bild gestalten und nutzen menschliche Subjekte technische Objekte. Bereits mit traditioneller Software, mehr noch mit KI, kommt es jedoch zu neuen Konstellationen. In vernetzten Systemen haben Menschen teils die Subjekt-, teils aber auch die Objektrolle inne. Wenn einerseits Entscheidungen über Menschen an Softwaresysteme delegiert werden, beispielsweise hinsichtlich der Gewährung von Krediten oder Sozialleistungen, werden Menschen zu Objekten der „Entscheidungen" dieser Systeme, die hier auftreten, als ob sie Subjekte seien. Andererseits kann die Subjektrolle von Menschen durch gute Software zur Entscheidungsunterstützung erhöht werden, beispielsweise wenn diese qualitativ hochwertige, diskriminierungsfreie und nachvollziehbare Informationen liefern, welche die Qualität menschlicher Entscheidungen und deren Begründbarkeit verbessern. Verschiebungen in den Subjekt-Objekt-Rollen zwischen Mensch und Technik müssen daher differenziert betrachtet werden. Sie hängen einerseits vom Ausmaß und diversen technischen und organisationalen Details ab; andererseits – und dies ist von besonderer ethischer Relevanz – manifestieren sie sich bei verschiedenen Personengruppen auf unterschiedliche Weise.

Die Gestaltung der Software und der damit operierenden Maschinen gibt jeweils die Alternativen vor, innerhalb derer Menschen handeln können. Optionen, die in dem Design nicht vorge-

sehen sind, werden ausgeschlossen. Algorithmen und Maschinen regulieren somit menschliches Handeln.[158] Derartige Prozesse finden in der Digitalisierung seit Jahrzehnten statt, werden aber durch KI-Systeme verschärft. Menschliche Akteure erleben dadurch einerseits eine Verminderung ihrer Autorschaft über das eigene Handeln und fühlen sich zunehmend eingeschränkt und fremdbestimmt. Andererseits werden KI-Systeme zielgerichtet von Akteuren eingesetzt, um die eigenen Handlungsmöglichkeiten zu erweitern. Ein Beispiel hierfür sind ADM-Systeme, die Klassifikationen und Prognosen vornehmen und beispielsweise durch das Errechnen von Risikoscores Menschen bei der Entscheidungsfindung unterstützen (vgl. Beispiele in Teil II). Es ist immer wieder ein erwünschter Effekt erweiterter Autorschaft, wenn menschliche Entscheidungen dadurch auf eine sachlichere Grundlage gestellt werden. Auf der anderen Seite jedoch droht das Risiko, dass Menschen den Ergebnissen der KI-Systeme, auch wenn diese nur als Vorschläge unterbreitet werden, einfach blind folgen. Dann würde die Person, die eine Entscheidung trifft, eher *reagieren* als aus eigener Einsicht heraus *agieren*, was ihre Autorschaft vermindern würde (*automation bias*). Neben den bereits vielfach diskutierten Herausforderungen an eine transparente und rechtssichere Zuschreibung von Verantwortung in komplexen Mensch-Technik-Systemen, etwa beim „autonomen" Fahren[159], ist von anthropologischer und ethischer Relevanz, inwieweit die digitalen Systeme Menschen unterstützen und dadurch die Möglichkeit der Entfaltung der menschlichen Fähigkeiten erweitern oder durch technische, von den Herstellern der Systeme vorgegebene Handlungsschemata diese Entfaltung behindern und vermindern.

Die angedeutete Komplexität neuer Mensch-Maschine-Wechselwirkungen und die erwähnten Verschiebungen in Subjekt-Objektrollen haben nicht nur in der öffentlichen Debatte, sondern auch in der Wissenschaft zu einer Aufweichung eines strikten Dualismus zwischen Mensch und Maschine geführt. So wird vor allem in der Wissenschafts- und Technikforschung (Science and Technology Studies) sowie der Techniksoziologie seit Jahrzehnten über die „Handlungsträgerschaft" von technischen Systemen diskutiert. Einige dieser Positionen, insbesondere die Akteur-Netzwerk-Theorie[160], die einen sehr schwachen Handlungsbegriff propagiert und diesen auf viele Entitäten ausweitet, stehen dabei in deutlicher Spannung mit der im vorigen Kapitel

[158] Orwat, C. et al. (2010): Software als Institution und ihre Gestaltbarkeit. In: Informatik-Spektrum 33, 626-633, 626.
[159] Ethik-Kommission (2017): Automatisiertes und vernetztes Fahren. Endbericht. Berlin. https://www.bmvi.de/SharedDocs/DE/Publikationen/DG/bericht-der-ethik-kommission.pdf?__blob=publicationFile [18.01.2023].
[160] Latour, B. (2005): Reassembling the Social. An Introduction to Actor-Network-Theory. Oxford; Law, J.; Hassard, J. (1999): Actor Network and After. Oxford.

ausführten philosophischen Handlungstheorie, die einen anspruchsvollen Handlungsbegriff beschreibt und diesen auf menschliche Akteure beschränkt. Jenseits disziplinärer Einzelperspektiven stellen sich in Anbetracht der zunehmenden Verschränkung und wechselseitigen Beeinflussung von Menschen und Maschinen unter anderem folgende Fragen:

- Reicht das technische Vokabular zur Beschreibung von KI-Systemen noch aus, um Phänomene komplexer Interaktionen von Mensch und Maschine zu beschreiben?
- Inwiefern kann bzw. sollte davon gesprochen werden, dass Maschinen handeln oder mithandeln können? Sollte also Maschinen die Rolle von Akteuren zugesprochen werden und wenn ja, unter welchen Bedingungen und mit welchen Implikationen?
- Kommt es zu Verschiebungen in den Möglichkeiten menschlicher Autorschaft und wenn ja, in welchen Richtungen?

Die Akteur-Netzwerk-Theorie beantwortet diese Fragen, indem sie technischen Systemen den Status von Akteuren mit eigenen Dynamiken zuspricht und von hybriden Handlungszusammenhängen zwischen Mensch und Maschine ausgeht.[161]Die Akteur-Netzwerk-Theorie als Beobachtungstheorie ohne spezifische normativ-anthropologische Prämissen hilft, vermeintlich autonome Wirkungen von Techniken, Artefakten oder Sachen und deren gesellschaftsverändernde Potenzial zu erkennen. So kann es gelingen, Phänomene aus der Beobachterperspektive in den Blick zu nehmen, die mit einem starken, die Autonomie betonenden Handlungsbegriff tendenziell ausgeblendet werden. Freilich bleibt die Frage nach dem Zusammenhang zwischen Handlung und Verantwortung hier offen.

Der Techniksoziologe Werner Rammert, einer der Pioniere der Thematisierung möglicher Handlungsträgerschaft von Technik, schlägt einen Mittelweg zwischen einer anspruchsvoll normativen Vorstellung von Handeln und der Vorstellung eigenmächtigen maschinellen Agierens vor. Stattdessen soll von einer verteilten Handlungsträgerschaft zwischen Mensch und Maschine gesprochen werden, um die Vorstellung zu vermeiden, Technik sei etwas außerhalb des Sozialen Stehendes.[162] So geht Rammert wie die Akteur-Netzwerk-Theorie von hybriden, sozio-technischen Konstellationen aus, in denen Menschen und Maschinen auf komplexe Weise wechselwirken. Das Handeln des Menschen in dieser Perspektive sieht er zwar von technischen Prozessen beeinflusst, jedoch nicht als determiniert an. Der Einfluss der Technik kann sich in

[161] Latour, B (2007): Eine neue Soziologie für eine neue Gesellschaft. Einführung in die Akteur-Netzwerk-Theorie. Frankfurt a. M.
[162] Rammert, W.; Schulz-Schaeffer, I. (2002): Technik und Handeln- wenn soziales Handeln sich auf menschliches Verhalten und technische Artefakte verteilt. In: Dies. (Hg.): Können Maschinen handeln? Frankfurt a. M., New York, 11-64.

beiden Richtungen auswirken: Individuelle Freiheitsspielräume und die Entfaltung der menschlichen Autorschaft des eigenen Lebens können sowohl erweitert als auch vermindert.

Vor diesem Hintergrund schlägt Rammert vor, die Wechselwirkung von Mensch und Technik dreistufig zu beschreiben, um sowohl die Komplexität dieser Wechselwirkungen empirisch zu erfassen als auch die Zuschreibung von Verantwortung auf Menschen zu begrenzen. Als Stufe 1 nennt er *Kausalität* im Sinne von, Veränderung bewirken zu können. Stufe 2 beschreibt er als *Kontingenz* mit der Bedeutung, auch anders agieren zu können. Stufe 3 schließlich ist durch *Intentionalität* gekennzeichnet, was beinhaltet, das eigene Verhalten deuten und steuern zu können. Kausalität und Kontingenz charakterisieren nach Rammert nicht nur die menschliche Intervention, sondern auch die von Technologien. Algorithmen „wählen" zwischen Alternativen; automatisierte Entscheidungssysteme „bestimmen" einen Risikoscore, auf dessen Basis entweder ein Mensch eine Entscheidung trifft, beispielsweise über eine Kreditvergabe, oder die Software „entscheidet" sogar selbst, welche Bewerbungen bereits vorab aussortiert und der Personalabteilung erst gar nicht angezeigt werden. Algorithmen wirken hier auf mannigfaltige und höchst komplexe Weise, ohne dass ihnen dafür Intentionalität unterstellt werden kann. Intentionalität nämlich, die höchste Stufe des „Agierens", ist dem Handeln von Menschen vorbehalten. Daher kann nach Rammert nur ihnen Verantwortung zugeschrieben werden.

Die genannten Schwierigkeiten bei der Zuschreibung von Verantwortung haben auch den Technikphilosophen Luciano Floridi motiviert, eine Theorie „verteilter Moralität"[163] zu entwickeln, die auf die Komplexität digitaler Mensch-Maschine-Systeme zugeschnitten ist. Basierend auf früheren Arbeiten[164] unterscheidet er zwischen moralischer Handlungsfähigkeit (*moral agency)* und moralischer Verantwortlichkeit (*moral responsibility)*.[165] Während er – wie Rammert – Intentionalität als notwendig für Verantwortlichkeit erachtet, ist sie keine Bedingung für Handlungsfähigkeit. Für Letzteres reichen Interaktivität, Autonomie und Adaptivität aus. KI-Algorithmen bzw. verteilte Mensch/KI-Systeme können demnach für ihre „Handlungen" zwar *accountable*, aber nicht *responsible* sein, da ihnen die Intentionalität fehle. Die moralischen Eigenschaften der Ergebnisse verteilter Handlungsträgerschaft *emergieren* aus den einzelnen Elementen, ohne dass ihnen eine Intention zugrunde liegt.

---

[163] Floridi, L. (2012): Distributed morality in the information society. In: Science and Engineering Ethics 19(3), 727-74. (DOI: 10.1007/s11948-012-9413-4).
[164] Floridi, L.; Sanders, J. W. (2004): On the morality of artificial agents. In: Minds and Machine 14, 349-379.
[165] Da die üblichen deutschen Übersetzungen nicht immer genau zu passen scheinen, sind hier die englischen Originalbegriffe erhalten.

Auch der Versuch von Floridi, die Interaktionen von Mensch und Maschine angemessen zu beschreiben, lässt die Frage der Zuschreibung von Verantwortung offen.[166] Es gibt moralisch bedeutsame Aktionen von KI-Systemen, insofern als moralisch problematische Resultate durch KI-Systeme verursacht werden. Den Systemen kann dafür aber keine Verantwortung zugeschrieben werden. Verantwortung muss daher anders geregelt werden, zum Beispiel durch Zuschreibung von Verantwortung an relevante Institutionen bzw. Organisationen. Dies können beispielsweise die Betreiber dieser Systeme sein, da sie aufgrund ihrer intentionalen Entscheidung zu deren Einsatz einschließlich der diffusen Verteilung von *agency* zugestimmt haben und damit – auch im Sinne von Floridi – verantwortlich sind. Der von der Europäischen Union geplante Artificial Intelligence Act (AI Act) wird folglich die Verantwortlichkeit konkret zuweisen müssen, um wirksam zu regulieren.

Die hier vorgestellten techniksoziologischen und -philosophischen Ansätze, die sich um ein differenziertes Verständnis der zunehmend komplexen Mensch-Technik-Wechselwirkungen in Bezug auf KI-Systeme bemühen, können mit den anthropologischen Positionen, die in Kapitel 3 dargelegt wurden, durchaus in Konflikt geraten. Entsprechend der dort ausgeführten anspruchsvollen philosophischen Handlungstheorie können Maschinen nicht handeln, weil sie nicht über Intentionen verfügen, und kommen daher als genuine Akteure nicht infrage, jedenfalls nicht nach gegenwärtigem Stand der Entwicklung. Aber auch, wenn technischen Systemen keine Handlungsfähigkeit und damit Verantwortung zugeschrieben werden kann, haben sie Einfluss auf menschliches Handeln. Menschliches Handeln ist weder völlig autonom noch völlig sozial oder technisch determiniert, sondern in zunehmendem Maß soziotechnisch situiert. Auch in der Digitalisierung und der KI ist dies empirisch durch zahlreiche Studien belegt, insbesondere zum sogenannten *Automation Bias* bzw. zu den Effekten von *Nudging*.[167] Diese Effekte sind höchst bedeutsam für ethische Fragen, zeigen sie doch, dass technologische Entwicklungen menschliche Handlungsfähigkeit beeinflussen und menschliche Autonomie und Autorschaft sowohl erweitern als auch vermindern können.

---

[166] Floridi, L. (2016): Faultless responsibility: on the nature and allocation of moral responsibility for distributed moral actions. In: Philosophical Transaction A 374: 20160112 (DOI: 10.1098/rsta.2016.0112).

[167] Unter Nudging versteht man die Formatierung einer Entscheidungssituation ohne an den Handlungsalternativen etwas verändern, sodass erwünschtes Verhalten wahrscheinlicher wird. Vgl. Thaler, R. H.; Sunstein, C. R. (2018): Nudge. Wie man kluge Entscheidungen anstößt. Übersetzt von Bausum, C. 13. Auflage. Berlin.

+++ Vorabfassung! Wird durch die lektorierte Fassung ersetzt. +++

4.4 Erweitern und Vermindern menschlicher Autonomie und Autorschaft

Angesichts der konstatierten intensiven und komplexen Wechselwirkungen zwischen Mensch und Technik stellt sich die Frage nach den Folgen des digital-technischen Fortschritts für die Bedingungen gelingenden Handelns und die Möglichkeiten, menschliche Autorschaft zu entfalten. Bisherige Erfahrungen zeigen Ambivalenzen und Dialektiken von ethischer Relevanz.[168] Neue Technologien sollen einerseits und vor allem, so die Erzählung spätestens seit Francis Bacon und der Europäischen Aufklärung, die Menschen von den Zwängen der Natur und der Tradition emanzipieren, Freiheitsräume durch neue Handlungsoptionen eröffnen und damit die Entfaltungsmöglichkeiten menschlichen Handelns erweitern. Entsprechende Effekte zeigen sich in der Tat im raschen Fortschritt der digitalen Technologien: globale Kommunikation in Echtzeit, schnelle Information, Mustererkennung durch Big Data, Effizienzsteigerung und Beschleunigung der Produktion, neue Dienstleistungen und Geschäftsmodelle, bessere medizinische Diagnosen und Therapien, Roboter und Algorithmen als künstliche Assistenten, selbstfahrende Autos, Minenräumroboter und vieles mehr. Speziell die KI eröffnet Möglichkeiten menschliches Handeln zu verbessern, so etwa durch Mustererkennung in großen Datenmengen für medizinische oder behördliche Zwecke, durch darauf aufbauende verbesserte Prognosemöglichkeiten, zum Beispiel zur Ausbreitung von Infektionskrankheiten oder für Prognosen in der Polizeiarbeit (*predictive policing*) , durch neue Möglichkeiten individualisierter Information und Werbung, aber auch durch Anwendungen im Bildungsbereich. Technik ist zentraler Teil menschlichen Lebens und gesellschaftlicher Vollzüge zumindest in den industrialisierten Regionen der Welt geworden und hat in vielen Fällen eindeutig positive Folgen in dem Sinne gezeitigt, dass die Möglichkeiten menschlicher Autorschaft erweitert wurden – jedenfalls für den Teil der Erdbevölkerung, der, vor allem im Globalen Norden, Zugang zu ihren Vorteilen hat.

Im Rahmen der Diffusion von Technik und Innovationen in die Gesellschaft, ihrer Nutzung und Veralltäglichung kommt es jedoch häufig zu Sekundäreffekten, die als negativ wahrgenommen werden. Zu den nicht intendierten Folgen wie Umweltproblemen und sozialen Verwerfungen zählen auch Begrenzungen menschlicher Entfaltungsmöglichkeiten. In die Erweiterung menschlicher Autonomie und Autorschaft im technischen Fortschritt ist ihre simultane Verminderung oft entweder bereits eingeschrieben oder entwickelt sich empirisch kontingent. Erweiterung und Verminderung sind häufig ineinander verschränkt, betreffen jedoch meist unterschiedliche Beteiligte in unterschiedlicher Weise, so etwa diejenigen, die Entscheidungen

[168] Grunwald, A. (2022): Technikfolgenabschätzung- Eine Einführung. Baden-Baden.

treffen, und diejenigen, die von diesen Entscheidungen betroffen sind. Während die Facetten der Erweiterung offensiv kommuniziert werden und oft auch deutlich sichtbar sind, etwa aufgrund neuer Fähigkeiten von IT-Systemen, ist ihre Kehrseite, die nicht immer aber immer wieder auftretende simultane Verminderung menschlicher Entfaltungsmöglichkeiten, oft nicht gut erkennbar.

Verminderungen qualifizierter Handlungsformen und Entfaltungsmöglichkeiten im Rahmen der Nutzung von Technik können etwa in folgenden Richtungen auftreten:

(1) Entstehende Abhängigkeiten: Mit dem Erfolg von Technik sind moderne Gesellschaften von ihrem reibungslosen Funktionieren abhängig geworden. Dies beginnt individuell mit der Abhängigkeit von Computer und Auto und reicht gesellschaftlich bis hin zur vollständigen Abhängigkeit vom Funktionieren der Energieversorgung und der weltweiten Datenkommunikationsnetze. Die individuelle wie kollektive Abhängigkeit von digitalen Technologien, insbesondere vom Internet, durchzieht sämtliche gesellschaftlichen Prozesse: ohne Internet keine funktionierende Weltwirtschaft, keine Finanztransaktionen, Kollaps internationaler Logistikketten, Zusammenbruch der öffentlichen wie privaten Kommunikation. Entsprechend führt jede Übertragung von Zuständigkeiten beispielsweise durch den Einsatz von Software zur Entscheidungsunterstützung, zu einer gewissen Abhängigkeit. Abhängigkeit jedoch vermindert menschliche Entfaltungsmöglichkeit, da sie Sachzwänge zum Weiterbetrieb der technischen Systeme nach sich zieht und die Vulnerabilität der Gesellschaft-gegenüber technischem Versagen und intendierten Störungen (z. B. Hacking) steigert.

(2) Anpassungsdruck: Technik nötigt zur Anpassung. Das ist auf der Ebene konkreter technischer Objekte wie Maschinen trivial; es müssen beispielsweise Bedienungsanleitungen beachtet werden, um die Technik sachgerecht nutzen zu können. Digitale Technik jedoch reguliert und ändert subtil menschliches Handeln und Verhalten. Softwaresysteme steuern vielfach explizit oder implizit Verhalten.[169] So strukturieren privat geführte digitale Infrastrukturen die politische Kommunikation, sortieren Suchmaschinen die Welt mithilfe der von ihnen gesetzten Filter und strukturieren Online-Plattformen Geschäftsprozesse. Dahinter steht beispielsweise die schlecht durch Daten belegbare Befürchtung, dass menschliches Denken und Handeln durch fortschreitende Anpassung an Softwaresysteme allmählich nach deren Anforderungen und Vorgaben reguliert und immer stärker im technischen Sinn normiert werden könnte. Menschliche

[169] Bowker, G. C.; Star, S. L. (1999): Sorting Things Out – Classification and Its Consequences. Cambridge (MA); Nguyen, C. T. (2021): How Twitter gamifies communication. In: Lacey, J. (Hg.): Applied Epistemology. Oxford, 410-436 (DOI: 10.1093/oso/9780198833659.003.0017).

Autorschaft würde leise und unbemerkt, sozusagen durch allmähliche Gewöhnung, unkritische Übernahme algorithmischer Vorschläge (Automation Bias) und Anpassung an technische Voreinstellungen, ausgehöhlt.

(3) Verschließen von Optionen: Immer wieder werden mit dem Eröffnen neuer Handlungsspielräume andere, bis dato etablierte Optionen abgewertet oder ganz verschlossen. In der Innovationstheorie gilt dies als „schöpferische Zerstörung".[170] Dies ist einerseits der normale Gang von Transformation und Wandel. Andererseits aber stürzen Innovationen vorhandene Anerkennungs- und Wertstrukturen durch disruptive Effekte um und ziehen Gewinner wie auch Verlierer nach sich. Von den neuen Optionen profitieren häufig andere Personen und Gruppen als die, die dann im Verschließen der traditionellen Optionen zu Verlierern des Wandels werden. Zum Verschließen von Optionen menschlicher Entfaltung durch technischen Fortschritt führen unterschiedliche Mechanismen. So werden Infrastruktursysteme häufig faktisch machtförmig, indem sie Lebensformen außerhalb dieser Systeme benachteiligen oder unmöglich machen. Beispielsweise wird mittlerweile häufig die Nutzung eines Smartphones vorausgesetzt, um an bestimmten Lebensvollzügen teilnehmen zu können. Diese Form der Verschließung von Optionen kann verschiedene Bevölkerungsgruppen unterschiedlich treffen und Gerechtigkeitsprobleme mit sich bringen, wie dies zum Beispiel unter dem Stichwort digitale Spaltung (*digital divide*) diskutiert wird. Ein anderer Mechanismus besteht in allmählicher Gewöhnung als Folge der oben erwähnten Anpassung. Technik, gerade die Digitaltechnik, macht vielfach das Leben angenehm und komfortabel. Sobald Routinehandlungen in Beruf oder Freizeit daran adaptiert wurden, gehört diese Technik so zum Leben, dass es ohne diese Technik oft kaum noch vorstellbar ist. Alternative Optionen verschließen sich, vermeintliche Sachzwang-Argumente erwecken den Anschein der Alternativlosigkeit, sind jedoch nur Ausdruck der schleichend eingetretenen Pfadabhängigkeit durch allmähliche Anpassung und Gewöhnung.

Diese Mechanismen, in denen Optionen menschlichen Handelns sich verschließen, können im Rahmen der Diffusion neuer Technologien in die Anwendung auftreten, ohne dass Intentionen von Akteuren dahinterstehen. Es geht nicht um eine Verminderung menschlicher Autorschaft durch bewusste Delegation vormals menschlicher Handlungsvollzüge an KI-Systeme, sondern um Effekte, die schleichend und teilweise unbewusst durch Verhaltensänderungen entstehen. Das Ersetzen als Endpunkt des Delegierens vormals menschlich ausgeübter Tätigkeiten an tech-

[170] Schumpeter, J. A. (2018): Kapitalismus, Sozialismus und Demokratie (9. Aufl.). Tübingen, S. 113 ff.

nische Systeme erfolgt jedoch intentional. Es betrifft Funktionen und Tätigkeiten, die technomorph beschrieben und sodann von KI-gesteuerten Systemen übernommen werden können, im Idealfall funktionsäquivalent oder „besser".[171] Motivationen, menschliche Tätigkeiten durch KI-Systeme zu ersetzen, sind zum Beispiel die Effizienzsteigerung behördlicher Funktionen, die Kostensenkung in der industriellen Produktion, die Routinisierung diagnostischer Auswertungen in der Medizin oder die Ermöglichung automatisierter Überwachung in Echtzeit.

Bezogen auf die Möglichkeit menschlicher Autorschaft stellt sich die Ersetzung menschlicher Tätigkeiten zunächst als Resultat menschlicher Entscheidungen dar. So wird beispielsweise in Unternehmen oder Behörden aus unterschiedlichen Gründen die Entscheidung getroffen, bestimmte Tätigkeiten, die zuvor von Menschen durchgeführt wurden, an maschinelle Systeme zu übertragen. Dies können beispielsweise ADM-Systeme in unterschiedlichen Anwendungen sein, etwa in der Medizin, im Sicherheitsbereich oder im Sozialwesen. Diese Übertragung ist für sich genommen ein Ausdruck der Wahrnehmung menschlicher Autorschaft in bestimmten institutionellen Kontexten und unter entsprechenden Randbedingungen. Die zentrale ethische Frage ist, ob und wie diese Übertragung die Möglichkeiten *anderer* Menschen beeinflusst, vor allem jener, über die entschieden wird. Es stellt sich hier also die Frage, wie die Delegation von Tätigkeiten an Technik die Handlungsmöglichkeit und Autorschaft der Betroffenen beeinflusst. Dies stellt sich in unterschiedlichen Anwendungsfeldern auf je andere Weise dar.

Bereits mehrfach hat sich damit gezeigt, dass mit einer erwünschten Erweiterung der Möglichkeiten menschlicher Autorschaft oft simultan eine Verminderung verbunden ist, häufig in Bezug auf andere Aspekte und Felder von Autorschaft bzw. andere Menschen. Auf jeden Fall verschieben KI-Systeme die Möglichkeiten der Wahrnehmung menschlicher Autorschaft. Dies betrifft unterschiedliche Bevölkerungsgruppen auf unterschiedliche Weise und weist daher eine soziale Dimension mit ethischen Fragen auf. Bei der Betrachtung von Chancen und Risiken etwa von Entscheidungsunterstützungssystemen im Sicherheitsbereich ist zu berücksichtigen, *für wen* es sich hier jeweils um Chancen oder Risiken, um Erweiterungen oder Verminderungen der Autorschaft handelt. Damit sind hier Aspekte sozialer Gerechtigkeit und Macht involviert. In der Digitalisierung und speziell bei der Entwicklung und Nutzung von KI ist grundsätzlich zu fragen, wer die die entsprechenden Prozesse bzw. der konkreten Software-Applikationen und KI-Algorithmen jeweils gestaltet und ob – und wenn ja mit welcher Legitimation – sie in

[171] Das Wort „besser" suggeriert, dass es Verbesserung „als solche" gebe, und ignoriert, dass „besser" semantisch grundsätzlich nur im Zusammenhang mit normativen Kriterien des „besser" sinnvoll ist – und diese Kriterien können umstritten und kontrovers sein.

die Autonomie und Autorschaft derjenigen, die diese Produkte nutzen, oder weiterer Betroffener eingreifen. Auch jenseits der intentionalen Manipulation durch die Gestalter sind Effekte von Beeinflussung, Gewöhnung und Abhängigkeiten bis hin zu digitalem Mediensuchtverhalten zu beobachten, in denen offenkundig menschliche Autorschaft eingeengt wird.

Die sich hier andeutenden ethischen Herausforderungen sind mit epistemologischen Herausforderungen verbunden. Allmähliche Verschiebungen, wie etwa die erwähnten Gewöhnungsprozesse an technisch normierte Handlungsmuster (Automation Bias), sind oft nur schwer aufzudecken und empirisch zu belegen. Es besteht das Risiko verspäteter Entdeckung, zu einem Zeitpunkt, zu dem möglicherweise nur noch schwer beeinflussbare Pfadabhängigkeiten bereits eingetreten sind, schlimmstenfalls ein Point of no Return überschritten wurde. Zwischen der hohen Relevanz dieser an die Dialektik von Herr und Knecht gemahnenden Situation und der epistemologisch schwierigen Nachweislage ist die Bewusstmachung möglicher ethisch bedenklicher Zukunftsentwicklungen dieses Typs eine Herausforderung. Denn angesichts starker Gegenwartspräferenzen vieler Akteure ist sie mit den bekannten Problemen vorsorgeorientierter Kommunikation konfrontiert. Von der einen Seite droht der Vorwurf der Übertreibung, Dramatisierung oder gar Technikfeindlichkeit, von der anderen der Vorwurf der Verharmlosung. Hohe epistemologische Unsicherheit macht vorsorgeorientierte Kommunikation anfällig für Ideologie, interessengeleitete Statements und Spekulation.

4.5 Fazit

Die ethische Analyse und Beurteilung des Einsatzes von KI-Systemen bedürfen über die begriffliche, anthropologische und handlungstheoretische Vergewisserung (vgl. Kapitel 3) hinaus eines genaueren Blicks auf die sich mit der Digitalisierung verändernden Konstellationen zwischen Mensch und Technik. Im Rahmen der philosophischen Handlungstheorie können Maschinen nicht handeln und kommen als genuine Akteure mit Verantwortung nicht infrage. Dennoch haben sie Einfluss auf menschliches Handeln, das in modernen Gesellschaften in zunehmendem Maß soziotechnisch situiert ist. Erfahrung und empirische Forschung zeigen, dass Technik einerseits von Menschen als Mittel nach Zwecken gestaltet wird, dass aber andererseits neue Technik und darauf aufbauende Innovationen oder Dienstleistungen menschliches Handeln und Verhalten beeinflussen. Ethisch relevant ist insbesondere, wie sich diese Wechselwirkungen auf die Möglichkeiten menschlicher Autorschaft und Verantwortungsübernahme auswirken und wie diese sich angesichts der zunehmenden Verbreitung von KI-Systemen verändern.

KI-Systeme können in vielen Feldern menschliche Handlungen und Entscheidungen unterstützen, dadurch zu besseren Ergebnissen beitragen und damit menschliche Autorschaft erweitern. Vor allem die durch Algorithmen eröffnete Möglichkeit, in großen Datenmengen (Big Data) Muster zu erkennen, die den Menschen ansonsten verborgen wären, ist die Basis für den unterstützenden Einsatz der KI zum Beispiel in der medizinischen Diagnostik, im Bildungsbereich aber auch im Medienbereich und in der Verwaltung. Gerade im Bereich der Sozialen Medien zeigt sich hier ein Phänomen, das als *Hypernudge*[172] beschrieben wurde: Datenbasierte algorithmische Systeme kuratieren dynamisch hochgradig personalisierte Informationsumgebungen, denen man sich nur schwer entziehen kann. Menschliche Autorschaft kann also durch KI nicht nur erweitert, sondern auch vermindert werden, entweder durch intendierte Delegation von Entscheidungen an automatische Systeme oder durch allmähliche Effekte der Gewöhnung und Anpassung an datengenerierte Empfehlungen von KI-Systemen.

KI-Systeme verschieben die Möglichkeiten der Wahrnehmung menschlicher Autorschaft. Die grundsätzlich erwünschte Erweiterung von Autorschaft ist häufig simultan mit einer Verminderung in Bezug auf andere Aspekte von Autorschaft bzw. andere Akteure verbunden. Insbesondere sind verschiedene Akteursgruppen in unterschiedlicher Weise betroffen. Die ethische Analyse von Chancen und Risiken etwa von ADM-Systemen in der öffentlichen Verwaltung (vgl. Kapitel 8) muss darauf achten, *für wen* es zu Erweiterungen oder Verminderungen der Autorschaft kommt, etwa in der Differenz von Entscheidern und Betroffenen. Es sind also mit dem Einsatz von KI-Systemen auch Fragen von Gerechtigkeit und Autonomie- bzw. Machtverteilung involviert. Speziell ist zu fragen, ob und wie die jeweiligen Gestalter der entsprechenden Prozesse bzw. der konkreten Software-Applikationen und KI-Algorithmen in die Autonomie und Autorschaft derjenigen eingreifen, die diese Produkte nutzen oder anderweitig von ihnen betroffen sind.

Weiterhin sind psychologische Effekte zu beachten, die spezifisch für digitale Instrumente und insbesondere für KI-Systeme sind. Hier ist vor allem der Automation Bias zu nennen. Menschen vertrauen, so empirische Untersuchungen, algorithmisch erzeugten Ergebnissen und automatisierten Entscheidungsprozeduren häufig mehr als menschlichen Entscheidungen. Vermutlich spielen dabei verbreitete Objektivitätsunterstellungen gegenüber Daten und Rechenverfahren eine Rolle, während menschliche Urteile tendenziell als subjektiv wahrge-

[172] Yeung, K. (2017): 'Hypernudge': Big Data as a mode of regulation by design. In: Information, Communication and Society 20 (1), 118-136 (DOI: 10.1080/1369118X.2016.118671).

nommen werden. Gerade bei Entscheidungen, die mit einer großen prognostischen Unsicherheit konfrontiert sind und zugleich gravierende Auswirkungen haben, besteht die latente Tendenz, den datenbasierten algorithmischen Auswertungen mehr zu vertrauen. Damit wird Verantwortung – zumindest unbewusst – auf diese „Quasi-Akteure" delegiert. Dieser Bias zugunsten der algorithmischen Verfahren kann beispielsweise dazu führen, dass auch bei einer handlungstheoretisch korrekten Organisation von Entscheidungsprozessen, in denen ein KI-System normativ strikt auf die Rolle der Entscheidungsunterstützung begrenzt und dem Mensch, der die Entscheidung trifft, die Verantwortung zugeschrieben wird, das KI-System allmählich in die Rolle des eigentlichen „Entscheiders" gerät und menschliche Autorschaft und Verantwortung ausgehöhlt werden. Bisweilen wird versucht, dieser Gefahr vorzubeugen, indem bei Verwendung eines Entscheidungsunterstützungstools ein entsprechender Warnhinweis gegeben wird. Eine weitere denkbare Vorkehrung wäre die Verpflichtung der entscheidenden Fachkräfte, die etwaige Übernahme des algorithmischen Entscheidungsvorschlages – etwa mit Verweis auf die eigene erfahrungsbezogen intuitive oder kollegial erörterte Prognose – ausdrücklich zu begründen. Auf jeden Fall bedarf dieser Überlappungsbereich normativer Regulierung und empirisch-psychologischer Effekte besonderer Aufmerksamkeit in den ethischen Analysen zu den Anwendungsfeldern in den folgenden Kapiteln.

Menschliche Entscheider haben kaum eine Möglichkeit, die epistemische Evidenz der Korrelationen und Muster kritisch zu beurteilen, sondern sind vielfach darauf angewiesen, sie so zu nehmen, wie sie von den Systemen bereitgestellt werden. Sie unterliegen damit einem verborgenen Nudging durch die Art und Weise, wie die KI-Systeme zu ihren Ergebnissen kommen, und werden in bestimmte Richtungen des Entscheidens gedrängt. Mögliche Einseitigkeiten, etwa auf Basis der Datenlage, sowie daraus möglicherweise resultierende Diskriminierungen geraten aus dem Blick und menschliche Autorschaft wird entleert.

In den Abwägungen zwischen Erweiterung und Verminderung menschlicher Autorschaft in ihrer sozialen Verteilung sind bereits auf einer abstrakten ethischen Ebene mehrere Dimensionen zu berücksichtigen. Erstens bedarf die Übertragung menschlicher Tätigkeiten auf KI-Systeme der Transparenz gegenüber den Betroffenen. Sie sollten darüber informiert sein, auf welche Weise Entscheidungen zustande kommen, von denen sie dann betroffen sind. Dies hat zweitens mit der Klarstellung von Verantwortungszuschreibungen zu tun. Um ein Verantwortungs- und gegebenenfalls auch Haftungsvakuum zu verhindern, muss die Verantwortungszuschreibung etwa über die Betreiber der Systeme oder die menschlichen Akteure, die die Übertragung an sie beschlossen haben, geregelt werden. Drittens bedarf es der Sicherstellung der Nachvollziehbarkeit in Bezug auf das zweckhafte Funktionieren der KI-Systeme. Viertens müssen mögliche

nicht intendierte Folgen wie beispielsweise schleichend einkehrende Abhängigkeiten von den KI-Systemen oder allmähliche Aushöhlung menschlicher Autorschaft sorgfältig beobachtet werden, um gegebenenfalls rechtzeitig korrigierend eingreifen zu können.

+++ Vorabfassung! Wird durch die lektorierte Fassung ersetzt. +++

TEIL II: AUSGEWÄHLTE ANWENDUNGEN UND SEKTORSPEZIFISCHE EMPFEHLUNGEN

In den folgenden Kapiteln des zweiten Teils dieser Stellungnahme sollen die zuvor angestellten Überlegungen anhand von Analysen in vier verschiedenen Sektoren vertieft werden: dem Bereich der Medizin (Kapitel 5), dem Bereich der (schulischen) Bildung (Kapitel 6), dem Bereich der öffentlichen Kommunikation und Meinungsbildung (Kapitel 7) sowie der öffentlichen Verwaltung (Kapitel 8).

KI-basierte Software-Systeme werden in allen vier Sektoren eingesetzt, wobei sich die Durchdringungsbreite und -tiefe durchaus unterscheiden. So liefern die für den Bereich der öffentlichen Kommunikation und Meinungsbildung zentralen Sozialen Medien ein Paradebeispiel einer sehr umfassenden Delegation vormals menschlicher Tätigkeiten an Algorithmen, beispielsweise zu Zwecken der Kuratierung und Moderation von Inhalten. Der Bereich der schulischen Bildung bildet den anderen Pol, gibt es hier doch eine vergleichsweise geringe Nutzung digitaler Technologien im Allgemeinen und von KI-Systemen im Speziellen, eine vollständige Ersetzung menschlicher Tätigkeiten scheint in weiter Ferne. Im Bereich der Medizin hingegen werden zunehmend einzelne Tätigkeiten oder gar ganze Funktionen an KI-basierte Softwaresysteme delegiert. Dies reicht vom Einsatz von KI-basierter Mustererkennung zu Zwecken der Krebsdiagnostik bis hin zur Verwendung von Chatbots in der Therapie, die schon einen möglichen Ersatz menschlichen Fachpersonals suggerieren. Auch in die öffentliche Verwaltung haben KI-basierte Softwaresysteme Einzug gehalten, insbesondere in Form von datenbasierter Software, die zur Erstellung von Risikoprofilen und zur Entscheidungsunterstützung herangezogen wird.

Die folgenden Analysen werden zeigen, dass Fragen des richtigen Ausmaßes der Delegation von Tätigkeiten und Funktionen an Softwaresysteme nur kontext-, anwendungs- und personenbezogen spezifiziert werden können. Dabei gilt als Richtschnur der Bewertung, ob die Delegation zu einer Erweiterung der Handlungsmöglichkeiten, insbesondere der Möglichkeiten für verantwortliches Handeln und Autorschaft der verschiedenen betroffenen Akteure führt oder ob es in Folge möglicherweise zu einer Verminderung solcher Handlungsmöglichkeiten sowie zu negativen Auswirkungen auf Möglichkeiten der Autorschaft und Verantwortungsübernahme kommt. Die Analyse des Einsatzes von KI-Systemen in den vier Sektoren enden jeweils in sektorspezifischen Empfehlungen. Unabhängig davon lassen sich aus dem Vergleich der Sektoren auch übergreifende Themen ausmachen, Aspekte, die sich in den vier Sektoren mal ähnlich, mal aber sehr unterschiedlich darstellen. Diese Themen werden in Teil III „Querschnittsthemen und übergreifende Empfehlungen“ aufgegriffen.

5 Medizin

5.1 Einleitung

Die zunehmende Durchdringung unserer Lebenswelt mit digitalen Produkten, die über KI-Komponenten verfügen, breitet sich auch im Gesundheitssystem immer weiter aus. Angesichts der Dynamik der technischen Entwicklung und der ökonomischen Triebkräfte dieses Transformationsprozesses ist es aus ethischer Perspektive erforderlich, die damit verbundenen Herausforderungen möglichst präzise wahrzunehmen, die jeweiligen Vor- und Nachteile des Einsatzes von KI-Instrumenten differenziert abzuwägen und einer problematischen Verselbstständigungstendenz des Technikeinsatzes frühzeitig entgegenzuwirken. Zwar gibt es bereits eine Vielzahl wertvoller ethischer Empfehlungen und Richtlinien, die sich auf unterschiedlichen Ebenen mit den Bedingungen eines verantwortbaren KI-Einsatzes beschäftigen[173], doch wird dabei entweder ein sehr genereller Überblick über mehrere Handlungsfelder gegeben, um die Chancen und Risiken der Schaffung und Bewirtschaftung integrierter Datenräume mit neuen KI-Technologien auszuloten, oder es werden die einschlägigen Herausforderungen im Gesundheitsbereich nur aus einer begrenzten – etwa ärztlichen – Perspektive reflektiert, um die Handlungssicherheit einer bestimmten Personengruppe im Umgang mit den neuen Instrumenten zu verbessern.[174]

Die vorliegende Stellungnahme knüpft an diese Vorarbeiten an, indem sie nicht nur die vielfach erhobene Forderung nach einem „Vorrang menschlichen Handelns“ mit Blick auf die verschiedenen Formen der Mensch-Maschine-Relation auf diesem besonders sensiblen Handlungsfeld näher spezifiziert und in Bezug setzt zu den bekannten sektorenübergreifenden Herausforderungen (z. B. des Datenschutzes; der Qualität, Robustheit und Sicherheit der eingesetzten Instrumente angesichts verschiedener Manipulationsmöglichkeiten; der Bias- bzw. Diskriminierungsgefahren und anderes infolge einseitiger Trainingsdaten und der Nachvollziehbarkeits- bzw. Opazitäts-Probleme besonders komplexer Algorithmen – siehe Kapitel 2).

[173] Europäische Kommission (2020): Weißbuch zur Künstlichen Intelligenz – Ein europäisches Konzept für Exzellenz und Vertrauen. https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:52020DC0065&from=DE [09.01.2023]; High Level Expert Group on Artificial Intelligence (2019): Ethics Guidelines for Trustworthy AI. https://www.aepd.es/sites/default/files/2019-12/ai-ethics-guidelines.pdf [09.01.2023].
[174] Zentrale Ethikkommission (2021): Entscheidungsunterstützung ärztlicher Tätigkeit durch Künstliche Intelligenz. In: Deutsches Ärzteblatt, 118 (33-34) (DOI: 10.3238/arztebl.zeko_sn_cdss_2021).

Aufgrund der Komplexität des medizinischen Versorgungssystems bedarf es auch mit Blick auf die folgende genauere Betrachtung von KI-Anwendungen im Gesundheitswesen einer wenigstens dreifachen Differenzierung:

Erstens sind mehrere Akteursgruppen zu unterscheiden, die bezüglich eines KI-Einsatzes unterschiedliche Funktionen und Verantwortlichkeiten besitzen, wie zum Beispiel Personen in der Softwareentwicklung, Kontroll- und Zertifizierungsinstanzen, medizinische Forschungsgruppen, klinisch tätige Personen, die von ihnen versorgten Menschen, medizinische Fachgesellschaften sowie Träger von medizinischen Versorgungseinrichtungen.

Zweitens umfasst das Gesundheitswesen mit der Grundlagen- und (prä-)klinischen Forschung sowie der konkreten medizinischen Versorgung durch Prävention, Kuration und Palliation, denen wiederum verschiedene diagnostische, therapeutische und prognostische Einzelmaßnahmen zuzuordnen sind, ganz unterschiedliche Anwendungsbereiche für KI-Produkte, deren jeweilige Chancen und Risiken kontextsensitiv zu beurteilen sind.[175]

Drittens sind hinsichtlich der derzeit verfügbaren KI-basierten Instrumente unterschiedliche *Grade der Ersetzung* menschlicher Handlungssegmente zu beobachten, die im Kernbereich medizinischer Behandlungen ganz verschiedene Auswirkungen auf die Arzt-Patienten-Beziehung haben können und dadurch die Handlungsmöglichkeiten der verschiedenen Akteure sowohl erweitern als auch vermindern können[176]. Das Spektrum der Einsatzmöglichkeiten dieser Instrumente reicht dabei von einer engen Nutzung, bei der lediglich ein einzelnes Segment ärztlichen Handelns technisch substituiert wird, über komplexere KI-Anwendungen, die mehrere ärztliche Handlungsschritte begleiten und unterstützen, bis hin zu einer vollständigen Ersetzung des Behandlers durch ein KI-System. Letztere trägt zwar gegenwärtig noch weithin fiktionale Züge, ist aber in einzelnen Sektoren der medizinischen Versorgung bereits Realität geworden.

Nachfolgend werden anhand exemplarischer Beispiele die jeweiligen Chancen und Risiken eines KI-Einsatzes in unterschiedlichen Bereichen des medizinischen Systems analysiert.

5.2 Einsatz von. KI-Systemen in der Medizin

Eine verantwortliche Verwendung von KI-Systemen im Bereich der Medizin setzt voraus, dass die gesamte Handlungskette – von der Entwicklung entsprechender Produkte über ihren Einsatz

[175] Der Bereich der privaten Nutzung entsprechender Produkte im Rahmen einer *digital self care* bleibt hier ausgeklammert.
[176] Zum Einsatz KI-gestützter robotischer Systeme im Rahmen der Pflege vgl. Deutscher Ethikrat (2020): Robotik für gute Pflege. Berlin.

in der Forschung bis hin zu ihrer Implementierung in den verschiedenen Sektoren der medizinischen Versorgung – ethischen Standards genügt, kontinuierlich überwacht und gezielt so weiterentwickelt wird, dass Vorteile sukzessive immer besser genutzt und Gefahren vermieden werden.

5.2.1 Entwicklung von KI-Systemen

Bereits die Entwicklung geeigneter KI-Komponenten für die medizinische Praxis, die hier als erster, dem klinischen Einsatz solcher Systeme notwendig vorausgehender, ethisch relevanter Teilbereich erwähnt werden soll, stellt eine anspruchsvolle Aufgabe dar. Sie erfordert nicht nur eine enge interdisziplinäre Zusammenarbeit verschiedener Sachverständiger aus unterschiedlichen Fachgebieten (z. B. Informatik, Ingenieurswissenschaft, Medizin, Recht), um geeignete Algorithmen für die zuverlässige Bewältigung bestimmter Aufgaben definieren zu können, sondern stellt auch hohe Anforderungen an die Qualität der verwendeten Trainingsdaten, um vermeidbare Verzerrungen der Ergebnisse von vorneherein auszuschließen oder zumindest zu minimieren.

Eine Besonderheit von KI-Systemen, die auf mit maschinellem Lernen aus Daten gewonnenen Modellen basieren, besteht darin, dass bei manchen Systemen selbst diejenigen, die diese Instrumente entwickeln, aufgrund der enormen Komplexität der Datenverarbeitungsprozesse nicht mehr rekonstruieren können, wie bestimmte Resultate zustande gekommen sind, da die Eingaben mit hochgradig nichtlinearen und verteilten Prozessen verarbeitet werden. Der Prozess wird zur Black Box. Zwar kann eine solche Opazität auch darin ihren Ursprung haben, dass bestimmte Algorithmen urheberrechtlich geschützt sind, doch gibt es neben dieser von äußeren Faktoren bedingten und insofern kontingenten Unzugänglichkeit auch eine rein technisch bedingte Opazität, die sich je nach Kontext unterschiedlich auswirken kann. Während es in manchen Bereichen sinnvoll oder sogar notwendig sein kann, ein Höchstmaß an Erklärbarkeit der jeweiligen Resultate anzustreben (Explainable AI), wofür gegebenenfalls ein hoher technischer und finanzieller Aufwand erforderlich ist, dürfte es in anderen Bereichen ausreichen sicherzustellen, dass die Personen, die diese Systeme anwenden, deren Resultate stets einer eigenen Plausibilitätsprüfung unterziehen, um den Gefahren eines ungerechtfertigten blinden Vertrauens in die Technik (Automation Bias) zu entgehen.

Die Forderung nach Transparenz lässt verschiedene Grade zu, sodass je nach Anwendungsbereich zu prüfen ist, was aus welchen Gründen für wen in welchem Umfang und zu welchem Zweck erklärbar sein muss. Tatsächlich gibt es in der Medizin viele Beispiele nicht nur für den

Einsatz technischer Geräte, deren genaue Wirkmechanismen denjenigen, die sie in ihrer Berufspraxis anwenden, allenfalls ansatzweise durchsichtig sind, sondern auch dafür, dass bestimmte Interventionen auch dann sinnvoll geplant und durchgeführt werden können, wenn das kausale Wissen um konkrete Wirkmechanismen begrenzt ist. Angesichts des Technisierungsgrades der modernen Medizin ist es weder möglich noch erforderlich, dass die behandelnden Personen die internen Prozesse der von ihnen genutzten technischen Hilfsmittel stets im Detail durchschauen, solange diese Prozesse in ausreichendem Umfang zumindest durch geeignete Stellen nachvollzogen und damit überprüft werden können.

In derart gelagerten Fällen ist es umso wichtiger, mittels „geeigneter Prüf-, Zertifizierungs- und Auditierungsmaßnahmen sicherzustellen“[177], dass die jeweiligen Systeme technisch einwandfrei funktionieren und verantwortungsvoll eingesetzt werden. Die Zertifizierung vertrauenswürdiger KI-Systeme erstreckt sich insbesondere auf die Wahrung von Mindest- und anwendungsbezogenen Spezialanforderungen bezüglich der Autonomie und Kontrolle, der Fairness, der Transparenz, der Verlässlichkeit, der Sicherheit und des Datenschutzes.[178]

Sowohl die medizinischen Fachleute, die solche „intelligenten“ Medizinprodukte verwenden, als auch die von ihnen behandelten Personen müssen darauf vertrauen können, dass nur hinreichend geprüfte und nach möglichst international konsentierten Maßstäben zertifizierte KI-Produkte zum Einsatz kommen und von der jeweiligen Gesundheitseinrichtung vorschriftmäßig gewartet und gegen Manipulation geschützt werden. Dies gilt umso mehr, als die Lebensdauer von Software-Produkten oft relativ kurz ist und die Systeme durch häufige Updates einen insgesamt fließenden Charakter aufweisen.

Angesichts der Dynamik der technischen Entwicklung und der wirtschaftlichen Interessen derjenigen – nicht selten als Startups aus universitären Forschungseinrichtungen heraus gegründeten – Unternehmen, die solche Produkte entwickeln und vertreiben, bedarf es eines geordneten Zusammenwirkens von staatlichen Aufsichts- und Kontrollbehörden, der Entwicklung neuer bzw. ergänzter Leitlinien für deren Einsatz durch die jeweils zuständigen medizinischen Fach-

[177] Zentrale Ethikkommission (2021): Entscheidungsunterstützung ärztlicher Tätigkeiten durch Künstliche Intelligenz. In: Deutsches Ärzteblatt, 118 (33-34) (DOI: 10.3238/arztebl.zeko_sn_cdss_2021), A5.
[178] Vgl. beispielsweise das Whitepaper des Projekts „Zertifizierte KI“, Cremers et al. (2019): Vertrauenswürdiger Einsatz von Künstlicher Intelligenz. Handlungsfelder aus philosophischer, ethischer, rechtlicher und technologischer Sicht als Grundlage für eine Zertifizierung von Künstlicher Intelligenz. https://www.iais.fraunhofer.de/content/dam/iais/KINRW/Whitepaper_KI-Zertifizierung.pdf [09.01.2023].

gesellschaften sowie kontinuierlicher Anstrengungen zur weiteren Aus-, Fort- und Weiterbildung des medizinischen Personals, um einen verantwortlichen Umgang mit den neuen Techniken zu gewährleisten.

Von besonderer ethischer Brisanz sind jene KI-Systeme, bei denen selbst von den Personen, die das System entwickeln und programmieren nicht mehr vollständig nachvollzogen werden kann, wie ein bestimmtes Ergebnis zustande kommt. Zumindest insofern KI-Systeme medizinische Entscheidungsvorschläge mit schwerwiegenden Konsequenzen für das Überleben und grundlegende Aspekte der Lebensqualität unterbreiten, müssen deren grundlegende Funktionsweisen und Arbeitsprozesse erklär- sowie interpretierbar sein, um einen selbstbestimmten Einsatz zu gewährleisten. Lassen sich die anwendungsbezogenen Transparenzanforderungen – selbst unter Rückgriff auf Erkenntnisse der Forschungsdisziplin „Explainable AI" – nicht erfüllen, verbietet sich jedenfalls ein Einsatz in diesem Bereich der medizinischen Versorgung. Die Festlegung derartiger Bereiche erfordert einen breiten gesellschaftlichen Diskurs, in den insbesondere die sektorbezogenen Chancen und Risiken des Einsatzes intransparenter KI-Systeme einzubeziehen sind.

5.2.2 KI in der medizinischen Forschung

Da die Planung und Durchführung medizinischer Forschung besonders hohe Anforderungen an den reflektierten Umgang mit unterschiedlichen Wissensformen und die methodischen Fertigkeiten derer stellt, die konkrete Forschungsfragen entwickeln und sie durch geeignete Experimente auf empirisch überprüfbare Weise zu beantworten suchen, sind die in dieser Weise am Forschungsprozess beteiligten Personen grundsätzlich nicht durch KI-Systeme ersetzbar. Trotz dieser prinzipiellen Grenzen kann die gezielte Implementierung von KI-Elementen auch im Kontext der medizinischen Forschung in mehrfacher Hinsicht vorteilhaft sein, sofern der Schutz derjenigen, die an den Studien teilnehmen, und ihrer personenbezogenen (Gesundheits-)Daten gewährleistet ist. Dazu bedarf es gerade in kollaborativen Forschungsprojekten nicht nur präziser Bestimmungen, „wer unter welchen Bedingungen Zugang zu den Daten erhält, um einen transparenten Austausch bei gleichzeitiger Gewährleistung hoher Datenschutzstandards zu ermöglichen", sondern auch „klarer Vorgaben, unter welchen Umständen Probanden und Patienten Zugang zu ihren Daten haben können, wie sie modular und dynamisch in ihre Nutzung einwilligen können … und wie die Daten langfristig erhalten bleiben" können.[179]

[179] Deutscher Ethikrat (2017): Big Data und Gesundheit – Datensouveränität als informationelle Freiheitsgestaltung. Berlin, 97.

Das Spektrum eines sinnvollen KI-Einsatzes in der medizinischen Forschung ist weit: Neben hilfreichen Vor- und Zuarbeiten etwa bei Literaturrecherchen oder der Durchsuchung großer Datenbanken können KI-Instrumente einen Beitrag dazu leisten, neue Korrelationen zwischen bestimmten Phänomenen zu entdecken, deren Bedeutung für die medizinische Praxis jedoch durch gezielte Analysen überprüft werden muss. Drei Beispiele mögen dies verdeutlichen:

Das erste Beispiel betrifft den Einsatz Künstlicher Intelligenz bei der Vorhersage raum- und zeitabhängiger medizinrelevanter Prozesse, wie etwa der Ausbreitung eines Virus in der Epidemiologie. In diesen Fällen ist die Anwendung mathematischer Algorithmen[180] häufig die einzig effiziente Methode, um prospektiv zu relevanten Erkenntnissen zu gelangen. Da man meist nicht alle relevanten Parameter kennt, sind die Vorhersagen mit einem Unsicherheitsfaktor verbunden, der – ähnlich einer Wettervorhersage – umso größer wird, je weiter die Vorhersage in der Zukunft liegt und je feiner die zeitliche und räumliche Auflösung ist. Es können jedoch Szenarien unter verschiedenen Annahmen berechnet werden, sodass ein Bereich wahrscheinlicher Entwicklungen sinnvoll eingegrenzt werden kann.

Als zweites Beispiel sei der KI-Einsatz in der Vorhersage komplexer Molekülstrukturen genannt, der bei der Erforschung von Krankheitsmechanismen und der Entwicklung vieler Therapeutika eine große Rolle spielt. Hier werden Algorithmen genutzt, die vor allem auf der von DeepMind (einer Tochter des Google-Mutterkonzerns Alphabet) entwickelten Software AlphaFold basieren, um beispielsweise effizient Kandidaten für Impfstoffe gegen Infektionskrankheiten (wie COVID-19, aber auch darüber hinaus[181]) zu identifizieren. Die Algorithmen sagen die Struktur von Proteinen vorher und ermöglichen so einen deutlich schnelleren Fortschritt in der Entwicklung bestimmter Substanzen, als dies über traditionelle labortechnische Verfahren möglich ist. Selbstverständlich müssen die identifizierten Therapeutika-Kandidaten weiterhin im Labor und in der Klinik getestet werden. Damit ist trotz des zeitlichen Gewinns die Sicherheit im Vergleich zu traditionellen Methoden nicht eingeschränkt.

Das dritte Beispiel stammt aus der Krebsforschung, in der die Verarbeitung großer Datenmengen etwa aus der Analyse des Genoms, des Transkriptoms und des Epigenoms von immer größerer Bedeutung wird. Einerseits bietet die systematische Verknüpfung und Integration unterschiedlicher Daten die Chance, Zusammenhänge zwischen einer Vielzahl von Variablen zu erfassen und im Blick auf mögliche kausale Einflussfaktoren zu analysieren, um zum Beispiel

---

[180] Dies bezieht die Modellierung über Differentialgleichungen mit ein, welche nicht in klassischem Sinne als Künstliche Intelligenz bezeichnet wird.
[181] Higgins, M. K. (2021): Can We AlphaFold Our Way Out of the Next Pandemic? In: Journal of Molecular Biology 433(20):167093 (DOI: 10.1016/j.jmb.2021.16709).

eine bessere Aufteilung großer Personenkollektive in relevante Subgruppen (Stratifizierung) zu ermöglichen – und damit letztlich gegebenenfalls eine besser auf diese Gruppen zugeschnittene Behandlung. Andererseits stellt die gesteigerte Komplexität und Heterogenität der Datenbasis neue Anforderungen an die methodische Kompetenz der Forschenden. Auch in der präklinischen und klinischen Forschung steht und fällt der mögliche Nutzen von KI-Anwendungen daher letztlich „mit der Expertise und Integrität der Personen und Institutionen, die Daten generieren, auswählen, verknüpfen und interpretieren“ .[182]

Zwar hat die automatisierte Bearbeitung großer Datenmengen durch KI-Komponenten aufgrund ihrer gesteigerten Rechenleistung das Potenzial, „Korrelationen zwischen wesentlich mehr Faktoren schneller und besser zu entdecken und dabei auch neue Hypothesen über Wirkzusammenhänge zu entwickeln“[183], doch wäre es „ein Missverständnis zu glauben, dass mehr Daten auch *automatisch* zu mehr Wissen über kausale Effekte führen“[184]. Aufgrund der Differenz zwischen Korrelation und Kausalität bedürfen Ergebnisse maschineller Datenanalyse daher stets der unabhängigen Überprüfung und Validierung, um in der Fülle der gefundenen Korrelationen die jeweils relevanten Kausaleffekte zu identifizieren und damit den Umfang des therapierelevanten Kausalwissens zu erweitern.

Trotz aller gebotenen Vorsicht gegenüber manchen allzu euphorischen Heilsversprechen einer KI-gestützten Forschung bietet gerade der Bereich der Onkologie inzwischen reichhaltiges Anschauungsmaterial dafür, wie in der Forschung entwickelte KI-Instrumente sukzessive auch in der medizinische Versorgung von Patientinnen und Patienten zur Diagnostik und Therapie eingesetzt werden können; dies zeigen auch einige der nachfolgend vorgestellten Beispiele.

5.2.3 KI in der medizinischen Versorgung

Obwohl sich viele der KI-basierten Systeme momentan noch in der Entwicklungs- und Erprobungsphase befinden, gibt es in den verschiedenen Bereichen der medizinischen Versorgung bereits erste Erfahrungen mit dem Einsatz dieser neuen Instrumente, die sich unter anderem durch den Grad der technischen Ersetzung einzelner oder mehrerer Handlungssequenzen voneinander unterscheiden lassen.

---

[182] Deutscher Ethikrat (2017): Big Data und Gesundheit – Datensouveränität als informationelle Freiheitsgestaltung. Berlin, 67.
[183] Deutscher Ethikrat (2017): Big Data und Gesundheit – Datensouveränität als informationelle Freiheitsgestaltung. Berlin, 71.
[184] Deutscher Ethikrat (2017): Big Data und Gesundheit – Datensouveränität als informationelle Freiheitsgestaltung. Berlin, 70.

Enge Ersetzung Grundsätzlich können KI-Systeme in allen Segmenten medizinischer Versorgung eingesetzt werden. Die größte praktische Verbreitung dürften derzeit Entscheidungsunterstützungssysteme in der Diagnostik haben, die versuchen, mittels computergestützter Analyse verschiedener Parameter der Labordiagnostik, der Bildbearbeitung sowie der automatisierten Durchsicht von Patientenakten und wissenschaftlichen Datenbanken Entscheidungsprozesse zu modellieren und zu automatisieren. Da die Entwicklung und Nutzung von Entscheidungsbäumen dem medizinischen Personal schon seit Langem dabei hilft, „diagnostische Informationen vollständig zu erheben, Therapieentscheidungen präzise zu treffen und damit insgesamt die bestmögliche Behandlung sicherzustellen", scheint dieses algorithmische Vorgehen geradezu „prädestiniert für KI-Anwendungen"[185] zu sein. Ihr Einsatz steht daher in struktureller Kontinuität zur Nutzung früherer technischer Hilfsmittel der Entscheidungsfindung (z. B. bestimmter Expertensysteme).

Neu ist der Umstand, dass die KI-basierten Systeme aufgrund der gesteigerten Rechenkapazitäten sehr große Datenmengen und eine Vielzahl relevanter Parameter berücksichtigen können und die einzelnen Rechenoperationen den Anwenderinnen und Anwendern nur begrenzt zugänglich sind. So bieten vor allem Fortschritte in der Bilderkennung große diagnostische Potenziale, da durch sie zum Beispiel neue Möglichkeiten einer frühzeitigen Detektion, Lokalisation und Charakterisierung verschiedener pathologischer Veränderungen in der Gewebestruktur eröffnet werden. Das Spektrum der derzeitigen Nutzung reicht von der Untersuchung des Augenhintergrundes über die Analyse von Hautläsionen in der Dermatologie bis hin zu verschiedenen Bereichen der Onkologie, die insofern von besonderer medizinischer Bedeutung sind, als die Früherkennung maligner Tumoren essenziell für eine erfolgreiche Therapie ist. Ein Beispiel bildet hier der KI-Einsatz in der Brustkrebsdiagnostik.

Infokasten 2: Enge Ersetzung am Beispiel Brustkrebsdiagnostik

Brustkrebs (Mammakarzinom) ist mit jährlich rund 71 375Neuerkrankten die häufigste Krebserkrankung der Frau; nur rund 1% der Brustkrebserkrankungen treten beim Mann auf.[186] Neben den vielen Einzelschicksalen, die hinter diesen Zahlen stehen, ist die kollektive Betrachtung des Brustkrebses wegen seiner Häufigkeit sowie der hohen Therapiekosten, der resultierenden Arbeitsausfälle sowie schlechter Heilungschancen bei der Erkennung des Tumors in einem späten Stadium von allgemeiner gesellschaftlicher Relevanz und sozioökonomischer Bedeutung.

[185] Schlemmer, H.-P.; Hohenfellner, M. (2021): Chancen von KI in der Onkologie am Beispiel der individualisierten Diagnostik und Behandlung von Prostatakrebs. In: Zeitschrift für medizinische Ethik 67 (3), 309-326, 318.
[186] Zentrum für Krebsregisterdaten (2019): Brustkrebs (Mammakarzinom). https://www.krebsdaten.de/Krebs/DE/Content/Krebsarten/Brustkrebs/brustkrebs_node.html [03.03.2023].

Eine frühzeitige Erkennung und die richtige Diagnose sind entscheidende Faktoren, um die Überlebenschancen zu erhöhen.

Um die Früherkennung zu unterstützen, wurden in vielen Ländern, so auch in Deutschland, Mammografie-Screening-Programme implementiert. Für die Mammografie werden Röntgenstrahlen eingesetzt, die mittels digitaler Detektoren nachgewiesen werden. Diese Information wird zur Rekonstruktion zweidimensionaler Bilder genutzt, die direkt für die Auswertung zur Verfügung stehen. Üblicherweise übernimmt dies erfahrenes, radiologisch qualifiziertes Personal, aber schon seit Längerem wird daran gearbeitet, die fachärztliche radiologische Beurteilung mittels computergestützter Verfahren zu unterstützen. Die ersten sogenannten CADe- und CADx-Algorithmen (*computer-aided detection* – CADe – und *computer-aided diagnose* – CADx) konnten sich zwar nicht durchsetzen, da die Anzahl an falsch-positiven Ergebnissen zu hoch war.[187] Durch den Einsatz von Künstlicher Intelligenz in Form von maschinellem Lernen, insbesondere von Deep Learning Convolutional Neural Networks zur Erkennung von Läsionen, konnten die Ergebnisse allerdings deutlich verbessert werden. Besondere Herausforderungen sind hierbei, dass sehr große Datensätze für das Trainieren der Netzwerkmodelle benötigt werden und oft eine Anpassung an lokale Gegebenheiten notwendig ist. Internationale Studien zeigen, dass beispielsweise durch den Einsatz der KI die Anzahl falsch-positiver Ergebnisse um 5,7% (USA) und um 1,2% (Großbritannien) gesenkt werden konnte und bessere Ergebnisse erzielt wurden als bei einer Beurteilung durch medizinisches Fachpersonal allein.[188]

Einige kommerzielle Anbieter haben bereits Programme auf den Markt gebracht wie beispielsweise Siemens Healthineers mit der interaktiven Entscheidungsunterstützung syngo.Breast Care bei der Mammografie. Die kommerziell angebotenen Programme wie zum Beispiel die Software Transpara™ bieten insbesondere eine Scoring-Funktion, welche die Mammografiebilder in Bezug auf das Vorliegen verdächtiger Läsionen stratifiziert. Dies ermöglicht es dem radiologischen Fachpersonal, in kürzerer Zeit komplexe Datensätze zu analysieren, und unterstützt die Entscheidungsfindung mit Blick auf weitere Behandlungsschritte. Insbesondere bei unklaren Befunden bleibt das Einholen einer Zweitmeinung einer weiteren Expertin/eines weiteren Experten allerdings unerlässlich. Insgesamt befinden sich KI-gestützte Systeme auch für die Diagnostik von Brustkrebs noch in der Entwicklung.

Ein verantwortlicher Umgang mit solchen Diagnose-Tools erfordert eine sorgfältige Abwägung der verschiedenen Chancen und Risiken, die mit ihrem Einsatz sowohl für die behandelnden Ärztinnen und Ärzte selbst als auch für die Betroffenen verbunden sind.

Aus *ärztlicher* wie auch aus Betroffenen-Perspektive dürfte der größte Vorteil einer Nutzung dieser Instrumente darin bestehen, pathologische Veränderungen der Zell- und Gewebestrukturen früher als bisher erkennen zu können und damit die Möglichkeiten einer erfolgreichen, an die individuellen Gegebenheiten der jeweiligen Betroffenen angepassten Therapie zu verbessern. Der aus der früheren Detektion des Tumors resultierende Zeitgewinn steigert nicht nur

[187] Sechopoulos et al. (2021): Artificial intelligence for breast cancer detection in mammography and digital breast tomosynthesis: State of the art. In: Seminars in Cancer Biology 72, 214-215 (DOI: 10.1016/j.semcancer.2020.06.002).
[188] McKinney, S. M. et al. (2020): International evaluation of an AI system for breast cancer screening. In: Nature 577, 89-94 (DOI: 10.1038/s41586-019-1799-6).

generell die Erfolgsaussichten der Behandlung, sie erweitert auch insofern die Therapieoptionen, als bestimmte Behandlungsmethoden (wie z. B. die operative Entfernung des Tumors) nur so lange möglich sind, wie eine Metastasierung noch nicht stattgefunden hat.

Ein weiterer Vorteil besteht darin, dass KI-gestützte Systeme die Auswertung digitaler Bilder beschleunigen und das ärztliche Personal von monotonen Routinearbeiten entlasten können, sodass diesem idealtypischerweise mehr Zeit für die Vermittlung der Befunde im persönlichen Gespräch zur Verfügung steht.[189]

Diesen Chancen stehen aber auch Risiken gegenüber. So können Ärztinnen und Ärzte infolge der fortschreitenden Delegation bestimmter Aufgaben an technische Systeme zum einen eigene Kompetenzen schleichend verlieren, da sie diese immer seltener selbst anwenden. Zum anderen könnten sie gerade aufgrund des verlorengegangenen eigenen Erfahrungswissens dazu neigen, ihre Sorgfaltspflichten im Umgang mit derartigen Instrumenten dadurch zu verletzen, dass sie deren Empfehlungen blind folgen (Automation Bias). Da auch KI-basierte Instrumente nicht fehlerfrei arbeiten, sondern mitunter nur andere Fehler als Menschen begehen und die Nichtentdeckung dieser Fehler fatale Konsequenzen haben kann, sollten KI-gestützte Entscheidungssysteme immer so gestaltet werden, dass die konkrete Art der Übermittlung ihrer Analyseergebnisse (etwa in Gestalt einer Wahrscheinlichkeitsangabe für das Vorliegen einer pathologischen Veränderung[190]) deutlich macht, dass hier noch eine ärztliche Plausibilitätsprüfung erforderlich ist. Da die Behandelnden moralisch und rechtlich dafür verantwortlich sind, den Betroffenen die aus ihrer Sicht beste Behandlung anzubieten, gehört auch die kritische Überprüfung der Ergebnisse der von ihnen eingesetzten technischen Instrumente zu den ärztlichen Sorgfaltspflichten, die als solche nicht delegierbar ist.

Ein weiteres Beispiel aus der Praxis im Krankenhaus soll die angesprochene Problematik illustrieren.[191] Ein KI-Gerät in der Radiologie unterstützt das Erkennen von Hirnblutungen bei Patientinnen und Patienten, indem es eine binäre Klassifikation anzeigt (ja, Blutung liegt vor bzw. nein, Blutung liegt nicht vor). Es stellt sich dabei die Frage, ob eine derart grobe Klassifikation

---

[189] Baltzer, P. A. T. (2021): Künstliche Intelligenz in der Mammadiagnostik – Anwendungsgebiete aus klinischer Perspektive. In: Der Radiologe 61, 192-198 (DOI: 10.1007/s00117-020-00802-2).

[190] Ein Bespiel für dieses Problem ist ein Algorithmus, der in der Notfallaufnahme zur Detektion einer Hirnblutung eingesetzt wird, und dessen Ausgabe lediglich ein „Ja“ oder „Nein“ umfasst. Hier liegt das Problem nicht in der Nutzung der KI als solcher, sondern in der undifferenzierten Art der Präsentation des Analyseergebnisses.

[191] Kiefer J.; May M. S. (2022): Diagnostic accuracy and analysis of an artificial intelligence algorithm for the detection of intracranial haemorrhage. In: RPS 717-3, ECR 2022 Book of Abstracts. Insights Imaging 13 (Suppl 4), 205 (DOI: 10.1186/s13244-022-01337-x).

ohne zusätzliche, differenziertere Angaben aus der Anwenderperspektive eine sinnvolle Entscheidungsunterstützung darstellt. Sowohl zur Einschätzung der Sicherheit des binären Klassifikationsergebnisses durch das ärztliche Personal als auch mit Blick auf die weitere Behandlungsplanung wären komplexere Aussagen wünschenswert, etwa zur Wahrscheinlichkeit der (verschiedenen) Diagnosen, der Lokalisation der Blutung im Gehirn oder zu möglichen Ursachen. Grundsätzlich erscheint ein Einsatz des Gerätes als maschineller „Zweitgutachter" aber durchaus sinnvoll.

Der Einsatz dieser KI-gestützten Klassifizierung stellt auch für erfahrene Fachkräfte trotz eines potenziellen Mehraufwandes eine Unterstützung dar, die zur besseren Urteilsbildung beiträgt. Praxisrelevant sind dabei technische Umsetzungsdetails wie beispielsweise eine direkte Integration im Scanner und eine schnelle Auswertung ohne gesonderte Einwilligung in den Datentransfer an anderer Stelle. Es sollte dabei sichergestellt werden, dass die KI-basierte Analyse das diagnostische Potenzial auf Grundlage der erhobenen Daten ausschöpft und den Behandelnden nachvollziehbar präsentiert und gegebenenfalls mit weiteren Daten (wie etwa typischen Vergleichsfällen und möglichen Konsequenzen der Läsion) in Verbindung bringt.

Für die *Patienten*-Perspektive ergeben sich noch weitere Aspekte. Denn obwohl die Chance einer verbesserten Früherkennung patientenseitig uneingeschränkt zu begrüßen ist, weil sie die Aussichten auf eine möglichst schonende Therapie und Heilung erhöht, bleibt die fortschreitende Technisierung einzelner Behandlungsschritte für die Arzt-Patienten-Beziehung keineswegs folgenlos. Personen, die im Rahmen ihrer medizinischen Behandlung mit neuen Diagnostik-Instrumenten konfrontiert werden, haben erfahrungsgemäß Fragen, etwa zur Sicherheit und Genauigkeit oder zum Datenschutz, und Befürchtungen, beispielsweise nicht mehr als Person wahrgenommen, sondern auf ihre Daten reduziert zu werden. Diese Fragen und Befürchtungen sind ernst zu nehmen und in einer umfassenden Aufklärung gezielt zu thematisieren, um den Patientinnen und Patienten auch die Möglichkeit zur Einholung einer Zweitmeinung zu eröffnen. Dieser Aufklärungsbedarf steigt zwangsläufig an, wenn nicht nur ein einzelnes Segment der ärztlichen Handlungssequenz technisch ersetzt wird, sondern weitere Ersetzungen möglich werden.

Da das Vertrauen in die Behandlung patientenseitig maßgeblich von der personalen Zuwendung der Behandelnden abhängt, dürfte nicht davon auszugehen sein, dass diese mit der fortschreitenden Implementierung von KI-Systemen in verschiedene Schritte des komplexen Behandlungsprozesses überflüssig werden. Vielmehr besteht teilweise die Hoffnung, die schon heute

bestehende Tendenz der Anonymisierung partiell zu korrigieren und den Ärztinnen und Ärzten wieder mehr Freiräume zur Erbringung ihrer eigentlichen Aufgaben zu schaffen.[192]

Ersetzungen mittleren Ausmaßes Da sich die digitale Transformation der Medizin dynamisch entwickelt/vollzieht, gibt es bereits heute eine Reihe von Phänomenen, bei denen die Grenze zwischen einer engen und einer mittleren Ersetzung einzelner Handlungsschritte durch KI-Komponenten zunehmend verschwimmt und die darauf hindeuten, dass es sich hierbei insgesamt um ein Kontinuum unterschiedlicher Nutzungsmöglichkeiten solcher Tools handelt, die zum gegenwärtigen Zeitpunkt erst teilweise realisiert sind.

Ein Beispiel hierfür ist ein Algorithmus zur Unterstützung des anästhesiologischen Fachpersonals zur Berechnung des Risikos, während einer Operation zu versterben. In einem ersten Schritt kann man die Nutzung solcher Instrumente als eine begrüßenswerte Erweiterung des ärztlichen Fähigkeitenspektrums deuten, da das bisherige operationsbegleitende Monitoring der Vitalparameter der operierten Person durch die Fachleute hier technisch so substituiert wird, dass eine frühzeitige Prognose kritischer Zustände möglich ist und eine entsprechende ärztliche Reaktion eingeleitet werden kann. Es ist aber in einem zweiten Schritt auch denkbar, dass der Algorithmus künftig nicht nur das Sterblichkeitsrisiko der behandelten Person in Echtzeit berechnet, sondern zusätzlich auch noch automatisiert die jeweils situativ angezeigten, bislang dem anästhesiologischen Personal vorbehaltenen Veränderungen der Operationsbedingungen (z. B. die Dosierung des Narkosemittels oder die zusätzliche Gabe von Katechulaminen) auslösen könnte.

In der Folge könnte sich das ärztliche Aufgabenspektrum künftig zunehmend auf die Wahrnehmung prä- und postoperativer Aufgaben (z. B. die Aufklärung der Patientinnen und Patienten) konzentrieren. Unabhängig davon, wie realistisch dieses Szenario ist, zeigen solche Beispiele, dass es eine Reihe von Misch- und Übergangsphänomenen gibt, die unter anderem auch daraus resultieren, dass die verschiedenen funktional differenzierbaren Segmente ärztlichen Handelns Teile eines einheitlichen Handlungszusammenhangs bilden.

Doch ist es nicht nur die innere Einheit von Diagnostik und Therapie, die solche Übergänge begünstigt. Auch externe Rahmenbedingungen und Effizienzüberlegungen zur Optimierung der Prozessqualität komplexer Behandlungsabläufe können Treiber einer fortscheitenden Ersetzungsdynamik sein, wie im Folgenden ein weiteres Beispiel aus der Onkologie verdeutlicht.

---

[192] Topos, E. (2019): The Topol Review. Preparing the healthcare workforce to deliver the digital future. Herausgegeben von Health Education England. https://topol.hee.nhs.uk/wp-content/uploads/HEE-Topol-Review-2019.pdf [22.02.2023].

Infokasten 3: Mittlere Ersetzung am Beispiel von Prostatakrebs

Prostatakrebs (Prostatakarzinom) ist mit rund 68 600 Neuerkrankungen jährlich die häufigste Krebserkrankung beim Mann. Die Ursache dieser Krebserkrankung noch weitgehend unbekannt, allerdings ist einer der wichtigsten Risikofaktoren das Alter. Aufgrund des demografischen Wandels hat die Anzahl an Prostatakrebserkrankungen deutlich zugenommen; gleichzeitig haben sich die Überlebensaussichten aber erheblich verbessert, da etwa zwei Drittel der Fälle bereits im Anfangsstadium diagnostiziert werden. In Screening-Verfahren zur Früherkennung wird das von den Karzinomzellen verstärkt gebildete prostataspezifische Antigen (PSA) gemessen.[193]

Erhebliche Fortschritte in der Prostatakrebsdiagnostik wurden zudem im Bereich der multiparametrischen Magnetresonanztomographie (mpMRT) in Kombination mit Ultraschall-gesteuerter Biopsie erzielt. Der KI-Einsatz unterstützt die Fusion der umfangreichen Datensätze und verbessert mittels Anwendung von Deep Learning Convolutional Neural Networks als Faltungsnetzwerk zur Bildsegmentierung, sogenannten U-Nets, die Erkennung und Lokalisation von suspekten Läsionen. Automatisierte Bildfusionsverfahren unterstützen die Gewebeentnahme, indem zuvor erstellte mpMRT-Aufnahmen in Echtzeit mit Ultraschallaufnahmen bei der Gewebeentnahme überlagert werden. Mit dieser Methode konnte durch eine präzise Biopsie die Entdeckung behandlungsbedürftiger Prostatakarzinome von ca. 50% auf 90% erhöht werden.[194] Für eine Therapieentscheidung ist letztlich jedoch eine pathologische Diagnose unerlässlich, die sich bisher weitgehend auf mikroskopische Untersuchungen stützt. Zunehmend werden aber, ähnlich der Auswertung von bildgebenden Verfahren wie Magnetresonanztomographie und Ultraschall, auch mikroskopische Bilder von Gewebeschnitten digitalisiert und automatisiert mit Unterstützung von KI-Algorithmen ausgewertet. Insbesondere bei der integrativen Auswertung aller Befunde kann KI darüber hinaus helfen, richtige Entscheidungen zu treffen, indem zunächst anhand eines „digitalen Zwillings" Therapieoptionen getestet werden, um möglichst erfolgversprechende Ansätze auszuwählen.

In der bildgeführten Therapie des Prostatakarzinoms kommen weitere (teil-)automatisierte Verfahren der Chirurgie und Strahlentherapie hinzu. Inzwischen werden in Deutschland an über 140 Kliniken Operationen mit dem Da-Vinci-Operationssystem durchgeführt.[195] Das roboterassistierte Operieren erweitert das Spektrum minimalinvasiver Operationen im Bauchraum. Die Geräte werden vom ärztlichen Personal gesteuert, das am Monitor die Gefäß- und Gewebestrukturen deutlich besser erkennen kann als mit bloßem Auge; die Schnitte werden von Roboterarmen ausgeführt. So wird die Präzision bei minimalinvasiven Operationen deutlich erhöht. Dies ermöglicht eine radikale und doch schonende und nervenerhaltende Prostataoperation (Prostatektomie).[196] Eine Therapiealternative bei Prostataoperationen, die sich noch in der Erprobungsphase befindet, ist die Cyberknife-Methode, eine

[193] Zentrum für Krebsregisterdaten (2019): Prostatakrebs (Prostatakarzinom). https://www.krebsdaten.de/Krebs/DE/Content/Krebsarten/Prostatakrebs/prostatakrebs_node.html [03.03.2023].
[194] Schelb, P. et al. (2019): Classification of Cancer at Prostate MRI: Deep Learning versus Clinical PI-RADS Assessment. In: Radiology 293, 607-617 (DOI: 10.1148/radiol.2019190938); Schlemmer, H.-P.; Hohenfellner, M. (2021): Chancen von KI in der Onkologie am Beispiel der individualisierten Diagnostik und Behandlung von Prostatakrebs. In: Zeitschrift für medizinische Ethik 67 (3), 309-326; Vgl. Gigerenzer G.; Mata J.; Frank R. (2009): Public knowledge of benefits of breast and prostate cancer screening in Europe. In: Journal of the National Cancer Institute 101 (17), 1216–1220 (DOI: 10.1093/jnci/djp237).
[195] Klinikradar (2023): Kliniken für Da-Vinci-OP-Roboter. Klinikliste 2023. https://klinikradar.de/da-vinci-op-roboter/kliniken/#klinikliste [11.01.2023]; Maurer, T.; Hoffmann, L. (2021): Prostatakrebs: Medizinroboter in der Therapie. In: Klinik Kompass. https://www.klinikkompass.com/der-medizinroboter-da-vinci-in-der-praxis/ [11.01.2023].
[196] https://www.leading-medicine-guide.de/behandlung/da-vinci-prostatektomie [11.01.2023].

vergleichsweise neue radiochirurgische Methode zur Bestrahlung von kleinen, gut lokalisierten Tumoren. Dabei werden Röntgenstrahlen gebündelt und höchst präzise auf den Tumor gelenkt, sodass das Strahlenbündel mit enormer Kraft die Funktion eines Operationsmessers übernimmt. Während die operierende Person die Strahlung reguliert, wird die Position in Echtzeit angepasst, was eine präzise Behandlung ermöglicht.[197]

Eine ethische Reflexion hat zunächst zu berücksichtigen, dass die angemessene Versorgung einer steigenden Anzahl von Prostatakrebspatienten vor einer doppelten strukturellen Herausforderung steht. Einerseits müssen die Arbeitsabläufe von Diagnostik und Therapie mit dem Ziel einer besseren Vernetzung aller beteiligten Akteure auf den unterschiedlichen Versorgungsebenen und eines beschleunigten Datenaustauschs weiter standardisiert werden. Andererseits soll die Behandlung für die einzelne Person durch eine bessere Stratifizierung relevanter Subgruppen personalisiert werden, um nutzlose oder sogar schädliche Therapieversuche zu vermeiden.

Nach Einschätzung vieler Sachverständiger [198] kann die verstärkte Nutzung von KI-Instrumenten die notwendigen Entscheidungen während des gesamten Behandlungsprozesses positiv unterstützen, da sich die einzelnen Segmente – von der klinischen Untersuchung und Labordiagnostik über Bildgebung und Biopsie bis hin zu Therapie und Nachsorge – in Gestalt von Entscheidungsbäumen strukturieren und damit prinzipiell einer Bearbeitung durch KI-Instrumente zugänglich machen lassen. Dabei sind es keineswegs nur reine Effizienzüberlegungen, die für den verstärkten Einsatz von KI-Systemen in diesem Bereich sprechen, sondern vor allem die dadurch ermöglichte Verbesserung der Behandlungsqualität, von der letztlich alle Beteiligten profitieren.

Mit den empirisch nachweisbaren Vorteilen – sowohl bei der Früherkennung von Prostatakrebs[199] als auch bei der Therapie durch neue KI-gesteuerte Operationsroboter, die mittlerweile eigenständig nicht nur die Ränder von Tumorgewebe präziser als ein Mensch identifizieren, sondern auch besonders gewebeschonend innerhalb des Organismus manövrieren können – gehen aber auch verschiedene Herausforderungen einher, die für das hier vorgestellte Beispiel ebenso gelten wie für viele ähnliche Anwendungsszenarien von KI im Gesundheitsbereich, bei

[197] https://radioonkologie.charite.de/leistungen/cyberknife/behandlungsspektrum/prostatakarzinom [11.01.2023].
[198] Zentrale Ethikkommission (2021): Entscheidungsunterstützung ärztlicher Tätigkeit durch Künstliche Intelligenz. In: Deutsches Ärzteblatt, 118 (33-34) (DOI: 10.3238/arztebl.zeko_sn_cdss_2021). Insbesondere auf den Seiten A2-A4 sind viele Studien verarbeitet.
[199] Schlemmer, H.-P.; Hohenfellner, M. (2021): Chancen von KI in der Onkologie am Beispiel der individualisierten Diagnostik und Behandlung von Prostatakrebs. In: Zeitschrift für medizinische Ethik 67 (3), 309-326, Anm. 5.

denen menschliche Tätigkeiten in engem bis mittlerem Ausmaß durch KI-gestützte Technik ersetzt werden.

Erstens ist für eine verlässliche Unterstützung der Diagnose und Therapie durch KI eine flächendeckende und einheitliche – oder zumindest qualitativ vergleichbare – technische Ausrüstung sowie entsprechend geschultes Personal und eine kontinuierliche Qualitätssicherung unabdingbar.

Zweitens ist aufgrund der immanenten Grenzen von KI-Systemen davon auszugehen, dass Ärztinnen und Ärzte bei der Diagnose und Behandlung von Prostatakarzinomen auch langfristig unverzichtbare Funktionen zu erfüllen haben, die nicht maschinell substituierbar sind. Beispielsweise müssen Befunde auf Plausibilität geprüft werden.

Drittens bedarf die konkrete Diagnose- und Therapieplanung einer Berücksichtigung der umfassenden Lebenssituation jedes einzelnen Betroffenen, seiner jeweiligen Vorerkrankungen, individuellen Interessen und Wertpräferenzen, zu der auch fortschrittliche KI-Systeme nicht in der Lage sind.

Viertens scheint es ärztlicherseits erforderlich, eine alle Versorgungsebenen und Behandlungsschritte einbeziehende Kommunikationsstrategie zu entwickeln, um der Gefahr zu wehren, dass die jeweiligen Behandelnden immer mehr hinter der Technik verschwinden und sich die von ihnen behandelten Personen mit ihren Fragen und Ängsten zunehmend allein gelassen fühlen. Da über den Einsatz solcher Systeme im Vorfeld aufgeklärt werden muss und auch während der Nutzung auftauchende Fragen zum Beispiel zur verantwortlichen Datennutzung, zum Datenschutz-oder zur Zuverlässigkeit der maschinell produzierten Empfehlungen beantwortet werden müssen, dürfte damit zu rechnen sein, dass die kommunikativen Anforderungen an eine vertrauensvolle Arzt-Patienten-Beziehung in der Zukunft sogar noch weiter zunehmen.

Fünftens ist damit zu rechnen, dass sich bestimmte Personen dem Einsatz dieser technischen Systeme selbst dann verweigern werden, wenn sie bisherigen Behandlungsformen aus medizinischer Sicht überlegen sind, sodass es im Rahmen der Interaktion zwischen ärztlichem Personal und den ihnen anvertrauten Personen klarer Regeln bedarf, die nicht nur die Patientenautonomie, sondern auch die ärztliche Therapiefreiheit schützen.

Sechstens ist zu prüfen, wie sich die immer stärkere Implementierung von KI-Systemen auf die Kostenentwicklung im Gesundheitswesen auswirkt.

Siebtens sind Herausforderungen zu beachten, die sich bei der für die (Weiter-)Entwicklung und Evaluierung solcher Systeme notwendigen Sammlung, Verarbeitung und Weitergabe von

gesundheitsbezogenen Daten ergeben. Im Bereich der KI-Nutzung ist das vom Ethikrat schon 2017 mit Blick auf die gesundheitsbezogene Datennutzung formulierte, übergreifende Ziel der Datensouveränität besonders schwierig zu erreichen.[200] Zum einen erfordern nahezu alle KI-Nutzungen im Gesundheitsbereich enorm große Big-Data-Datensätze. Die teils sehr restriktive individuelle Auslegungspraxis geltender Datenschutzbestimmungen und einzelne teils überholte Datenschutzregeln[201] können deren Gewinnung und damit teils auch sinnvollen Entwicklungen entgegenstehen. Zum anderen erlauben die Anwendungen oft noch unmittelbareren Zugriff auf die Privatsphäre von Betroffenen als andere Anwendungen der datenreichen Medizin. Darüber hinaus besteht das Problem der Verwendung solcher Daten für andere Zwecke. Entsprechend entbindet das Ziel, neue Erkenntnisse für bessere zukünftige Behandlungsmethoden im Bereich der öffentlichen Forschung zu generieren, keinesfalls davon, die Personen, die ihre Daten für die Forschung zur Verfügung stellen, möglichst umfassend (gegebenenfalls auch über den Grad der Fremdnützigkeit des jeweiligen Forschungsprojektes) aufzuklären. Erschwerend kommt hinzu, dass etwa die Sekundärnutzung von klinischen Daten für KI-getriebene klinische Forschung in einem komplexen Forschungsfeld erfolgt, in dem die Grenzen zwischen der im Gesundheitswesen verbreiteten, rein öffentlichen oder kooperativen öffentlich-privaten Forschung mit Gemeinwohlorientierung einerseits und der Forschung von Akteuren ganz außerhalb dieses insgesamt stark regulierten Systems andererseits zunehmend verschwimmen.

Weitreichende Ersetzung

Einer der wenigen medizinischen Handlungsbereiche, in denen KI-basierte Systeme zum Teil ärztliches bzw. anderes Gesundheitspersonal jedenfalls in bestimmten Kontexten und Gruppen bereits weitgehend oder sogar vollständig ersetzen, ist die Psychotherapie. In den letzten Jahren ist eine Fülle von Instrumenten zur (Teil-)Diagnose und Behandlung verschiedener psychischer Probleme entstanden, meist in Form von Bildschirm-basierten Apps, etwa Chatbots, mit denen auf algorithmischer Basis eine Art von Therapie – zumeist kognitive Verhaltenstherapie – mit den Betroffenen abläuft.[202]

---

[200] Datensouveränität verstanden „als eine den Chancen und Risiken von Big Data angemessene verantwortliche informationelle Freiheitsgestaltung“, Deutscher Ethikrat (2017): Big Data und Gesundheit – Datensouveränität als informationelle Freiheitsgestaltung. Berlin, 251-280.

[201] Ein Beispiel hierfür ist § 27 Abs. 4 des Bayerischen Krankenhausgesetzes. Dieser erlaubte es bis Mitte 2022 bayerischen Krankenhäusern nicht, dass Patientendaten das Krankenhaus verlassen. Erst Mitte Juni 2022 wurde der Paragraph geändert.

[202] Übersichten in Fiske, A. et al. (2019): Your Robot Therapist Will See You Now: Ethical Implications of Embodied Artificial Intelligence in Psychiatry, Psychology, and Psychotherapy. In: Journal of Medical Internet Research 21 (5), (DOI:10.2196/13216); Gratzer, D.; Goldbloom, D. (2020): Therapy and E-therapy – preparing

Vielfach werden solche Instrumente als Elemente einer ärztlich bzw. therapeutisch geführten Behandlung eingesetzt, etwa um durch ein kontinuierliches Monitoring Klarheit über die (auch kontextspezifische) Ausprägung, Variabilität und Häufigkeit von Symptomen zu gewinnen (sogenanntes *ecological momentary assessment* als Alternative zu retrospektiven Selbstberichten), Aspekte bestimmter Therapien zu verstärken (Biofeedback, Virtual Reality als imaginäre Konfrontationstherapie), Verhaltensweisen zu üben oder therapeutisch wirksame Aufgaben zu erledigen (beispielsweise Entspannungsübungen oder Selbstinstruktionen).[203] In dieser Form bleiben diese Anwendungen also unter einer gewissen professionellen Aufsicht und unterscheiden sich wenig von den bereits seit einigen Jahrzehnten bekannten therapeutischen Begleitinstrumenten, wie etwa papier- oder bildschirmbasierten psychotherapeutischen Aufgaben auf Fragebogen- oder Entscheidungsbaumbasis. Hier werden also allenfalls einzelne Elemente im Sinn einer engen Ersetzung auf algorithmischer Basis übernommen, gegebenenfalls mit etwas mehr Benutzungsfreundlichkeit und einer Menge (ihrerseits ethisch nicht immer unproblematischen) Möglichkeiten der Datensammlung[204], ohne dass die therapeutische Beziehung notwendigerweise grundlegend verändert würde.

Allerdings sind viele dieser Apps frei in Appstores erhältlich und werden daher auch außerhalb des medizinisch-therapeutischen Kontextes (laut Anbieterangaben) millionenfach eingesetzt. In dieser Form erfolgt – wenn Betroffene solche Apps nicht nur als einen niedrigschwelligen Einstieg in eine therapeutische Behandlung durch medizinisches Fachpersonal nutzen – tatsächlich eine Therapie, ganz ohne dass menschliches Personal hinzukäme.[205] Die Chatbots treten dabei in „Gesprächskontakt“ und „diagnostizieren“ den psychischen Zustand, in dem sich eine Person befindet. Sie bieten verschiedene therapeutische Wege sowie Übungen an und interagieren regelmäßig und aufsuchend mit den sie nutzenden Personen (es sei denn, diese haben die entsprechende Funktion ausgestellt).[206] Die meisten dieser Chatbots weisen darauf hin, dass etwa suizidale Gedanken unmittelbare, intensivere Behandlung erfordern, und „schlagen vor“,

---

future psychiatrists in the era of apps and chatbots. In: Academic Psychiatry, 44(2), 231-234 (DOI: 10.1007/s40596-019-01170-3).

[203] Van Daele, T. et al. (2021). Dropping the E: the potential for integrating e-mental health in psychotherapy. In: Current Opinion in Psychology, 41, 46-50. (DOI: 10.1016/j.copsyc.2021.02.007).

[204] Bei verschiedenen in App-Stores erhältlichen Apps ist unklar, ob und gegebenenfalls wie Nutzerdaten weiterverwendet werden. Da Anwendungen wie etwa der Chatbot Woebot kostenlos sind, ist anzunehmen, dass Daten weiterverkauft oder jedenfalls einer kommerziellen Nutzung zugeführt werden.

[205] Ebert, D. D. et al. (2018): Internet-and mobile-based psychological interventions: applications, efficacy, and potential for improving mental health. In: European Psychologist, 23, 167-187 (DOI: 10.1027/1016-9040/a000318).

[206] Mehta, A. et al. (2021): Acceptability and Effectiveness of Artificial Intelligence Therapy for Anxiety and Depression (Youper). In: Journal of Medical Internet Research, 23(6): e26771 (DOI: 10.2196/26771).

Krisentelefon-Services in Anspruch zu nehmen. Sie sind jedoch nicht mit Notdiensten o. Ä. verbunden und fungieren daher nicht als eigenständiges Warnsystem.

Hier ergibt sich also tatsächlich eine weite Ersetzung insofern, als die Chatbots und diejenigen, die mit ihnen interagieren, in einer Art direkten, therapeutischen Beziehung stehen. Bisher gibt es wenige empirische Studien oder theoretische Abhandlungen, die dieses Phänomen untersuchen.[207] Dennoch liegen ethische Vorteile und Bedenken auf der Hand. Positiv diskutiert wird insbesondere, dass solche Apps angesichts ihrer Niedrigschwelligkeit und ständigen Verfügbarkeit Menschen in Erstkontakt mit therapeutischen Angeboten bringen können, die sonst oft zu spät oder gar keine Therapie erhalten. Dies gilt insbesondere für Gruppen, die oft schwer (z. B. wohnortbedingt) mit anderen Angeboten erreicht werden (Gateway-Phänomen).[208]

Hinzu kommt ein weithin für die Nutzung etwa von Aufklärungsalgorithmen in Krankenhäusern beschriebenes Phänomen[209], dass Menschen maschinellen „Therapeuten" mehr von sich berichten–und weniger Scham verspüren als im Gespräch von Mensch zu Mensch. Gerade Scham oder die Sorge vor Stigmatisierung hält bekanntermaßen viele Menschen davon ab, frühzeitig Therapieangebote aufzusuchen. Diesbezüglich könnte also gerade die Ersetzung menschlichen Fachpersonals dazu führen, dass überhaupt ein therapieähnliches Angebot in Anspruch genommen wird. Hier ist allerdings zu berücksichtigen, dass die Frage, ob und wieweit sich Menschen öffnen – und welche Effekte dies hat –, nicht nur von der Anonymität der Interaktion, sondern auch vom Vertrauen in die Kontaktperson abhängt. Der anonyme Chatbot ersetzt nicht ohne Weiteres den als vertrauenswürdig wahrgenommenen Therapeuten oder die Therapeutin, denn die Wirkung von Psychotherapie entfaltet sich in vielen Fällen erst auf der Grundlage der Therapeut-Klienten-Beziehung.[210] Und schließlich wird insbesondere mit Blick auf Regionen, in denen nicht ausreichend Therapieplätze zur Verfügung stehen, argumentiert, dass angesichts

---

[207] Holohan, M.; Fiske, A. (2021): "Like I'm Talking to a Real Person": Exploring the Meaning of Transference for the Use and Design of AI-Based Applications in Psychotherapy. In: Frontiers in Psychology, 12:720476 (DOI: 10.3389/fpsyg.2021.720476).

[208] Parviainen, J.; Rantala, J. (2021): Chatbot breakthrough in the 2020s? An ethical reflection on the trend of automated consultations in health care. In: Medicine, Health Care and Philosophy, 25, 61-71 (DOI: 10.1007/s11019-021-10049-w).

[209] Pugh, A. (2018): Automated health care offers freedom from shame, but is it what patients need? In: The New Yorker. https://www.newyorker.com/tech/annals-of-technology/automated-health-care-offers-freedom-from-shame-but-is-it-what-patients-need [28.02.2023]; Dennis, A. et al. (2020): User reactions to COVID-19 screening chatbots from reputable providers. In: Journal of the American Medical Informatics Association, 27 (11), 1727–1731 (DOI: 10.1093/jamia/ocaa167).

[210] Croes, E.; Antheunis, M. (2020): 36 Questions to Loving a Chatbot: Are People Willing to Self-disclose to a Chatbot? In: Følstad, A. et al. (Hg.): Chatbot Resaerch and Design. 4th International Workshop, 81-95; Gerhardinger, S. (2020). Entwicklung der Therapeutenpersönlichkeit. Berlin, Heidelberg. Vgl. hierzu das Kapitel ‚E-Mental Health: Psychotherapie ohne Psychotherapeuten', 225-239.

des immer weiter steigenden Versorgungsbedarfs[211] gerade milderer psychischer Probleme Chatbot-Apps jedenfalls „besser als nichts“ sind.[212]

Allerdings ergeben sich insbesondere im Falle einer Ersetzung therapeutischer Fachkräfte durch Maschinen aus ethischer Sicht wichtige Bedenken und Probleme. Offenkundig sind Probleme wie die mangelnde Qualitätskontrolle der Bots (v. a. wenn sie nicht als digitale Gesundheitsanwendungen[213] zugelassen sind), Fragen zur Datensammlung und -weiterverwendung, zum Schutz der Privatsphäre sowie die erwähnte fehlende Warnfunktion, etwa bei eindeutiger Suizidalität.

Hinsichtlich der Betroffenenperspektive kann zudem angenommen werden, dass gerade vulnerable Personen, die ohnehin schlecht versorgt sind, den Eindruck gewinnen, mit vermeintlich zweitklassigen therapeutischen Surrogaten abgespeist zu werden. Weiterhin kann nicht unterstellt werden, dass alle, die solche Apps verwenden, umfassend über deren Charakter informiert sind. Es ist möglich, dass manche die App mit einer Art telemedizinischem Angebot verwechseln und davon ausgehen, dass auf der anderen Seite doch ein Mensch therapeutisch agiert. Hier würde also die Ersetzung als solche verkannt – mit potenziell problematischen Folgen.

Schließlich ist noch so gut wie gar nicht untersucht, ob innerhalb der therapeutischen Beziehung zwischen einer App und denjenigen, die sie nutzen, Phänomena von Übertragung – ein wesentliches Element bei vielen therapeutischen, insbesondere psychoanalytischen Ansätzen – erfolgen. Es ist zumindest anzunehmen, dass Menschen eine Art emotionale Beziehung zur therapeutischen App aufbauen könnten, die von Zuneigung, Sich-Verlassen bis hin zu Abhängigkeitsaspekten reichen könnte. Daraus ergibt sich nicht nur die Frage, ob und inwieweit hier Empathie und Verstehen technisch substituiert werden können oder sollen, sondern auch, ob dies auf Dauer Effekte auf die Beziehungsfähigkeiten der Betroffenen haben könnte. Therapiebegünstigende (weil reflektierte) Gegenübertragungen seitens der therapeutischen Fachkraftkönnen von KI-Anwendungen hingegen nicht erwartet werden. Dies verdeutlicht, dass digitale Anwendungen auf dem Gebiet der Psychotherapie nicht unabhängig von der je-

[211] Vgl. hierzu auch Deutscher Ethikrat (2022): Pandemie und psychische Gesundheit. Berlin.
[212] Fiske, A. et al. (2019): Your Robot Therapist Will See You Now: Ethical Implications of Embodied Artificial Intelligence in Psychiatry, Psychology, and Psychotherapy. In: Journal of Medical Internet Research 21 (5): e13216. (DOI:10.2196/13216), 5.
[213] Siehe die Digitale-Gesundheitsanwendungen-Verordnung (DiGAV). https://www.bundesgesundheitsministerium.de/service/gesetze-und-verordnungen/guv-19-lp/digav.html [22.02.2023].

weiligen Therapieform betrachtet werden können – ebenso wenig, wie die Therapieformen unabhängig von der jeweiligen Nutzerpersönlichkeit und der jeweils infrage stehenden Problematik sowie ihrem unmittelbaren Bedrohungspotenzial betrachtet werden können.

Auch aus gesellschaftlicher Perspektive ergeben sich Fragen. Kontrovers diskutiert wird etwa, ob die zunehmende Nutzung solcher Apps einem weiteren Abbau von therapeutischem Fachpersonal Vorschub leistet und damit die Reduktion von Versorgungsbereichen beschleunigt. Ebenso wird eine Verstärkung gesundheitlicher Ungleichheiten und von digital divides angemahnt sowie auch ein möglicher weiterer Prestigeverlust der *sprechenden Medizin*, die in ungerechtfertigter Weise als durch Algorithmen einfach ersetzbar erscheinen könnte.

5.3 Fazit und Empfehlungen

Die Erfahrung, dass sich die digitale Transformation unseres Gesundheitssystems auf verschiedene Bereiche des medizinischen Handelns bislang sehr unterschiedlich auswirkt, bestätigt sich auch im Blick auf den speziellen Bereich von KI-Anwendungen. Dies gilt nicht allein für den Grad der Durchdringung eines Handlungsfeldes mit den neuen technischen Möglichkeiten und die Dynamik der daraus resultierenden Veränderungen für die Arzt-Patienten-Beziehung, sondern auch für die Gründe, die diese Entwicklung jeweils vorantreiben.

Neben naheliegenden immanenten Faktoren wie der Existenz immer größerer Datenmengen, die einer maschinellen Bearbeitung zugänglich sind, können auch kontingente Faktoren die Nutzung von KI-Anwendungen begünstigen. Dazu gehören beispielsweise Versorgungsengpässe aufgrund von Personalmangel, aber auch die Möglichkeit bzw. Erwartung präziserer Diagnostik oder neuerer, gegebenenfalls niedrigschwelligerer therapeutischer Anwendungen in einzelnen Bereichen.

Die in den vorigen Abschnitten vorgestellten Beispiele machen deutlich, dass es dabei sehr auf die jeweils konkreten Bedingungen des Einsatzes von KI-basierten Tools ankommt, ob diese beispielsweise innerhalb oder außerhalb einer etablierten Arzt-Patienten-Beziehungen zum Einsatz kommen. Bei der Zertifizierung zukünftiger Anwendungen sind solche kontextspezifischen Fragen zu berücksichtigen.

Ungeachtet der daraus resultierenden Notwendigkeit einer Binnendifferenzierung des weiten Handlungsfeldes der Medizin, legen sich einige übergreifenden Empfehlungen für den Gesundheitssektor nahe.

Empfehlungen

- *Empfehlung Medizin 1:* Bei der Entwicklung, Erprobung und Zertifizierung medizinischer KI-Produkte bedarf es einer engen Zusammenarbeit mit den relevanten Zulassungsbehörden sowie insbesondere mit den jeweils zuständigen medizinischen Fachgesellschaften, um Schwachstellen der Produkte frühzeitig zu entdecken und hohe Qualitätsstandards zu etablieren.
- *Empfehlung Medizin 2:* Bei der Auswahl der Trainings-, Validierungs- und Testdatensätze sollte über bestehende Rechtsvorgaben hinaus mit einem entsprechenden Monitoring sowie präzise und zugleich sinnvoll umsetzbaren Dokumentationspflichten sichergestellt werden, dass die für die betreffenden Patientengruppen relevanten Faktoren (wie z. B. Alter, Geschlecht, ethnische Einflussfaktoren, Vorerkrankungen und Komorbiditäten) hinreichend berücksichtigt werden.
- *Empfehlung Medizin 3:* Bei der Gestaltung des Designs von KI-Produkten zur Entscheidungsunterstützung ist sicherzustellen, dass die Ergebnisdarstellung in einer Form geschieht, die Gefahren etwa von Automatismen *(*Automation Bias) transparent macht, ihnen entgegenwirkt und die die Notwendigkeit einer reflexiven Plausibilitätsprüfung der jeweils vom KI-System vorgeschlagenen Handlungsweise unterstreicht.
- *Empfehlung Medizin 4:* Bei der Sammlung, Verarbeitung und Weitergabe von gesundheitsbezogenen Daten sind generell strenge Anforderungen und hohe Standards in Bezug auf Aufklärung, Datenschutz und Schutz der Privatheit zu beachten. In diesem Zusammenhang verweist der Deutsche Ethikrat auf seine 2017 in Kontext von Big Data und Gesundheit formulierten Empfehlungen, die sich am Konzept der Datensouveränität orientieren, das für den Bereich von KI-Anwendungen im Gesundheitsbereich gleichermaßen Gültigkeit entfaltet.[214]
- *Empfehlung Medizin 5:* Bei durch empirische Studien sorgfältig belegter Überlegenheit von KI-Anwendungen gegenüber herkömmlichen Behandlungsmethoden ist sicherzustellen, dass diese allen einschlägigen Patientengruppen zur Verfügung stehen.

[214] Datensouveränität verstanden „als eine den Chancen und Risiken von Big Data angemessene verantwortliche informationelle Freiheitsgestaltung“, Deutscher Ethikrat (2017): Big Data und Gesundheit – Datensouveränität als informationelle Freiheitsgestaltung. Berlin, 251-280.

- *Empfehlung Medizin 6:*Für erwiesen überlegene KI-Anwendungen sollte eine rasche Integration in die klinische Ausbildung des ärztlichen Fachpersonals erfolgen, um eine breitere Nutzung vorzubereiten und verantwortlich so gestalten zu können, dass möglichst alle Patientinnen und Patienten davon profitieren und bestehende Zugangsbarrieren zu den neuen Behandlungsformen abgebaut werden. Dazu ist die Entwicklung einschlägiger Curricula/Module in Aus-, Fort- und Weiterbildung notwendig. Auch die anderen Gesundheitsberufe sollten entsprechende Elemente in die Ausbildung aufnehmen, um die Anwendungskompetenz bei KI-Anwendungen im Gesundheitsbereich zu stärken.
- *Empfehlung Medizin 7:*Bei routinemäßiger Anwendung von KI-Komponenten sollte nicht nur gewährleistet werden, dass bei denjenigen, die sie klinisch nutzen, eine hohe methodische Expertise zur Einordnung der Ergebnisse vorhanden ist, sondern auch strenge Sorgfaltspflichten bei der Datenerhebung und -weitergabe sowie bei der Plausibilitätsprüfung der maschinell gegebenen Handlungsempfehlungen eingehalten werden. Besondere Aufmerksamkeit erfordert die Gefahr eines Verlustes von theoretischem wie haptisch-praktischem Erfahrungswissen und entsprechenden Fähigkeiten (*deskilling*); dieser Gefahr sollte mit geeigneten, spezifischen Fortbildungsmaßnahmen entgegengewirkt werden.
- *Empfehlung Medizin 8:* Bei fortschreitender Ersetzung ärztlicher, therapeutischer und pflegerischer Handlungssegmente durch KI-Komponenten ist nicht nur sicherzustellen, dass Patientinnen und Patienten über alle entscheidungsrelevanten Umstände ihrer Behandlung vorab informiert werden. Darüber hinaus sollten auch gezielte kommunikative Maßnahmen ergriffen werden, um dem drohenden Gefühl einer zunehmenden Verobjektivierung aktiv entgegenzuwirken und das Vertrauensverhältnis zwischen den beteiligten Personen zu schützen. Je höher der Grad der technischen Substitution menschlicher Handlungen durch KI-Komponenten ist, desto stärker wächst der Aufklärungs- und Begleitungsbedarf der Patientinnen und Patienten. Die verstärkte Nutzung von KI-Komponenten in der Versorgung darf nicht zu einer weiteren Abwertung der sprechenden Medizin oder einem Abbau von Personal führen.
- *Empfehlung Medizin 9:* Eine vollständige Ersetzung der ärztlichen Fachkraft durch ein KI-System gefährdet das Patientenwohl und ist auch nicht dadurch zu rechtfertigen, dass schon heute in bestimmten Versorgungsbereichen ein akuter Personalmangel besteht. Gerade in komplexen Behandlungssituationen bedarf es eines personalen Gegenübers, das durch tech-

nische Komponenten zwar immer stärker unterstützt werden kann, dadurch selbst als Verantwortungsträger für die Planung, Durchführung und Überwachung des Behandlungsprozesses aber nicht überflüssig wird.

+++ Vorabfassung! Wird durch die lektorierte Fassung ersetzt. +++

6 Bildung

6.1 Einleitung

Auch im Bildungsbereich kommen zunehmend digitale Technologien und algorithmische Systeme zum Einsatz.[215] Dies kann sowohl zur Standardisierung von Lernprozessen führen, als auch im Falle von stark datenbasierten KI-Systemen – ähnlich wie in der Medizin – mehr Personalisierung ermöglichen, d.h. einen stärkeren Zuschnitt auf individuelle Bedingungen und Neigungen der Lernenden. Personalisiertes Lernen und Lehren wird dabei unter anderem dadurch möglich, dass durch umfangreiche maschinelle Datenerfassung und -auswertung ein Vielfaches an Informationen zum Lerngeschehen im Vergleich zu früheren Möglichkeiten zur Verfügung steht und von Lehrkräften, Lernenden und weiteren Beteiligten zur Personalisierung von Lehr- und Lernprozessen verwendet werden kann.[216] Wie auch im Anwendungsfeld Medizin reichen die Einsatzmöglichkeiten von sehr eng umrissenen punktuellen Angeboten für Lernende und Lehrkräfte, bis hin zu Szenarien, in denen digitale Werkzeuge umfassend eingesetzt werden und beispielsweise datenbasierte KI-gestützte Lehr-Lernsysteme zeitweise oder gänzlich eine Lehrkraft ersetzen können.[217] Die damit verbundenen Veränderungspotenziale, Chancen und Risiken werden hier für die Schule als exemplarischem Lernort und Ort der Persönlichkeitsbildung vorgestellt.[218]

Jenseits der auch anderweitig thematisierten technischen Herausforderungen rund um die Nutzung digitaler und algorithmischer Systeme in der Schule[219] gilt das Interesse in dieser Stellungnahme vor allem den anthropologischen und ethischen Fragen. Im Mittelpunkt steht hierbei die Auseinandersetzung damit, wie die maschinelle Ersetzung bestimmter menschlicher Handlungssegmente Lern- und Lehrprozesse verändert, Handlungsmöglichkeiten aller Akteure er-

---

[215] OECD Digital Education Outlook (2021): Pushing the Frontiers with Artificial Intelligence, Blockchain and Robots. In: OECD Publishing. Paris (DOI: 10.1787/589b283f-en).
[216] Dies ist natürlich nur unter der Voraussetzung gegeben, dass sowohl mobile Endgeräte als auch kostenloses WLAN für die Schülerinnen und Schüler von staatlicher Seite zur Verfügung gestellt werden. Bildungsungleichheit bei den Startchancen darf dabei nicht weiter vertieft, sondern könnte durch das spezifische Eingehen auf die Heterogenität der Schülerschaft möglicherweise grundlegend angegangen werden.
[217] Majumdar, M.; Zahorsky, I. (2020): KI als helfende Lehrkraft während des Lockdowns. In: eGovernment. https://www.egovernment-computing.de/ki-als-helfende-lehrkraft-waehrend-des-lockdowns-a-926866/ [11.01.2023].
[218] Somit sind andere Orte der Bildung und des Lernens wie Hochschulen, Kindertagesstätten oder andere Formen der Aus- und Weiterbildung nicht explizit berücksichtigt.
[219] Vgl. zum Überblick OECD Digital Education Outlook (2021): Pushing the Frontiers with Artificial Intelligence, Blockchain and Robots. In: OECD Publishing. Paris (DOI: 10.1787/589b283f-en).

weitert oder auch vermindert. Zudem ist die Frage, wie dies das Selbstverständnis und Miteinander der verschiedenen Beteiligten am Lernort Schule und dem Zuhause der Schulpflichtigen beeinflusst und gegebenenfalls darüber hinaus gesellschaftliche Auswirkungen entfaltet.

Wie in Kapitel 3 ausgeführt, orientiert sich das hier zugrunde gelegte Verständnis von Bildung an der Fähigkeit des Menschen zu freiem und vernünftigem Handeln, das nicht auf behavioristisches oder funktionalistisches Modelle zu reduzieren ist. Die vielfältigen Möglichkeiten des Einsatzes digitaler und algorithmischer Angebote in der Bildung sind auf diese grundlegende Ausrichtung von Bildung zu beziehen. Wenn daher von Chancen und Risiken datenbasierter KI-Lehr- und Lernsystemen die Rede ist, sind diese nicht einfach an technologischen Optimierungsvorstellungen zu orientieren. Alle technologischen Möglichkeiten der Gestaltung der Bildungsprozesse sind vielmehr daraufhin zu überprüfen, ob sie dem eingangs entwickelten Verständnis des Menschen als einer zur Selbstbestimmung und Verantwortung fähigen Person entsprechen oder ob sie diesem Verständnis entgegenstehen. Die Notwendigkeit einer solch grundlegende Betrachtung wird auch in den aktuellen Debatten rund um ChatGPT sichtbar.[220] Unmittelbar nachdem dieser KI-basierte und hoch performante Textgenerator frei zugänglich gemacht wurde, wurden disruptive Entwicklungen nicht zuletzt für den Bildungsbereich diskutiert. Ging es vordergründig zunächst um die Frage, wie Prüfungen fair gestaltet werden können, wenn man nicht weiß, ob Schülerinnen und Schüler ihre Hausarbeiten selbst geschrieben oder von ChatGPT haben produzieren lassen, so erfordert ChatGPT darüber hinauseine erneute Auseinandersetzung damit, was Bildung ist und sein soll. Es geht um eine Vergewisserung und Neubestimmung dessen, was Ziel und Wert von Bildung ist, was relevantes Wissen ist und welche Fertigkeiten und Fähigkeiten Lernende weiterhin benötigen und welche vielleicht an Relevanz verloren haben.

6.2 Zum Bildungsbegriff

Oberstes Ziel von Bildung ist es, den Menschen zu einer mündigen und freien, und das heißt einer zur Verantwortung fähigen Person heranzubilden. Bildungsbegriffe und -konzepte sind zwar immer auch vom jeweiligen Kulturkreis abhängig, insofern die Sicht auf den Menschen, seine Sozialität und damit die Verhältnisbestimmung von Individuum und Gesellschaft auch das Verständnis von Bildung prägt. Die folgenden Überlegungen gehen davon aus, dass Bil-

[220] Da die inhaltliche Befassung mit der Stellungnahme zum Zeitpunkt der Veröffentlichung von ChatGPT bereits abgeschlossen war, wird hier nicht detaillierter auf ChatGPT und die damit verbundenen Folgen für schulische Bildung eingegangen.

dung sich an einem Verständnis des Humanum orientiert, das auf eine umfassende Persönlichkeitsbildung zielt, die auf Urteilsfähigkeit und verantwortliche Teilhabe an der Gesellschaft angelegt ist.

Solche Bildung erfordert den Erwerb von Orientierungswissen als Bedingung reflexiver Urteilskraft und Entscheidungsstärke. Orientierungswissen entsteht einerseits aus der kritischen Rezeption empirischer Inhalte – von Alltagserfahrungen bis zu Ergebnissen wissenschaftlicher Forschung – und andererseits aus der Wahrnehmung normativer Inhalte – von sozialen Üblichkeiten bis zur reflektierten moralischen Urteilsbildung. Aus beiden Quellen bildet sich die Fähigkeit zur Übernahme von Verantwortung des Menschen gegenüber seiner sozialen Umgebung und sich selbst.

So umfasst Bildung neben der Vermittlung von Informationen ebenso Kontextwissen, das technische Kompetenzen vor dem Hintergrund verschiedener Erfahrungen zu beurteilen und einzusetzen vermag. Zu diesen Erfahrungen gehören politische, kulturelle oder existenzielle Erfahrungen, die in einem kulturellen Lernen intergenerationell weitergegeben und angeeignet werden können. Wenn über die Kriterien der Bildung angesichts voranschreitender Digitalisierung auch im schulischen Unterricht nachgedacht wird, ist das aus empirischen wie normativen Aspekten zusammengesetzte Orientierungswissens als Grundlage kulturellen Lernens zu berücksichtigen.[221] Es ist als Kriterium von Bildung auch für die Chancen und Risiken digitaler Bildung relevant.

Die Digitalisierung ist nicht Selbstzweck und die mit der digitalen Bildung oft einhergehende Vorstellung der Optimierung von Lernprozessen (oder des Lehr-Lern-Geschehens) ist kritisch auf das grundlegende Bildungsziel urteilsfähiger Selbstbestimmung des Menschen und seiner Handlungsverantwortung zu überprüfen.

Bildung, die auch im Rahmen von digitaler Bildung als Bildung zum Humanum zu begreifen ist, geht einher mit Identitätsstiftung. Aus den Überlegungen der vorangegangenen Kapitel ergibt sich, dass in diesem Feld der KI mögliche Lehrroboter eine Bildung, die sich als Kompetenz zur Selbstbestimmung und Urteilsfähigkeit auszeichnet, nicht ersetzen können. Sie können dem Kompetenzerwerb sogar abträglich sein, denn zum hier vertretenen Bildungsverständnis gehört es, Ideen zu entwickeln, Geschichten zu entwerfen, Kunstwerke zu entdecken, aber

[221] Nicht weiter eingegangen wird auf interkulturelles Lernen.

auch Kritikfähigkeiten auszubilden, um Ideologien zu durchschauen. Das abwägende und kompetente Urteil der mündigen Person bildet hier die Voraussetzung für die Beurteilung der digitalen Prozesse.

Zu beachten ist daher ferner: Bildung erschöpft sich nicht in einer kognitiven Vermittlung von Information, sondern sie schließt affektive Dimensionen und soziale Kontakte ein. Schulisches Lernen ist daher nicht auf kognitiv-technische Operationen zu reduzieren, sondern beinhaltet auch emotionale und motivationale Aspekte. Für die Lehrkräfte ist es von Bedeutung, sich klarzumachen, dass das Lehr- und Lerngeschehen als dynamische Interaktion mit anderen Personen zu begreifen ist, als Interaktion mit Lehrenden wie auch mit anderen Lernenden. Lehrkräfte können im besten Fall als Identifikationsfiguren ein motivierendes und identitätsstiftendes Potenzial entfalten, das den Wissensdurst in Schülerinnen und Schülern wecken und Begeisterungsfähigkeit entfachen kann. Der Informationsgewinn durch Vermittlung von Sachwissen wird damit reflexiv auf die eigene Lebenswirklichkeit bezogen. Solches reflexive Lernen ist entscheidend für die Vorgänge des Verstehens und Aneignens.

Digitale Bildung ist daraufhin zu überprüfen, ob sie solch reflexives Lernen in der Interaktion zwischen Lehrkräften und Lernenden fördert oder behindert. Denn erst in diesem reflexiven Aneignen von Wissen werden die Potenziale jener Bildung freigesetzt, die den Menschen als freie und selbstverantwortliche Person ansprechbar werden lassen. Diese Kompetenzen reflexiven Verstehens und Aneignens sind daher Voraussetzung für einen urteilsfähigen Umgang mit den Möglichkeiten digitalen Lernens.

Bildung umfasst damit mehr als Informationszuwachs und logisches Schließen. Dazu zählt die immer mitlaufende Bezogenheit auf soziale und kommunikative Verständigung. Werden Teile der Interaktion zwischen Lernenden und Lehrenden oder auch der Lernenden untereinander durch digitale Angebote ersetzt oder vermittelt, stellt sich die Frage, was dies für individuelle und soziale Lernprozesse bedeutet. Einer Automatisierung von Lernprozessen, die den reflexiven Bezug der Aneignung von Wissen und des Verständnisses erschwert oder gar unterbindet, wäre daher eine Absage zu erteilen. Damit würde das Lernen nicht mehr im vorgestellten umfassenden Sinne der Bildung der Person erfolgen. Ob Bildung sich ohne ein nicht zu vernachlässigendes personales Gegenüber vollziehen kann, ist fraglich. Denn für den Aneignungs- und Verstehensprozess ist der kommunikative Aspekt des Lernens, der ein personales Gegenüber voraussetzt, unverzichtbar.

Der Eingang der Digitalisierung in den Bereich der Bildung muss hinsichtlich der Veränderungsprozesse, den er für die Bildung als solche einleitet, daher genau reflektiert werden. Das

gilt sowohl für Prozesse aufseiten der Lehrenden wie aufseiten der Lernenden. Dafür ist es zum Beispiel nötig, dass Lehrkräfte auch die Funktionsweise datenbasierter KI-gestützter Lehr-Lern-Software hinsichtlich ihrer didaktischen Leistungsfähigkeit einordnen können, um ihren Einsatz konstruktiv vornehmen, aber auch kritisch begleiten zu können. Es geht mit Blick auf die Zukunft darum, den Einsatz der neuen Techniken an den Entwicklungspotenzialen der Lernenden zu orientieren. Das Ziel muss sein, die Bildung zur Verantwortungsfähigkeit nicht zu vermindern, sondern zu erweitern. Aufgrund der Dynamik und Komplexität der Entwicklung solcher digitalen Lerntechniken ist es daher erforderlich, möglichst bald entsprechend der Zertifizierung klassischer Lehr- und Lernmaterialien auch deren digitale Pendants zu zertifizieren.[222]

6.3 Einsatzmöglichkeiten datenbasierter, KI-gestützter Lehr-Lern-Systeme

Grundsätzlich eröffnet die Entwicklung zunehmend leistungsfähiger und kostengünstiger Geräte, die mobil verwendet, vernetzt und mit potenziell vielfältigen Sensoren ausgestattet oder verknüpft werden können, eine Fülle an (neuen) Einsatzmöglichkeiten im Lernraum Schule. Sie decken ein Spektrum ab, das von der Erfassung relevanter Daten und dem Erkennen von Mustern über deren Analyse zur Diagnostik von zum Beispiel Wissensständen und Lernfortschritten bis hin zu didaktischen Interventionen reicht.[223] Ausgangspunkt ist immer die Sammlung und Auswertung vieler Daten der Lernenden und mitunter auch der Lehrenden[224]. Qualität und Umfang der verfügbaren Daten haben einen zentralen Einfluss auf die Leistungsfähigkeit von KI-Systemen.[225] Für Lehr-Lern-Systeme bedeutet das, dass die Genauigkeit von Prognosen über den Lernverlauf mit der Menge der verfügbaren Daten steigt.[226] Jede intensive – und mitunter auch invasive – Datenerhebung für den Einsatz von KI-gestützter Lehr-Lern-Software sollte je nach Zweck und Ziel des Einsatzes beurteilt und auf ihr sachlich angemessenes und ethisch vertretbares Maß überprüft werden. Hier stellen sich Fragen nach dem sinnvollen Grad

---

[222] Vgl. die österreichischen Bemühungen: Bundesministerium Digitalisierung und Wirtschaftsstandort (2021): Strategie der Bundesregierung für Künstliche Intelligenz Artificial Intelligence Mission Austria 2030 (AIM AT 2030). Wien, 30.
[223] Molenaar, I. (2021): Personalisation of Learning. Towards hybrid human-AI learning technologies. In: OECD Digital Education Outlook 2021: Pushing the Frontiers with Artificial Intelligence, Blockchain and Robots. Paris, 57-77 (DOI: 10.1787/589b283f-en), 62.
[224] Auch für Lehrkräfte stehen angesichts vieler erhobener Datenmengen vielseitigere Evaluationsmöglichkeiten zur Verfügung.
[225] Deutsches Forschungszentrum für Künstliche Intelligenz (2017): Künstliche Intelligenz. Wirtschaftliche Bedeutung, gesellschaftliche Herausforderungen, menschliche Verantwortung. Berlin, Kaiserslautern, 66, 187.
[226] Martini, M. et al. (2020): Automatisch erlaubt? Fünf Anwendungsfälle algorithmischer Systeme auf dem juristischen Prüfstand. Herausgegeben von der Bertelsmann Stiftung. Gütersloh (DOI 10.11586/2019067) , 23.

und Ausmaß der Datenerhebung sowie deren wünschenswerten Verwendungsweisen. Dabei geht es beispielsweise darum, wie Datenerfassung die Lernenden in ihrem individuellen Lernprozess bestmöglich unterstützt, ohne dass diese Daten zur Überwachung oder Stigmatisierung von einzelnen Lernenden missbraucht werden können. Hier ist ein ethischer Umgang mit Daten auch im schulischen Bereich gefordert, der größtmögliche Datensouveränität auch für Schülerinnen und Schüler ermöglicht.[227]

Auf Grundlage der erhobenen Daten können individualisierte Rückmeldungen über Lern- und Lehrprozesse erfolgen sowie entsprechende Reaktionen oder Empfehlungen des Softwaresystems, die in der Folge die Grundlage für didaktische wie auch pädagogische Interventionen und auch Innovation bilden können. Dabei kommen zunehmend Elemente maschinellen Lernens zum Einsatz, bei denen die fortlaufende Analyse der Eingabedaten dafür genutzt wird, die diagnostischen und gegebenenfalls prädiktiven Ausgaben des Softwaresystems wie auch seine Empfehlungen dynamisch anzupassen. So können die Systeme durch Auswertung von beispielsweise Lerngeschwindigkeit, typischen Fehlern, Stärken und Schwächen das Lernprofil der Lernenden erkennen und die Lerninhalte entsprechend anpassen.[228] Die subjektiven Eindrücke der Lehrkräfte zu Lernfortschritt und Aufmerksamkeitsspanne individueller Lernender können dadurch datenbasiert untermauert, aber auch korrigiert werden.

Der Einsatz dieser Systeme kann dabei je nach Ziel, Gestaltung, und technischen Grundlagen sehr unterschiedlich ausfallen. Er reicht beispielsweise von Programmen für personalisiertes Lernen auf individuellen Endgeräten, die den Lernenden Wissen vermitteln oder bestimmte Fähigkeiten trainieren (adaptive Lernprogramme), über die Erfassung und Rückmeldung bestimmter Fortschritte, Schwächen und Stärken der Lernenden an Lehrkräfte, um individuelle Förderung zu erleichtern (Lernstandkontrolle), bis hin zur Überwachung der Aufmerksamkeit oder gar der Emotionen der Lernenden und Lehrenden durch Sensoren und Kameras. Hinzu

---

[227] Deutscher Ethikrat (2017): Big Data und Gesundheit – Datensouveränität als informationelle Freiheitsgestaltung. Berlin. Um den Datenumgang verantwortlich zu gestalten sind Grundkenntnisse über die Bedeutung und den Wert von Big Data und die damit verbundenen Risiken vorausgesetzt. D.h. der kritische Umgang mit Daten erfordert Datensouveränität. Dies hat der Ethikrat bereits in seiner Stellungnahme zu Big Data von 2017 festgehalten: „Da bereits Kinder digitale Anwendungen nutzen und dabei Daten generieren sollte eine entsprechende Nutzerkompetenz schon in der Schule vermittelt werden. Über die reinen technischen Aspekte der gängigen Digitalisierungsstrategie im schulischen Unterricht hinaus sollte dies als Querschnittsaufgabe für alle Fächer des schulischen Curriculums ausgestaltet sein, um der gerade bei Kindern und Jugendlichen virulenten informationellen selbst Gefährdung entgegenzuwirken und schon früh ein Bewusstsein für die rechtlichen, sozialen und ethischen Implikationen zu schaffen. Die Vermittlung solcher Nutzerkompetenz sollte daher zukünftig Teil der Lehraus- und -fortbildung werden.“ (271 f.)

[228] Strathmann, M. (2016): Alleine zur Fremdsprache. In: Zeit Online. www.zeit.de/digital/internet/2016-03/babbel-sprachkurs-start-up-berlin?utm_referrer=https%3A%2F%2F [11.01.2023].

kommen Anwendungen aus dem Bereich der Robotik, *Virtual Reality* oder *Augmented Reality*. Zu manchen Entwicklungen in diesen Bereichen gibt es erhebliche Kontroversen (vgl. Abschnitt 6.4).

Häufig enthalten Lernprogramme spielerische Elemente (sogenannte *Gamification*), um die Motivation zur Nutzung und damit zum Lernen zu steigern.[229] Durch Nudging sollen die Lernenden angeregt werden, das *Richtige* zu tun, das heißt, das gewünschte Lernverhalten auszuüben. Die Frage nach dem *richtigen* Lernen hat bereits Anlass zu Diskussionen dahingehend gegeben, ob Kinder und Jugendliche dadurch möglicherweise in ein „behavioristisches Dressursetting" hineingezogen werden.[230] Auch stellt sich die Frage, wer festlegt, was als *richtig* gelten darf. Im hier verwendeten Menschen- wie Bildungsbegriff ist bereits angelegt, an dieser Stelle kritisch gegenüber einem hohen Maß an Gamification und Nudging im Bildungsprozess zu sein. Auch hier sollte die Zielrichtung jedweden Einsatzes in der Erweiterung menschlicher Handlungsmöglichkeiten und Autorschaft liegen.

Auch wenn soziale Lernprozesse nicht durch KI ersetzt werden, werden soziale Beziehungen in Lernkontexten zunehmend durch KI vermittelt, beispielsweise durch die Messung von Leistung oder aber auch der Aufmerksamkeit oder Emotionen, die dann an Lehrkräfte und Lernende zurückgespiegelt wird. So können einerseits Informationen über die Gruppe, zum Beispiel eine Schulklasse, ausgewertet werden, um daraus Hinweise auf die Lerngeschwindigkeit und inhaltlichen Schwächen abzuleiten, mittels derer die Lehrkräfte ihre kurz-, mittel- und langfristigen Strategien besser an die jeweilige Gruppe anpassen können. Auf der individuellen Ebene können KI-gestützte Systeme dazu beitragen, individuelle Schwächen oder Rückstände zu identifizieren oder gar die Gefahr des Schulabbruchs frühzeitig anhand von Mustern zu erkennen und gezielte Unterstützungsangebote zu unterbreiten. Auf die Risiken solcher Ansätze wird im Folgenden noch eingegangen.

---

[229] Rachels, J.; Rockinson-Szapkiw, A. (2018): The effects of mobile gamification app on elementary students' Spanish achievement and self-efficacy. In: Computer Assisted Language Learning (31), 72–89.

[230] Hartong, S. (2019): Learning Analytics und Big Data in der Bildung- Zur notwendigen Entwicklung eines datenpolitischen Alternativprogramms. Herausgegeben von der Gewerkschaft, Erziehung und Wissenschaft. Frankfurt a. M. www.gew.de/index.php?eID=dumpFile&t=f&f=91791&token=702ec8d5f9770206a4aa8a1079750ec9021b90bf&sdownload=&n=Learning-analytics-2019-web-IVZ.pdf [11.01.2023] , 11 f.

6.4 Mensch-Maschine-Relationen in der schulischen Bildung: Ersetzen, Erweitern, Vermindern

Ähnlich wie im Medizinbereich muss auch im Bildungsprozess von unterschiedlichen Beziehungen ausgegangen werden, die durch den KI-Einsatz betroffen sind: Interaktionen zwischen Lehrenden und Lernenden wie auch der Lernenden untereinander im Klassenraum, im weiteren Sozialraum Schule, aber auch darüber hinaus. Hinzu kommen mitunter spezifische Untergruppen, wie Lernende mit besonderen Bedürfnissen und gegebenenfalls weitere Beteiligte wie zusätzliches (sozial-)pädagogisches Personal, Eltern, und Personen in der Schulverwaltung sowie jene, die an der Entwicklung und Planung von Curricula und dem Einsatz digitaler Techniken beteiligt sind. Parallelen zum Handlungsfeld Medizin lassen sich auch insofern feststellen, als es in der Schulbildung ebenfalls zu engen, mittleren und weitgehenden Ersetzungen bestimmter Handlungssegmente und Interaktionen kommen kann, die sich jeweils situationsabhängig erweiternd oder vermindernd auf die Handlungsmöglichkeiten der Beteiligten auswirken können.

Im Folgenden werden exemplarisch einige KI-basierte Lern- und Lehrsysteme für den schulischen Kontext dargestellt. Die genannten Möglichkeiten durch KI im System Schule sollten vor dem Hintergrund des zuvor dargelegten Menschenbildes und Bildungsbegriffs bewertet werden. Hier gilt insbesondere, dass die für Bildungsprozesse und Persönlichkeitsentwicklung grundlegenden Aspekte der direkten Ansprache sowie der personalen Beziehungen nicht vernachlässigt werden dürfen. Mit Blick auf sehr weitreichende Ersetzungen ist dabei auch auf die sich stellende Gefahr des Animismus, das heißt der Zuschreibung mentaler Eigenschaften an Maschinen, zu verweisen, die bei sich noch entwickelnden Kindern und Jugendlichen umso schwerer wiegt. Engere und gegebenenfalls mittlere Ersetzungen können aber durchaus mit den normativen Weichenstellungen vereinbar sein und den Bildungsprozess unterstützen.

Eine *enge Ersetzung* liegt etwa vor, wenn ein Softwaresystem für einen genau bestimmten Lernabschnitt eingesetzt wird wie das Vokabellernen im Fremdsprachenunterricht oder das Formellernen in der Mathematik. Ein Beispiel für solch ein eng umrissenes algorithmisches System, dass beim Erlernen einer ganz spezifischen Kompetenz zum schulischen Einsatz kommt, ist das Intelligente Tutorsystem (ITS) Subkraki[231], mit dessen Hilfe die Lernenden in der Grundschule das schriftliche Subtrahieren üben können. Das Programm stellt Übungsaufgaben, erkennt welche Art von Fehlern die Lernenden bei der Lösung der gestellten Aufgabe

[231] Zeller, C.; Schmid, U. (2016): Automatic Generation of Analogous Problems to Help Resolving Misconceptions in an Intelligent Tutor System for Written Subtraction. https://fis.uni-bamberg.de/handle/uniba/41882 [11.01.2023]. Für eine Demo-Version von Subkraki siehe https://cogsys.uni-bamberg.de/ITS [02.03.2023].

machen und gibt dann konstruktives Feedback, indem es beispielsweise spezifisch auf das einem beobachteten Fehler zugrundeliegende Verständnisproblem reagiert. Fehlerhafte Lösungen werden nicht einfach vom System korrigiert. Stattdessen werden die Lösungsschritte anhand einer ähnlichen Rechenaufgabe erläutert und den Lernenden damit die Möglichkeit gegeben, Fehler eigenständig zu korrigieren, daraus zu lernen und Selbstwirksamkeit zu entwickeln.

Aufwendigere und datenintensivere ITS (siehe Infokasten 4) können auch komplexere Lerninhalte in verschiedenen Fächern im Zusammenwirken mit Lernenden vermitteln und so breitere Teilaspekte des Unterrichtsgeschehens ersetzen (mittlere Ersetzung) oder im Einzelfall die Funktion einer Lehrkraft vollständig übernehmen *(*vollständige Ersetzung*)*. Eine weitgehende Ersetzung kann noch zugespitzt werden, wenn ein als Software realisiertes KI-System in bestimmten Situationen nicht nur als eigenständige Lehrkraft auftritt, sondern über einen virtuellen Avatar oder in einem Roboter mit besonders personalen Eigenschaften wie Mimik oder natürlicher Sprache in Erscheinung tritt.

Infokasten 4: Intelligente Tutorsysteme

Schon in den 1970er-Jahren wurden Intelligente Tutorsysteme (ITS) zur Unterstützung individueller Lernprozesse entwickelt. Aktuelle Fortschritte bei datengetriebenen, KI-gestützten Ansätzen haben die Möglichkeiten solcher Softwaresysteme deutlich erweitert. Ziel von ITS ist es, Inhalte auf eine Art und Weise zu vermitteln, die sich dynamisch an die individuellen Besonderheiten und Bedürfnisse der Lernenden anpasst.

Ein ITS besteht typischerweise aus vier Komponenten[232]: Sein Kern ist ein (1) Wissensmodell, das sämtliche Konzepte, Regeln und Problemlösungsstrategien für das zu vermittelnde Wissen umfasst. Aufgabenstellungen können vom Wissensmodell eigenständig gelöst werden. Es entspricht also einem Expertensystem für den zu vermittelnden Wissensbereich – sei es analytische Geometrie, Hebelgesetze oder die Entstehung von Regen. Das Computerprogramm gleicht dieses Wissensmodell fortlaufend mit den verfügbaren Daten der Lernenden ab, die es in einem (2) Studierendenmodell zusammenfasst. Dabei analysiert das System, wie Lernende sich zum Wissensmodell verhalten – welche Fortschritte oder Fehler sie beispielsweise machen – und aktualisiert das Studierendenmodell immer wieder entsprechend. Im einfachsten Fall ist dies ein Fortschrittsprofil. Es können aber auch komplexe intelligente Methoden zum Einsatz kommen, die die Identifikation von Fehlkonzepten ermöglichen. In der (3) Didaktikkomponente wählt das System auf Grundlage dieses Abgleichs bestimmte Rückmeldungen an die Lernenden (sowie gegebenenfalls auch an Lehrkräfte) aus, etwa Einschätzungen zum Fortschritt, Hinweise auf Fehler und Hilfestellungen zu ihrer Überwindung oder eine Auswahl weiterführender Schritte oder Aufgaben. An (4) der Benutzerschnittstelle schließlich findet der Austausch zwischen dem Tutorsystem und den Lernenden statt, beispielsweise in Form von Texteingaben und -ausgaben.

[232] Nkambou, R., Mizoguchi, R., & Bourdeau, J. (2010). Advances in intelligent tutoring systems. Heidelberg, 3-5.

+++ Vorabfassung! Wird durch die lektorierte Fassung ersetzt. +++

Während in frühen ITS sämtliche möglichen Bezüge zwischen dem Wissens- und dem Studierendenmodell ebenso wie alle Ausgabemöglichkeiten vorab bestimmt und programmiert werden mussten, zum Beispiel in Form von Entscheidungsbäumen, erlauben moderne datenintensive und KI-gestützte Ansätze flexiblere und differenziertere Funktionen. Je nach Menge und Vielfalt der Daten, die Lernende in das System einspeisen, können der individuelle Wissensstand sowie Fortschritte und gegebenenfalls auch Präferenzen beim Lernen genauer analysiert werden. Dadurch kann das Programm präzisere Vorhersagen für den weiteren Lernprozess treffen und seine Ausgaben entsprechend individualisieren.

Mit seinen vier Komponenten bildet ein ITS einige Kernfunktionen nach, die sonst von menschlichen Lehrkräften übernommen werden – (1) die Bereithaltung von Fachwissen, (2) die Einschätzung des Lernstands und Förderbedarfs von Lernenden, (3) eine daran angepasste Selektion didaktischer Strategien und (4) die Interaktion mit den Lernenden. ITS können daher mit dem Anspruch eingesetzt werden, die Funktion einer Lehrkraft für den inhaltlich abgedeckten Bereich ganz oder teilweise zu ersetzen. Genauso ist jedoch ein Einsatz in hybriden Formaten und/oder mit Schwerpunkten bei der Unterstützung der Lehrkraft (siehe Haupttext) möglich.

Als Beispiel für ein ITS-System mit weitreichendem Ersetzungspotenzial kann AutoTutor genannt werden. Es wurde von Arthur Graesser am Institute for Intelligent Systems der Universität von Memphis entwickelt. Das Programm simuliert eine menschliche Lehrkraft, indem es mit den Lernenden ein natürliches Gespräch führt.[233] Dabei erfolgt eine Anpassung an die Lernenden, indem ihre Emotionen und Reaktionen (Dialogmuster, Mimik, Körperhaltung) erfasst werde. Bisher mit dieser Methode unterstützte Inhalte aus dem schulischen und außerschulischen Bereich stammen dabei unter anderem aus der Informatik, Physik, Mathematik, Elektronik, Philosophie und Biologie. AutoTutor bietet darüber hinaus die Möglichkeit, jeweils die Über- und Unterforderung der Lernenden zu registrieren und darauf zu reagieren. Auf diese Weise eingesetzt, bieten intelligente Tutorsysteme das Potenzial, positive Lernumgebungen zu schaffen. ITS bieten die Möglichkeit, spezifische Lerninhalte von Mathematik bis Fremdsprachenerwerb in spezifischen Aufgabenkontexten zu üben und dabei unmittelbares und individualisiertes Feedback zu erhalten. Auf diese Weise sind sie geeignet, den Unterricht mit einer Lehrkraft im Klassenverband gezielt um Phasen des individualisierten Lernens zu ergänzen. Kompetenzen werden konstruktiv (learning by doing) erworben und ergänzen das learning by being told des üblichen Unterrichts.

Intelligente Tutorsysteme und verwandte Daten- und KI-gestützte adaptive Lerntechnologien werden zunehmend weniger mit dem Ziel entwickelt, eine menschliche Lehrkraft möglichst weitgehend zu ersetzen, sondern immer stärker für den Einsatz in hybriden Lehr-Lern-Systemen konzipiert.[234] Solche Systeme werden in zahlreichen Ländern bereits mit gut dokumentierten Beiträgen zur Verbesserung von Lernerfolgen bei unterschiedlichen Altersgruppen eingesetzt. Das 2012 aus einer Elterninitiative in den Niederlanden hervorgegangene Angebot

[233] Graessner, A. C. et al. (2012): AutoTutor. In: McCarthy, P. M.; Boonthum-Denecke, C. (Hg.): Applied Natural Language Processing: Identification, Investigation and Resolution. Hershey, 169-187 (DOI: 10.4018/978-1-60960-741-8.ch010).

[234] Vgl. Molenaar, I. (2021): Personalisation of learning: Towards hybrid human-AI learning technologies. In: OECD Digital Education Outlook 2021: Pushing the frontiers with AI, blockchain, and robots. Paris, 57-77 (DOI: 10.1787/589b283f-en).

Snappet beispielsweise wird inzwischen in mehreren Ländern, darunter auch Deutschland, zur Unterstützung des Unterrichts in der Grundschule verwendet. Die Software wird von Lernenden im Präsenz-, Distanz- oder Wechselunterricht zur Bearbeitung von Aufgaben eingesetzt, die auf die Lehrpläne und den individuellen Lernfortschritt in allen Fächern zugeschnitten sind. Sie spielt die Auswertung der dabei gesammelten Daten an die Lehrkräfte zurück, um diese bei der weiteren Unterrichtsplanung und differenzierten Förderung der Kinder zu unterstützen. In einer sechsmonatigen Studie an 79 Grundschulen konnten mit dem Einsatz von Snappet unter anderem in einer standardisierten Prüfung verbesserte Ergebnisse und eine höhere intrinsische Motivation der Lernenden nachgewiesen werden.[235]

Auch eine weiter reichende Nutzung von Technik zur erfolgreichen Teilhabe am Schulunterricht kann im Einzelfall sinnvoll sein. Dies kann mit dem im Kasten 5 beschriebenen Szenario veranschaulicht werden, in dem ein Telepräsenzroboter Lernenden, die nur von zu Hause am Unterricht teilnehmen können, mehr Interaktion mit der Lerngruppe ermöglicht.

Infokasten 5: Teilnahme am Unterricht mittels eines Telepräsenzroboters

Mittels eines Telepräsenzroboters (z. B. AV-1-Roboter des norwegischen Startups No isolation) können Lernende im Falle längerer Abwesenheiten, beispielsweise aufgrund von Krankheit, ihre Klasse virtuell besuchen. Der Roboter übernimmt den Platz des Lernenden in der Klasse, ausgestattet mit Internetverbindung, Kamera, Lautsprecher, Mikrofon. Die Lernenden können damit im Klassenzimmer in Echtzeit hören, sehen und sprechen (via App am Smartphone oder Tablet von zu Hause aus, ohne Datenspeicherung). Die Funktionen des Roboters umfassen das Herumblicken im Klassenzimmer, das Handheben und Sprechen oder auch Mimik, z. B. um Unverständnis auszudrücken).[236]

Soziale Isolation zum Beispiel infolge einer chronischen Erkrankung kann durch diesen Telepräsenzroboter abgemildert, eine Wiedereingliederung unterstützt bzw. der Platz im Klassenzimmer symbolisch vom Telepräsenzroboter besetzt gehalten werden. Der Telepräsenzroboter unterstützt so nicht nur kognitive, sondern in beschränktem Ausmaß auch emotionale und soziale Funktionen, indem er beispielsweise das Tuscheln mit anderen Kindern aus der Klasse ermöglicht. Es geht also nicht nur darum, im Lernpensum nicht zurückzubleiben, sondern auch darum, weiterhin als Teil der Klassengemeinschaft wahrgenommen zu werden. Während der AV-1-Roboter menschenähnliche Züge aufweist, wird hierauf in anderen Projekten explizit verzichtet, um anthropomorphen Fehlverständnissen vorzubeugen.

[235] Faber, J. H. et al. (2017): The effects of a digital formative assessment tool on mathematics achievement and student motivation: Results of a randomized experiment. In: Computers & Education 106, 83-96 (DOI: 10.1016/j.compedu.2016.12.001).
[236] Belpaeme, T.; Tanaka, F. (2021): Social Robots as Educators. In: OECD Digital Education Outlook 2021: Pushing the Frontiers with Artificial Intelligence, Blockchain and Robots. Paris, 143-157 (DOI: 10.1787/589b283f-en), 148.

Neben den zuvor beschriebenen Möglichkeiten für den Einsatz datenbasierter und KI-gestützter Technologien in der personalisierten Gestaltung und Auswertung konkreter Lerninhalte und -prozesse, gibt es mittlerweile auch Bestrebungen, KI zur Analyse des Verhaltens im Klassenraum oder ganzer Einrichtungen einzusetzen. Wie umfassend dies geschehen kann, zeigen Beispiele aus China.

Infokasten 6: Besonders weitreichender Einsatz digitaler Technologien im Schulunterricht

Teilweise kann ein besonders weitreichender Einsatz von datenintensiven und KI-gestützten Tools im Schulunterricht beobachtet werden, der über die Verwendung einer geschlossenen Lehr-Lern-Software hinausgeht.[237] So werden beispielsweise an manchen chinesischen Schulen mithilfe zahlreicher und vielfältiger Sensoren umfangreiche Daten auf dem Campus gesammelt. Dies geschieht einerseits, um die Ressourcen auf dem Gelände zu verwalten. Dazu gehören unter anderem die Sicherheitseinrichtungen, die Beleuchtung, die Regelung der Wasser- und Luftqualität, aber auch das Sammeln von Bewegungsdaten. Im Klassenzimmer erfolgt eine ähnlich umfassende Datensammlung, beispielsweise über Kameras und Mikrofone und teilweise ergänzt um individuelle, von den Lernenden getragene Geräte (Wearables), um ein möglichst umfassendes Bild eines jeden Individuums zu erstellen. Es werden Körperdaten (Herzfrequenz, Körpertemperatur) ermittelt, aber auch Daten zum Sozialverhalten, der emotionalen und psychischen Gesundheit anhand von Analysen der Mimik, Körperhaltung und Sprache sowie natürlich der schulischen Leistungen. Genannte Ziele dieser Maßnahmen umfassen die Sicherstellung der Aufmerksamkeit der Schülerinnen und Schüler, die individuelle Förderung der Lernenden sowie die Unterstützung der Lehrkräfte, beispielsweise bei Aufgaben wie der Unterrichtsvorbereitung, dem Vergeben von Hausaufgaben und der Auswertung von Lernprozessen.

Ein Schwerpunkt solcher Ansätze wird mit der Bezeichnung *classroom analytics* beschrieben. Damit sind Konzepte gemeint, welche die Dynamik ganzer Lerngruppen umfassend zu dokumentieren und auszuwerten versuchen, unter Berücksichtigung des Verhaltens und der Interaktion der Lernenden und Lehrkräfte im Klassenraum[238], einschließlich der räumlichen Gegebenheiten und Prozesse vor Ort[239]. Classroom-Analytics-Ansätze sind aufgrund der für sie notwendigen Erfassung vielfältiger Daten unter anderem über das Verhalten von Schülerinnen und Schülern sowie Lehrkräften umstritten (siehe unten). Befürworter solcher Ansätze verweisen auf die Chancen, die sich durch Classroom Analytics hinsichtlich einer Verbesserung von

[237] Vincent-Lancrin, S. (2021): Frontiers of smart education technology: Opportunities and challenges. In: OECD Digital Education Outlook 2021: Pushing the Frontiers with Artificial Intelligence, Blockchain and Robots. Paris, 19-42 (DOI: 10.1787/589b283f-en), 24 f.
[238] Dillenbourg, P. (2021): Classroom analytics: Zooming out from a Pupil to a Classroom. In: OECD Digital Education Outlook 2021: Pushing the Frontiers with Artificial Intelligence, Blockchain and Robots, Paris, 105-122 (DOI: 10.1787/589b283f-en).
[239] Martinez-Maldonado, R. et al. (2021): Moodoo the Tracker: Spatial Classroom Analytics for Characterizing Teachers` Pedagogical Approaches. In: International Journal Artificial Intelligence in Education 32, 1025–1051 (DOI: 10.1007/s40593-021-00276-w).

Pädagogik und Didaktik ergeben können. Beispielsweise könne Classroom Analytics die Kapazität von Lehrkräften steigern, alle Lernenden gleichzeitig im Blick zu haben und sie so im gesamten Management ihrer Lehrkonzeption zu unterstützen, etwa bei der Bildung von Lerngruppen oder bei der Frage, wann in die Lernprozesse der Schülerinnen und Schüler eingegriffen werden soll.

Mittels Classroom Analytics können der Lehrkraft Informationen an die Hand gegeben werden, die auch zur kritischen Selbstreflexion des Lehrverhaltens genutzt werden können. Für computergestützte Analysen ist beispielsweise das Verhalten des lehrenden Fachpersonals schwierig zu erfassen, wenn nur die verbale Kommunikation zugrunde gelegt wird. Indem jedoch mittels Classroom Analytics auch das Verhalten im Klassenraum beobachtet wird, zu dem auch die Häufigkeit der persönlichen Kontakte mit den Schülerinnen und Schülern gehören, kommen zusätzliche Parameter ins Spiel, die pädagogische Ansätze und Lehrstrategien messbarer werden lassen. Damit kann den Lehrkräften in konstruktiver Weise ein Feedback gegeben werden. Ein solches Instrument ist etwa das TeachActiveDashboard. Es kann die pädagogischen Bemühungen durch visuelle Unterstützung verbessern. Den Lehrkräften wird mit den Analysen gespiegelt, ob und wie ihre pädagogischen Praktiken zum aktiven Lernerfolg der Lernenden beitragen.[240] Classroom Analytics eröffnen Lehrkräften zudem Chancen auf eine objektivere Wahrnehmung der Lernenden und können es ermöglichen, implizite Vorurteile der Lehrkräfte gegenüber Lernenden, etwa zu Herkunft und Gender, aufzudecken.[241] Wenn solches Feedback den Lehrkräften zurückgespiegelt wird, kann dies als kritisches Korrektiv für das eigene Handeln dienen.

In diesem Zusammenhang sind unterschiedliche ethische Fragen relevant. Da im Rahmen von Classroom Analytics mitunter besonders umfangreiche Daten erfasst und ausgewertet werden, sind zum einen negative Auswirkungen solcher Messungen auf die Privatsphäre und Autonomie aller Beteiligten zu diskutieren. Neben den direkten Auswirkungen der Datenerfassung und Prognostik auf die Privatsphäre und Autonomie der Beobachteten, ist hier insbesondere auch auf sogenannte *Chilling*-Effekte hinzuweisen. Im Kontext von Überwachung beschreiben Chilling-Effekte- das Phänomen, dass bereits die Sorge, bestimmte Daten über Personen könnten

[240] Alzoubi, D. et al. (2021): TeachActive Feedback Dashboard: Using Automatic Classroom Analytics to Visualize Pedagogical Strategies at a Glance. In: ACM Conference on Human Factors in Computing Systems (CHI), Artikel 312, 1-6 (DOI: 10.1145/3411763.3451709).
[241] Reinholds, D. et al. (2020): Walking The Walk: using Classroom Analytics to Support instructors to address implicit bias in teaching. In: The International Journal for Academic Development 25, 259-272 (DOI: 10.1080/1360144X.2019.1692211).

erfasst und ihr Verhalten analysiert werden, negative Rückwirkungen auf ihr Befinden und Verhalten haben kann. Dies könnte eine Verminderung von Handlungsfreiheit, Motivation und anderen für gute Bildung wichtigen Bedingungen mit sich bringen. Darüber hinaus können Daten in unangemessener Weise zweck- und kontextentfremdet genutzt werden, wenn zum Beispiel Informationen über mangelhafte Schulleistungen oder Lehrkraftevaluierungen zu Strafen oder anderweitigen Sanktionen führen. Ein weiteres Problem stellen systematische Verzerrungen (Bias-Problematik) sowohl in der Datenanalyse als auch in der Interpretation von Daten und Prognosen dar, welche zu ungerechten Urteilen über die Leistung von Lernenden oder Lehrenden führen können.

Um diesen Herausforderungen zu begegnen, bedarf es eines verantwortlichen Umgangs mit Daten und Analysen, der die Bedürfnisse von Lernenden und Lehrenden stets im Blick hat. Diese Personengruppen sollten bereits während der Entwicklung digitaler Bildungsinstrumente mit einbezogen werden. Die ethische Angemessenheit von Lösungen hängt dabei auch immer von technischen und institutionellen Details ab: welche Daten über wen in welcher Granularität erfasst werden (z. B. individuell, gruppen- oder aufgabenbezogen) und wer in welcher Granularität und für welche Zwecke Zugang zu den Daten, Analysen oder Prognosen erhält.

Ein besonders kontrovers diskutierter Aspekt von Classroom Analytics betrifft in diesem Zusammenhang die mögliche Erfassung der Aufmerksamkeit (Attention Monitoring) oder der emotionalen Verfasstheit (Affect Recognition) der im Klassenraum interagierenden Personen, insbesondere basierend auf der Analyse von Video- oder Audiodaten aus Klassenräumen. Die Bedeutung von Emotionen für Lernprozesse wurde in den letzten Jahrzehnten immer stärker berücksichtigt und entsprechend ist es nicht verwunderlich, dass nicht nur die Messung von Aufmerksamkeit, sondern auch von Emotionen in schulischen Kontexten voranschreitet. Die Verfügbarkeit von Sensoren wie Kameras und Mikrofonen in Laptops und Tablets, aber auch deren Einbau in Schulräumen scheinen die Möglichkeit der Erfassung von Aufmerksamkeit und Emotionen nahezulegen, um neben der Erfassung von Leistung zusätzliche Einsichten in das Lerngeschehen zu gewinnen. Auch wenn dies durchaus mit dem Ziel einer Verbesserung von Lernergebnissen verbunden sein kann, so ist der Einsatz solcher Technologien in schulischen Kontexten aufgrund erkenntnistheoretischer und ethischer Herausforderungen kritisch zu betrachten.

Die *erkenntnistheoretische* Kritik insbesondere an Technologien zur Affekterkennung bezieht sich hier vor allem auf das vorherrschende reduktionistische Modell menschlicher Emotionen, nach dem es sechs Grundemotionen gibt, die als universell stabil gelten und die zuverlässig aus

äußerem Verhalten wie insbesondere der Mimik erschlossen werden können. Der Kritik an diesem Modell wird mittlerweile in der Forschung durchaus Rechnung getragen, etwa mit der Erfassung situativer Kontexte oder der Zusammenführung verschiedener Datentypen, beispielsweise visueller und verbaler Daten mit dem Ziel, die Validität der Affektmessung zu erhöhen. Aber selbst wenn Emotionen in Zukunft zuverlässig durch KI-Systeme erfasst werden könnten, so sind dennoch *ethische* Bedenken hinsichtlich ihres Einsatzes in schulischen Kontexten zu berücksichtigen, da das hierfür notwendige Monitoring von Lernenden und Lehrenden Auswirkungen auf deren Privatsphäre und Autonomie haben kann und Fragen der Gerechtigkeit berührt (siehe unten).

Die Bewertung des Einsatzes von Technologien zur Affekt- und Aufmerksamkeitserkennung in schulischen Kontexten hängt daher von der Beantwortung folgender Fragen ab: Können und sollten Aufmerksamkeit und Emotionen genau, zuverlässig und ohne systematische Verzerrung gemessen werden? Ob sie messbar *sind*, hängt von der wissenschaftlichen und technologischen Basis der eingesetzten Systeme ab.

Ob solche Instrumente eingesetzt werden *sollen*, hängt jedoch von einer zusätzlichen Bewertung des potenziellen Nutzens und Schadens für alle Beteiligten ab. Erstens müsste nachgewiesen werden, dass das Monitoring von Aufmerksamkeit und Emotionen insbesondere den Schülerinnen und Schülern nützt, zum Beispiel indem es ihre Lernprozesse oder Lernergebnisse verbessert. Zweitens müsste bewertet werden, ob diese potenziellen Vorteile die potenziellen Nachteile überwiegen, wie etwa die negativen Auswirkungen der notwendigen Datenerfassung auf die Privatsphäre, auf die Autonomie und Freiheit der Lernenden. Dazu gehören auch Fragen zu Chilling-Effekten und den langfristigen Folgen des Messens und Spiegelns insbesondere von Emotionen für Lernende und Lehrende. Drittens ist es nicht nur ein erkenntnistheoretisches, sondern auch ein ethisches Problem, wenn Technologien zur Aufmerksamkeits- und Emotionserfassung ungenau oder verzerrt sind. Denn sobald solche Analysen zu weiteren Handlungen bzw. zur Bewertung der Leistung führen und diese Analysen ungenau oder verzerrt sind, wirft dies auch Fragen der Gerechtigkeit auf.

Der Kontext von Schulen bringt zwei zusätzliche Herausforderungen mit sich. Zum einen sind Schulkontexte durch Hierarchien und asymmetrische Machtverhältnisse gekennzeichnet. Zum anderen sind Kinder beteiligt, die besonders vulnerabel sind, weshalb höhere Anforderungen an den möglichen Nutzen solcher Technologien für die beteiligten Kinder zu stellen sind.

Auch wenn ein Einsatz aktuell verfügbarer Technologien im schulischen Kontext in Anbetracht der oben genannten Kriterien nicht empfehlenswert erscheint, ist nicht prinzipiell auszuschließen, dass zukünftige Entwicklungen zu einer didaktisch sinnvollen Verbesserung des Lernprozesses beitragen.

In Anbetracht der erkenntnistheoretischen und ethischen Herausforderungen und unter Abwägung potenzieller Nutzen und Schäden stehen die Mitglieder des Deutschen Ethikrates dem Einsatz von Audio- und Videomonitoring im Klassenzimmer insgesamt skeptisch gegenüber. Insbesondere erscheint die Analyse von Aufmerksamkeit und Emotionen per Audio- und Videoüberwachung des Klassenraums mittels aktuell verfügbarer Technologien nicht vertretbar. Ein Teil des Ethikrates schließt den Einsatz von Technologien zur Aufmerksamkeits- und Affekterkennung zukünftig jedoch nicht vollständig aus, sofern sichergestellt ist, dass die erfassten Daten eine wissenschaftlich nachweisbare Verbesserung des Lernprozesses bieten und das hierfür notwendige Monitoring von Schülerinnen und Schülern sowie Lehrkräften keine inakzeptablen Auswirkungen auf deren Privatsphäre und Autonomie hat. Ein anderer Teil des Ethikrates hingegen befürwortet ein Verbot von Technologien zu Aufmerksamkeitsmonitoring und Affekterkennung in Schulen.

6.5 Grundsätzliche Diskussion: KI im schulischen Bildungsprozess

Die öffentliche Debatte zu KI in der Bildung hat sich in den letzten Jahren intensiviert und wird hauptsächlich in den Fach-Communities und von Stakeholdern betrieben. So hat sich die Gewerkschaft Erziehung und Wissenschaft im Projekt „Bildung in der digitalen Welt“ mit dem Thema *Learning Analytics* auseinandergesetzt.[242] Auch im Abschlussbericht der Enquete-Kommission Künstliche Intelligenz des Deutschen Bundestages[243] wird das Potenzial von Learning-Analytics-Systemen hervorgehoben.

Auf der Seite der Chancen ist das personalisierte Lernen anzuführen, das im Sinne des hier skizzierten Feldes bedeutet, dass durch personalisierte Abstimmung auf Einzelne Lernchancen erhöht werden können. So können auf unterschiedliche Weise Stärken wie Schwächen des je einzelnen Lernenden wie auch einer Lerngruppe erkannt und auf existierende Schwierigkeiten beim Aufgabenlösen eingegangen werden. Dies kann einerseits Lernende unterstützen, sowohl

[242] Hartong, S. (2019): Learning Analytics und Big Data in der Bildung – Zur notwendigen Entwicklung eines datenpolitischen Alternativprogramms. Frankfurt. a.M.
[243] Deutscher Bundestag (2020): Bericht der Enquete-Kommission Künstliche Intelligenz – Gesellschaftliche Verantwortung und wirtschaftliche, soziale und ökologische Potenziale. Bundestagsdrucksache 19/23700. Berlin. https://dserver.bundestag.de/btd/19/237/1923700.pdf [22.02.2023], 306.

in Bezug auf individuelle Stärken als auch besondere Bedürfnisse, da Lernende fokussiert in ihren Bedarfen von einem nicht müde oder ungeduldig werdenden System begleitet werden. Mithilfe dieser Prozesse werden auch Lehrkräfte unterstützt, die ihre Ressourcen dann sinnvoll im Gesamtlehrprozess einsetzen können. Als Beispiel sei das schon erwähnte Subkraki angeführt, das als eine Form der engen Ersetzung[244] der Lehrenden das Lehr-Lern-Geschehen punktuell unterstützen kann und daher eine Entlastung der analogen Lehre für alle Beteiligten bietet. Eine weitere Chance besteht darin, dass KI-gestützte Lehr-Lern-Systeme für bestimmte Bevölkerungsgruppen die Zugangschancen zu Bildung verbessern können, selbst wenn sie nicht als Teil des Unterrichtsprogramms von der Schule angeboten werden. Auch wenn derartige Programme, etwa zum Spracherwerb oder zur Nachhilfe, meist nicht kostenfrei verfügbar sind, liegen die Kosten in der Regel weit unterhalb der Kosten für private Nachhilfe oder Sprachkurse.

Soweit mit dem Einsatz KI-gestützter Lehr-Lern-Systeme einzelne Lernende in den Fokus gerückt werden, ist ihr Einsatz im Besonderen auch unter dem Gesichtspunkt der Inklusion bedenkenswert. So können KI-basierte Lehr-Lern-Systeme für Lernende mit besonderen Bedürfnissen, deren individuelle Lerngeschichten adaptiver begleitet werden müssen, Chancen bieten.[245] KI kann bei der Entwicklung sozialer Fähigkeiten von autistischen Kindern helfen, Lernende mit Dysgrafie diagnostizieren und unterstützen sowie die Teilhabe für blinde oder sehbehinderte Kinder durch Bereitstellung grafischen Materials fördern.

Es bleibt deshalb grundsätzlich festzuhalten: Die neuen Systeme können hilfreiche, auf das Individuum zugeschnittene Einzeltools darstellen, sind aber nicht die generelle Lösung für Fragen von Inklusion, und sie müssen unter Berücksichtigung möglichen Missbrauchs und möglicher Risiken angewendet werden. Um diese zu erkennen und im Blick zu behalten, ist es wichtig, das digitale Lehr- und Lerngeschehen permanent zu evaluieren.

Eine weitere Chance, die der Einsatz technischer Systeme für Inklusion mit sich bringt, liegt in der Flexibilität und einem erweiterbaren Aktionsradius. Vor allem das Beispiel mit dem Telepräsenz-Roboter zeigt, wie Technik Inklusion erhöhen kann, indem sie es beispielsweise chronisch kranken Lernenden ermöglicht sich von zu Hause aus am Unterrichtsgeschehen zu beteiligen.

---

[244] Kontrolle von LernApps geringer verglichen mit der Kontrolle von Schulbüchern.
[245] Good, J. (2021): Serving students with special needs better: How digital technology can help. In: OECD Digital Education Outlook 2021: Pushing the Frontiers with Artificial Intelligence, Blockchain and Robots. Paris, 123-142 (DOI: 10.1787/589b283f-en).

Die Verwendung von Lehr- und Lerntechnologien ist also mit großen Chancen verbunden. Dazu gehört auch der berechtigte Wunsch nach einer objektiveren und faireren Bewertung der Lernergebnisse. Da Software nicht die möglichen Vorurteile der Lehrkräfte „hegt", sei es bezüglich der Intelligenz der Lernenden, seien es emotionale Vorbehalte, die eine Gleichbehandlung trüben können, steigt die Chance, dass Lernende mithilfe von KI-Anwendungen neutraler unterrichtet werden können. Allerdings darf auch hier eine solch vorgebliche Neutralität oder Objektivität nicht als gegeben angenommen werden, sondern bedarf der Überprüfung. Ein besonderes Risiko bei datenbasierten KI-Systemen besteht nämlich in systematischen Verzerrungen (Bias), woraus sich hohe Anforderungen an die Validität und Qualität der verwendeten Trainingsdaten und die Angemessenheit der verwendeten Methoden ergeben. Auch weitere allgemeine Risiken, wie beispielsweise Fragen der Sicherheit, Privatheit und Transparenz zeigen sich natürlich ebenso beim Einsatz von KI in der Bildung.

Darüber hinaus ergeben sich im Kontext der (schulischen) Bildung weitere sektorspezifische Herausforderungen. Hier sind in besonderer Weise die sozialen Aspekte hervorzuheben, die beim Ersatz des Lehrpersonals eintreten können, insbesondere Gefahren der Isolation und Vereinsamung von Lernenden.

Außerdem ist zu berücksichtigen, dass bereits der Einsatz digitaler Medien das qualitative Ergebnis des Lernprozesses verändern kann. So werden Texte, die über elektronische Medien rezipiert werden, anders durchdrungen als Texte, die mittels analoger Medien (auf Papier) gelesen werden.[246] Auch beim Einsatz von KI muss in Betracht gezogen werden, dass sich Lernverhalten qualitativ verändert. So könnten sich etwa grundsätzliche Auswirkungen auf die Motivation und Fähigkeit von Schülerinnen und Schülern, komplexere Aufgaben zu lösen, ergeben. Wenn KI vor allem auf die Lösung kleiner, überschaubarer Aufgaben und schnelle Erfolgserlebnisse ausgerichtet ist, könnte langfristig die Bereitschaft vermindert werden, komplexere Probleme anzugehen, deren Lösung in der Regel nicht innerhalb kurzer Zeit gefunden werden kann.

Die Erforschung der Auswirkungen des Einsatzes digitaler Medien wie auch KI-basierter Methoden muss daher vielfältige, auch indirekte Dimensionen erfassen und darf sich nicht nur an wenigen offensichtlichen Markern schulischen Erfolgs (wie etwa den Schulnoten oder dem Abfragen von Fakten) orientieren. Sind diese Aspekte jedoch erst einmal hinreichend verstanden,

[246] COST E-READ (2018): Stavanger Declaration Concerning the Future of Reading. In: cost – European Cooperation in Science and Technology. https://ereadcost.eu/wp-content/uploads/2019/01/StavangerDeclaration [12.01.2023].

so bieten sich gerade Methoden der KI an, um spezifische Defizite eines stärker digitalisierten Lernens zu kompensieren und etwa die individuelle Kompetenz zum Lösen komplexerer Aufgaben zu fördern.

Auch die Messung von Aufmerksamkeit und Emotionen kann ein Mittel sein, Lernprozesse und Lernerfolge besser zu verstehen. Allerdings sind Technologien der Aufmerksamkeits- und Emotionserkennung beispielsweise durch Kameras oder andere Instrumente wissenschaftlich umstritten[247] und greifen in invasiver Weise in die Autonomie und Privatheit der Lernenden ein. Eine solche invasive Form der Datenerfassung ist mit dem hier leitenden Bildungsbegriff, der sich am freien und urteilsfähigen Menschsein orientiert, nicht kompatibel und muss daher abgelehnt werden.[248]

Neben Fragen des Zugriffs auf Daten (z. B. durch Lehrkräfte, Eltern, Behörden) und dessen Auswirkungen auf Fragen der Privatsphäre müssen auch die Auswirkungen einer so engmaschigen Überwachung für die Lernatmosphäre und die Selbstwahrnehmung der Lernenden berücksichtigt werden. Hier spielen auch Fragen der Arbeitsplatzüberwachung eine Rolle, wenn auch Lehrpersonen in den Fokus der Technologie geraten. Hinzu kommen Zweifel an der Validität bestimmter Daten und Analysen. Diskutiert wird etwa, ob der Einsatz von EEG-Headsets und anderen Technologien zur Messung von Aufmerksamkeit oder emotionalem Engagement überhaupt aussagekräftige Daten generiert.[249] Insbesondere Bemühungen, den individuellen Lernerfolg auf Grundlage solcher Daten vorherzusagen[250], sind kritisch zu beurteilen. Lernverhalten und -erfolg dürfen nicht auf einfach erfassbare Daten reduziert werden. Einem solchen Reduktionismus liegt eine behavioristische Betrachtungsweise des Menschen zugrunde. Kog-

---

[247] Tcherkassof, A.; Dupré, D. (2022): The emotion–facial expression link: evidence from human and automatic expression recognition. In: Psychological Research 85, 2954–296 (DOI: 10.1007/s00426-020-01448-4); Mohammad, S. M. (2021): Ethics sheet for automatic emotion recognition and sentiment analysis. In: Computational Linguistics 48 (2), 239–278 (DOI: 10.1162/coli_a_00433).

[248] Es handelt sich aber um einen Verhaltenskodex für Hochschulen: Hansen, J. (2020): Verhaltenskodex für Trusted Learning Analytics. Version 1.0. Entwurf für die hessischen Hochschulen. Frankfurt a. M. (DOI: 10.25657/02:18903); Hartong, S. (2019): Learning Analytics und Big Data in der Bildung – Zur notwendigen Entwicklung eines datenpolitischen Alternativprogramms. Frankfurt. a. M. Der Einsatz umfangreicher Datenerhebungen und -auswertungen bringt vielfache Herausforderungen mit sich. Entsprechend spielen Datenschutzfragen und Sorgen vor einem möglichen Missbrauch von KI-gestützten Lehr-Lern-Systemen, vor allem bei flächendeckender Einführung, eine große Rolle im Fachdiskurs, ebenso wie eine mögliche Regulierung zur Verhinderung solcher Probleme. Im Vorschlag für einen Verhaltenskodex für *Trusted Learning Analytics* wird auf Kompetenzaufbau als Voraussetzung für einen verantwortungsvollen Umgang mit KI-gestützten Technologien gesetzt statt auf Regulierung.

[249] Liu, L. (2019): Orwellian Nonsense or Innovation in the Classroom? In: EETimes China. https://www.eetasia.com/orwellian-nonsense-or-innovation-in-the-classroom/ [12.01.2023].

[250] Gray, C. C.; Perkins, D. (2019): Utilizing early engagement and machine learning to predict student outcomes. In: Science Direct 131, 22-32 (DOI: 10.1016/j.compedu.2018.12.006).

nitive Verstehensprozesse und konstruktive Lernszenarien werden dabei unter Umständen zurückgedrängt. Wenn es jedoch um die Reichweite von Bildung geht, die auch der Persönlichkeitsentwicklung und der Bildung junger Menschen zu verantwortlichen Personen mit Urteilskraft dienen soll, dann sind die personale Vermittlung und die personalen Aspekte von Bildung nicht zu vernachlässigen.

Datengetriebene, KI-gestützte Lehr-Lern-Systeme können den jeweiligen Lernprozess unterstützen. Die Bildungsvorteile hinsichtlich der Wissens- und Informationsvermittlung durch den Einsatz digitaler Werkzeuge sind nicht zu unterschätzen. Sie ersetzen aber nicht die personale Vermittlung und die personalen Aspekte von Bildung. Das analoge Gespräch ist unverzichtbar für das motivations- und identitätsstiftende Potenzial, das der Unterricht bereitstellen sollte. Verstehensprozesse, deren Überprüfung ein personales Gegenüber braucht, können von digitalem Lernen alleine nicht angestoßen werden. Die Verantwortung für die Lernprozesse liegt bei den Menschen, die die Wahrung und Förderung der Selbstbestimmung auch aufseiten der Lernenden im Blick haben sollten.[251] Wenn Bildung als Beitrag zu einem mündigen Person-Sein Bestand haben soll, kann sie nicht vollständig automatisiert und an Maschinen delegeiert werden, da personale Bildung (stets auch) auf interpersonale und pädagogische Aktion und Interaktion angewiesen ist. Die Relevanz der Schule als Sozialraum der Interaktion zwischen Menschen ist dabei nicht zu unterschätzen. All diese Aspekte in ihrer Bedeutung für den Bildungsbereich hat die Corona-Pandemie unterstrichen, vor allem die Notwendigkeit des sozialen Miteinanders.[252]

6.6 Fazit und Empfehlungen

In dieser Stellungnahme wird das Schulsystem als ein wesentlicher Ort von Bildung in den Fokus genommen. Ausgeklammert wurden die enormen praktischen Herausforderungen in diesem System, die ganz offenkundig nicht einfach durch den Einsatz von KI-Technologien gelöst werden können, dennoch aber den Kontext der zukünftigen Gestaltung mitbestimmen. Personalmangel, sich verschärfende Lern-Defizite bei den Schülerinnen und Schülern – auch verstärkt in Folge der Corona-Pandemie – sowie verschiedene infrastrukturelle Schwächen sind gut bekannt und gleichwohl noch immer nicht adäquat gelöst. Sie bilden den Hintergrund für

[251] Weiterführende Fragen stellen sich in diesem Feld, die vor allem auf der Makroebene angesiedelt sind: Wer ist der Industriepartner oder übernimmt dies die öffentliche Hand, vor allem da es sich um kostenintensive Entwicklungen und teilweise auch Anwendungen handelt? Sind vor allem bei dieser Entwicklung auch Lernende mit besonderen Bedürfnissen im Blick?

[252] Deutscher Ethikrat (2022): Vulnerabilität und Resilienz in der Krise – Ethische Kriterien für Entscheidungen in einer Pandemie. Berlin.

jede Debatte, ob und gegebenenfalls wie sich KI-Technologien in ethisch verantwortlicher Weise in das deutsche Schulwesen integrieren lassen.

Übergreifend lässt sich diese Situation dabei sowohl für als auch gegen einen verstärkten Einsatz solcher neuen Systeme ins Feld führen: dafür, weil durch solche Technologien gegebenenfalls einige Defizite und Mängel zumindest partiell gemildert bzw. aufgefangen werden könnten. Ebenso lässt sich angesichts des teils schlechten Zustandes des Schulwesens in Deutschland auch gegen den Einsatz argumentieren, vor allem, wenn dieser dazu dient, das grundsätzliche Ziel, Bildung personell, finanziell und strukturell besser zu gestalten, zu umgehen bzw. von der Agenda zu nehmen. Entsprechend gilt es zu betonen, dass die infrastrukturellen und personellen Herausforderungen des Schulsystems weiter – und schon sehr lange – ihrer Verbesserung harren und prioritär angegangen werden sollten; ein Thema, das weit über den Fokus dieser Stellungahme hinausreicht. Zugleich sollten technologische Entwicklungen realistisch und mit Blick darauf untersucht und erwogen werden, ob und gegebenenfalls inwieweit sie auch in dieser Hinsicht Abhilfe schaffen könnten. Potenziale sollten nicht deswegen verschenkt werden, weil sie eventuell nur symptomatische, aber keine tiefgreifenden Lösungen bieten. Einfache Ersatzoptionen oder leichte Lösungen der strukturellen Herausforderungen sind wiederum von den hier diskutierten Technologien nicht zu erwarten und sollten nicht zum Nachlassen eines Reformdrucks im Schulsystem führen.

Eingedenk des hier verwendeten Verständnis des Bildungsbegriffs sowie der Analyse zur Erweiterung und zum Vermindern menschlicher Fähigkeiten durch Delegation von vormals menschlichen Tätigkeiten an Technologien, sollte im schulischen Kontext zudem unterstrichen werden, dass die grundlegende pädagogische Arbeit von den vorgestellten Systemen nicht ersetzt werden kann und sollte. Die Relevanz der Schule als Sozialraum der Interaktion zwischen Lehrenden und Lernenden sollte unbedingt erhalten bleiben. Allerdings bieten sich durchaus Potenziale, diese Interaktionen zu erweitern. Das Hauptziel beim Einsatz von KI-Systemen im schulischen Unterricht sollte immer die Förderung der Lernenden auf der einen Seite und die Unterstützung wie auch Entlastung der Lehrenden auf der anderen Seite sein.

Wie in anderen Anwendungsbereichen auch, lässt sich also nicht pauschal beantworten, welche der vorgestellten Tools – die vom Vokabeltrainer und IST-System über Classroom Analytics bis hin zum Telepräsenzroboter reichen – wann, von wem und in welchem Fach eingesetzt werden sollten. Gesellschaftliche Bildungsvorstellungen, die sich in den Institutionen und ihren Lernzielen niederschlagen, unterschiedliche Verantwortlichkeiten, aber auch die (in der Aus- und Weiterbildung erworbene) Kompetenz der Lehrkräfte, die Methoden gemäß den Lernzielen

festzulegen und die digitalen Techniken in ihrer Ambivalenz wahrzunehmen, sowie die Bedarfe aufseiten der Lernenden – all dies sind wichtige Einflussfaktoren, die auf die Notwendigkeit einer einzelfallbezogenen bzw. auf einzelne Anwendungsbereiche bezogenen Abwägung verweisen.

Da das hier angelegte Verständnis der Bildung des Menschen davon ausgeht, dass diese nicht nur in optimierbarer und berechenbarer Anhäufung von Wissen, sondern vor allem in einem konstruktiven und verantwortlichen Umgang mit erlerntem Wissen besteht, ist besondere Aufmerksamkeit geboten, die Lernprozesse, die zentral für die Persönlichkeitsbildung des Menschen sind, nicht auszulagern. Entscheidungen über das Setting des genauen Einsatzes des jeweiligen Tools sollten daher letztendlich immer beim Menschen liegen. Zudem muss eine Beeinträchtigung der Privatsphäre der Lernenden (und der Lehrenden) ausgeschlossen werden.

Empfehlungen

- *Empfehlung Bildung 1:* Digitalisierung ist kein Selbstzweck. Der Einsatz sollte nicht von technologischen Visionen, sondern von grundlegenden Vorstellungen von Bildung, die auch die Bildung der Persönlichkeit umfassen, geleitet sein. Die vorgestellten Tools sollten deshalb im Bildungsprozess kontrolliert und als ein Element innerhalb der Beziehung zwischen Lehrenden und Lernenden eingesetzt werden.
- *Empfehlung Bildung 2:* Für jedes Einsatzgebiet gilt es, eine angemessene Abwägung von Chancen und Risiken vorzunehmen. Insbesondere sollten Autonomie und Privatheit von Lehrenden und Lernenden hohen Schutz erfahren. Besondere Chancen ergeben sich im Bereich der Inklusion und Teilhabe, wo das Potenzial dieser Systeme genutzt werden sollte, um etwa sprachliche oder räumliche Barrieren abzubauen.
- *Empfehlung Bildung 3:* Tools, die einzelne Elemente des Lehr- und Lernprozesses ersetzen bzw. ergänzen (enge Ersetzung) und nachweislich Fähigkeiten, Kompetenzen oder soziale Interaktion der Personen, die sie nutzen, erweitern, wie etwa einige intelligente Tutor-Systeme oder Telepräsenz-Roboter für externe Lehrbeteiligung, sind prinzipiell weniger problematisch als solche, die umfassendere bzw. weitere Teile des Bildungsprozesses ersetzen. Je höher der Ersetzungsgrad, desto strenger müssen Einsatzbereiche, Umgebungsfaktoren und Nutzenpotenziale evaluiert werden.

- *Empfehlung Bildung 4:* Es gilt standardisierte Zertifizierungssysteme zu entwickeln253, die anhand transparenter Kriterien des Gelingens von Lernprozessen im genannten umfassenden Sinne Schulämter, Schulen und Lehrkräfte dabei unterstützen können, sich für oder gegen die Nutzung eines Produkts zu entscheiden. Hier kann sich auch der Empfehlung zur dauerhaften Einrichtung länderübergreifender Zentren für digitale Bildung, wie es im jüngsten Gutachten „Digitalisierung im Bildungssystem. Handlungsempfehlungen von der Kita bis zur Hochschule“ von der Ständigen Wissenschaftlichen Kommission der Kultusministerkonferenz angesprochen wurde254, angeschlossen werden.
- *Empfehlung Bildung 5:* Bei der Entwicklung, Erprobung und Zertifizierung entsprechender KI-Produkte bedarf es einer engen Zusammenarbeit mit den relevanten Behörden, mit den jeweils zuständigen pädagogischen Fachgesellschaften sowie der Partizipation von Beteiligten, um Schwachstellen der Produkte frühzeitig zu entdecken und hohe Qualitätsstandards zu etablieren. Bekannte Herausforderungen KI-getriebener Technologien wie beispielsweise Verzerrungen bzw. Bias oder Anthropomorphisierungstendenzen sollten bei der Entwicklung und Standardisierung berücksichtigt werden.
- *Empfehlung Bildung 6:* Um den verantwortlichen Einsatz von KI-Technologien im Bildungsprozess zu gewährleisten, muss die Nutzungskompetenz insbesondere der Lehrkräfte erhöht werden; es bedarf der Entwicklung und Etablierung entsprechender Module und Curricula in der Aus-, Fort- und Weiterbildung. Insbesondere die Gefahren eines verengten pädagogischen Ansatzes und eines Deskillings in der Lehre sollten dabei aktiv in den Blick genommen werden. Ebenso sollte die digitale Nutzungskompetenz von Lernenden sowie Eltern gestärkt und um KI-Aspekte erweitert werden.
- *Empfehlung Bildung 7:* Im Sinne der Beteiligungsgerechtigkeit sollten KI-basierte Tools Lernenden grundsätzlich auch für das Eigenstudium zur Verfügung stehen.
- *Empfehlung Bildung 8:* Die Einführung von KI-Tools im Bildungsbereich erfordert ferner den Ausbau verschiedener flankierender Forschungsbereiche. Sowohl theoretische Fundierung als auch empirische Evidenz zu Effekten, etwa auf die Kompetenzentwicklung (z. B.

---

[253] Es werden allgemeine Zertifizierungskriterien entwickelt, die dann auch für die Zertifizierung des Einsatzes im schulischen Alltag Anwendung finden könnten. Heesen, J. et al. (2020): Zertifizierung von KI-Systemen – Kompass für die Entwicklung und Anwendung vertrauenswürdiger KI-Systeme. Whitepaper aus der Plattform Lernende Systeme. München. https://www.acatech.de/publikation/zertifizierung-von-ki-systemen-kompass-fuer-die-entwicklung-und-anwendung-vertrauenswuerdiger-ki-systeme/ [07.02.2023].

[254] Ständige Wissenschaftliche Kommission der Kultusministerkonferenz (2022): Digitalisierung im Bildungssystem. Handlungsempfehlungen von der Kita bis zur Hochschule. Gutachten der Ständigen Wissenschaftlichen Kommission der Kultusministerkonferenz (SWK). Bonn (DOI: 10.25656/01:25273).

Problemlösen) oder zur Beeinflussung der Persönlichkeitsentwicklung bei Kindern und Heranwachsenden, müssen weiter ausgebaut werden. Dabei sollte nicht nur stärker in Forschung und entsprechende Produktentwicklung investiert, sondern vor allen Dingen auch die praktische Erprobung und Evaluation im schulischen Alltag verstärkt werden.

- *Empfehlung Bildung 9:* Des Weiteren stellt sich hier die Problematik der Datensouveränität. Zum einen sind bei der Sammlung, Verarbeitung und Weitergabe von bildungsbezogenen Daten strenge Anforderungen an den Schutz der Privatsphäre zu beachten. Zum anderen sollte die gemeinwohlorientierte, verantwortliche Sammlung und Nutzung von großen Daten, etwa in der prognostischen lehrunterstützenden Anwendung, ermöglicht werden.
- *Empfehlung Bildung 10:* Eine vollständige Ersetzung von Lehrkräften läuft dem hier skizzierten Verständnis von Bildung zuwider und ist auch nicht dadurch zu rechtfertigen, dass schon heute in bestimmten Bereichen ein akuter Personalmangel und eine schlechte (Aus-)Bildungssituation herrschen. In der komplexen Situation der schulischen Bildung bedarf es eines personalen Gegenübers, das mithilfe technischer Komponenten zwar immer stärker unterstützt werden kann, dadurch selbst als Verantwortungsträger für die pädagogische Begleitung und Evaluation des Bildungsprozesses aber nicht überflüssig wird.
- *Empfehlung Bildung 11:* In Anbetracht der erkenntnistheoretischen und ethischen Herausforderungen und unter Abwägung potenzieller Nutzen und Schäden stehen die Mitglieder des Deutschen Ethikrates dem Einsatz von Audio- und Videomonitoring im Klassenzimmer insgesamt skeptisch gegenüber. Insbesondere erscheint die Analyse von Aufmerksamkeit und Emotionen per Audio- und Videoüberwachung des Klassenraums mittels aktuell verfügbarer Technologien nicht vertretbar. Ein Teil des Ethikrates schließt den Einsatz von Technologien zur Aufmerksamkeits- und Affekterkennung zukünftig jedoch nicht vollständig aus, sofern sichergestellt ist, dass die erfassten Daten eine wissenschaftlich nachweisbare Verbesserung des Lernprozesses bieten und das hierfür notwendige Monitoring von Lernenden und Lehrkräften keine inakzeptablen Auswirkungen auf deren Privatsphäre und Autonomie hat. Ein anderer Teil des Ethikrates hingegen hält die Auswirkungen auf Privatsphäre, Autonomie und Gerechtigkeit hingegen generell für nicht akzeptabel und befürwortet daher ein Verbot von Technologien zu Aufmerksamkeitsmonitoring und Affekterkennung in Schulen.

7 Öffentliche Kommunikation und Meinungsbildung

7.1 Einleitung

Durch die digitale Transformation und die in der Folge veränderten politisch relevanten Kommunikationsprozesse wird die Demokratie als Herrschafts- und als Lebensform verändert.[255] Die rasante Verbreitung digitaler Plattformen und Sozialer Medien mit ihren algorithmisch vermittelten Informations- und Kommunikationsangeboten wirkt sich nicht nur auf einzelne gesellschaftliche Sphären aus, sondern potenziell auch auf große Teile der öffentlichen Kommunikation und Meinungsbildung – mit Konsequenzen für das demokratische Legitimationsgefüge, im Positiven wie im Negativen.

Über die seit der Pionierzeit der Digitalisierung bestehenden hohen Erwartungen, durch mehr Beteiligung, weitreichende Transparenz, bessere Kommunikation und schnellere Bewältigung hoheitlicher Aufgaben die Verwandlung demokratischer Strukturen zum Besseren oder gar demokratische Umbrüche erst anzustoßen oder zu ermöglichen, hat sich ein Schatten gelegt: die immer deutlicher zu Tage tretende systemische Verletzlichkeit politischer Institutionen durch Desinformations- und Manipulationsversuche und Polarisierungstendenzen. Was zu Beginn durch mehr Partizipation, erhebliche Vergrößerungen der Zugänge zu Informationen, Abbau von Hierarchien vor allem als Ausweitung und Vertiefung der Möglichkeiten eines demokratischen Gemeinwesens und seiner miteinander agierenden Bürgerinnen und Bürger angelegt zu sein schien, erweist sich in Teilen schon jetzt als gefährdend. Zunehmende Diskursverrohung und Vertrauenserosion können dazu führen, dass sich die realen Freiräume zur Partizipation und dem gemeinsamen Ringen um das bessere Argument eher verringern.

Nimmt man die Chancen wie Risiken zusammen, erscheint die Algorithmisierung politischer Kommunikationsprozesse oder gar demokratisch legitimierter Verfahren als gestaltungsbedürftig. Dabei sind die vielfach als größte Bedrohung wahrgenommenen Gefährdungen nicht die Möglichkeiten politisch motivierter Manipulation oder Angriffe auf digitale Infrastrukturen. Die Herausforderung liegt tiefer. Sie berührt das Selbstverständnis mündiger, in ihren politischen Entscheidungen grundsätzlich freier Personen. Es geht also um die Gefahr einer Verminderung menschlicher Autorschaft. Vor allem die Informations- und Diskursqualität, die für die

[255] Vgl. Nationale Akademie der Wissenschaften Leopoldina (2021): Digitalisierung und Demokratie. Stellungnahme. Halle (Saale) https://www.leopoldina.org/publikationen/detailansicht/publication/digitalisierung-und-demokratie-2021/ [03.02.2023].

je eigene Willensbildung eminent wichtig sind, werden von den Sozialen Medien herausgefordert.

7.2 Das Internet als soziotechnisches System: Funktionsweise Sozialer Medien

7.2.1 Neue soziotechnische Infrastrukturen

Die Möglichkeiten zur Informationsbeschaffung und Kommunikation im Internet haben sich in den vergangenen Jahrzehnten drastisch verändert, insbesondere durch die rasch angestiegene Verbreitung und Marktmacht interaktiver Plattformen, über die inzwischen große Teile des digitalen Informationsaustauschs ablaufen. Hierzu gehören nicht nur Soziale Netzwerke wie Facebook, Instagram, Twitter, YouTube, TikTok und LinkedIn, sondern auch Suchmaschinen, allen voran Google, und Messenger-Dienste zum Austausch von Sofortnachrichten wie WhatsApp, Signal, Telegram oder WeChat.

Die Übergänge zwischen Sozialen-Netzwerk-Plattformen, Suchmaschinen und Messenger-Diensten entwickeln sich dabei fließend, und es kommt auch bei inhaltlichen und funktionalen Merkmalen zunehmend zu Konvergenzen. Viele Plattformen bieten inzwischen einander ähnelnde Möglichkeiten dahingehend an, multimediale Inhalte zu erstellen und zu veröffentlichen oder live zu verbreiten; auf die Inhalte anderer zu reagieren, sie zu bewerten, zu kommentieren und – auch plattformübergreifend – weiterzuverbreiten; sich über direkte Nachrichten und Konferenzschaltungen mit anderen Personen auf der Plattform auszutauschen; die Plattform nach Inhalten, Profilen, Gruppen oder sonstigen Angeboten zu durchsuchen und bestimmte Inhalte zu abonnieren. Auch Optionen, eigene Inhalte gezielt zu bewerben und Produkte und Dienstleistungen direkt anzubieten oder zu kaufen, sind vielfach vorhanden.

Fast alle weiter verbreiteten Plattformen und Dienste werden von privaten Unternehmen aus den USA oder China betrieben und die größten Sozialen Netzwerke gehören nur wenigen Firmen. Von den zehn am meisten genutzten Angeboten wurden im Januar 2023 vier – Facebook, WhatsApp, Instagram und Messenger – vom US-Konzern Meta betrieben. Das zweitgrößte Soziale Netzwerk, die Videoplattform YouTube, gehört zum US-Konzern Google, der gleichzeitig mit seiner Suchmaschine und Diensten wie Google Photos, Gmail, Chrome, Maps, Android, Cloud und Drive große Teile des Internets dominiert. Drei weitere der zehn meistgenutzten

Dienste werden von den chinesischen Unternehmen Tencent (WeChat) und ByteDance (TikTok und seine chinesische Version Douyin) angeboten.[256]

Aufgrund dieser Marktmacht sowie der Vielseitigkeit und Integration der von einzelnen Konzernen bereitgehaltenen Dienste funktionieren die Angebote der großen Konzerne inzwischen als reichhaltige soziotechnische Infrastrukturen, in denen sich ein Großteil des Online-Nutzungsverhaltens nach den Vorgaben weniger Konzerne abspielt. Zahlreiche Einbindungsmöglichkeiten für externe Webseiten und Dienste verstärken diesen Effekt noch, wie beispielsweise das Teilen von Inhalten anderer Webseiten auf der Plattform, die direkte Verknüpfung externer Dienste mit einer Plattform, und die Option, sich per Einmalanmeldung (Single Sign-on) mit den Anmeldedaten großer Portale wie Google oder Facebook auch bei anderen Seiten zu authentifizieren. Wie weitumfassend einzelne Plattformen die Interneterfahrung prägen können, zeigt sich beispielsweise an Facebooks Initiative internet.org, das seit 2014 in ausgewählten Ländern über Mobilgeräte kostenlosen Zugang ins Internet anbietet, aber die Registrierung bei Facebook voraussetzt. Ebenso ist die chinesische App WeChat aufgrund ihres großen Funktionsumfangs einschließlich Bezahlfunktion für viele Menschen in China zum Zentrum fast all ihrer Onlineaktivitäten geworden.

7.2.2 Informationsauswahl und Kuratierung

Mit der Fülle der Informationen und Interaktionsmöglichkeiten in Sozialen Medien gehen technische Herausforderungen und ökonomische Potenziale einher, die gemeinsam zur Gestaltung aktueller Funktionsweisen und Geschäftsmodelle beigetragen haben. Im Gegensatz zu klassischen Medien, die Inhalte mit einer begrenzten Zahl mitarbeitender Personen nach einem bestimmten Konzept erstellen, auswählen, prüfen und veröffentlichen, produzieren in Sozialen Medien neben professionellen Akteuren auch die Nutzerinnen und Nutzer selbst die dort zirkulierenden Beiträge. So entstehen schnell riesige Mengen an Inhalten von höchst unterschiedlicher Qualität. Dies stellt die Plattformen wie auch ihre Kundschaft vor das Problem der Informationsauswahl, das sich mit traditionellen redaktionellen Strukturen, in denen Profis Inhalte aussuchen und qualitativ prüfen, nicht mehr bewältigen lässt. Aktuelle Ansätze zur Lösung dieses Problems delegieren daher einen Großteil der redaktionellen Arbeit, vor allem die Kuratierung, das heißt die Auswahl von Inhalten, an maschinelle Systeme. Insbesondere die personalisierte Kuratierung, die dazu führt, dass jeder Person beim Besuch einer Plattform auf sie

[256] World's Most Used Social Media Platforms January 2023 DataReportal https://datareportal.com/reports/digital-2023-global-overview-report [10.03.2023].

persönlich zugeschnittene Inhalte in einer bestimmten Reihenfolge angezeigt werden, erfolgt inzwischen weitgehend durch Algorithmen, die menschliche Redaktionsarbeit auf dieser Ebene somit nahezu vollständig ersetzen und unsichtbar im Hintergrund arbeiten.

Die Kriterien, nach denen solche Algorithmen ihre Auswahl treffen, sind eng mit ökonomischen Faktoren verknüpft. Die meisten Plattformen und Dienste folgen einem werbebasierten Geschäftsmodell, bei dem Werbetreibende dafür bezahlen, möglichst viele Menschen zu erreichen, die an ihren Produkten interessiert sein könnten. Dies gelingt am besten, wenn die Interessen der einzelnen Nutzerinnen und Nutzer erstens möglichst präzise bekannt sind und Menschen zweitens möglichst viel Zeit auf der Plattform verbringen. Je mehr Zeit und Aufmerksamkeit eine Person der Plattform widmet, desto besser sind die Voraussetzungen, ihr auf persönliche Interessen zugeschnittene Werbung zu präsentieren. Dies geschieht auf Basis der Annahme, dass derart maßgeschneiderte Werbung größere Erfolgsaussichten hat als Werbung, die weniger eng orientiert an der Aufmerksamkeit und den Interessen ihrer Zielgruppe präsentiert wird.

Vor diesem Hintergrund lohnt es sich für Plattformen, so viele Datenspuren wie möglich über den persönlichen Hintergrund, die Interessen, das Nutzungsverhalten und das Soziale Netzwerk der Personen, die es nutzen, zu sammeln und diese Daten dazu zu verwenden, deren Aufmerksamkeit mithilfe personalisierter Inhalte möglichst lange zu fesseln und dabei auf ebenfalls personalisierte Werbeinhalte zu lenken. Dies gelingt technisch mithilfe von Empfehlungsdiensten (*collaborative filtering recommender systems*), das heißt Systemen, die neben den eigenen Informationen und nutzungsbasierten Vorlieben jeder Person zusätzlich die entsprechenden Daten von Menschen aus deren Netzwerk oder mit ähnlichem Profil sowie die jeweiligen Interaktionen berücksichtigen (siehe Infokasten 7). Auf Grundlage dieser wechselseitigen Interessenanalyse können solche Algorithmen mit hoher und im Zuge der Nutzungsdauer ständig wachsender Präzision vorhersagen, welche Inhalte jemand angeboten bekommen sollte, um das weitere Verweilen dieser Person auf der Plattform zu maximieren. Weitere Mechanismen, die von Plattformen eingesetzt werden, um die Zeit, die Menschen auf eine Plattform verbringen, auszudehnen, sind das ständige Nachladen neuer Inhalte am Ende einer Seite (*infinite scroll*), und der Einsatz von Benachrichtigungen, die auf neue abonnierte Inhalte, Kommentare oder sonstige Reaktionen hinweisen und so dazu motivieren, die Plattform erneut aufzurufen.

Die Übergänge zwischen Werbung und regulären Inhalten und Interaktion (sogenannte organische Inhalte) sind oft fließend. Viele Firmen, Organisationen und sonstige Anbieter, die Wer-

bung auf Plattformen schalten, betreiben gleichzeitig Profile, auf denen sie auch reguläre Beiträge veröffentlichen und mit anderen im Netzwerk interagieren. Reguläre Beiträge können auch nachträglich gegen Geld eine größere Reichweite erhalten und erscheinen dann ebenfalls als Anzeige. Hinzu kommt, dass Privatpersonen, die mit erfolgreichen Inhalten auf einer Plattform eine große Zahl an Profilen, die ihre Neuigkeiten abonnieren (Follower), und somit viel Reichweite erlangt haben, zunehmend in Kooperationen mit Werbetreibenden als Influencer agieren, die Produkte und Dienstleistungen gegenüber ihrem Publikum gezielt rezensieren und bewerben.

Infokasten 7: Facebook-Algorithmus

Die Auswahlkriterien von Empfehlungsalgorithmen auf sozialen Plattformen werden in der Regel nicht vollumfänglich veröffentlicht, wohl jedoch bestimmte Akzentuierungen und Änderungen, wie folgendes Beispiel der Facebook-Software[257] zeigt.

Nachdem die Plattform 2006 in der Ursprungsversion ihres Nachrichtentickers (Newsfeed, mittlerweile nur noch Feed) nur Inhalte aus unmittelbar vernetzten Profilen in chronologischer Reihenfolge angezeigt hatte, wurde 2007 mit der Einführung des Like Button erstmals die Möglichkeit angeboten, Inhalte zu bewerten. Diese Interaktionen wurden ab 2009 dafür genutzt, Beiträge mit vielen Likes im Newsfeed nach oben zu spülen. Im Laufe der folgenden Jahre kamen diverse weitere Kriterien dazu, darunter:

- höhere Priorität für Beiträge aus Profilen, mit dem man viel interagiert (2013)
- weniger Reichweite für Profile, die in ihren nicht als Anzeigen geschalteten Beiträgen Werbung verbreiten (2015)
- weniger Reichweite für Beiträge, die von vielen als unwahr markiert werden (2015)
- höhere Priorität für Beiträge, die Ähnlichkeit mit Inhalten haben, mit denen jemand zuvor viel Zeit verbracht hat (2016)
- höhere Priorität für Beiträge, die Ähnlichkeit mit Inhalten haben, auf die jemand nicht mit dem klassischen „Gefällt mir“ reagiert hat, sondern mit einem der neueren Emojis „Liebe“, „Umarmung“, „Haha“, „Wow“, „Traurig“ und „Wütend“ (2017)
- mehr Reichweite für Beiträge, die besonders häufig geteilt werden und viele Reaktionen und Diskussionen in den Kommentaren auslösen (2018)

Laut Angaben von Facebook[258] sind die „drei hauptsächlichen Signale“, die der Algorithmus aktuell bei der personalisierten Priorisierung von Inhalten verwendet:

(1) wie häufig eine Person mit der Quelle, aus der ein Beitrag stammt, zuvor interagiert hat
(2) welche Art von Inhalten eine Person zuvor am häufigsten genutzt hat (z. B. Fotos, Videos oder Links)
(3) wie viele Interkationen ein Beitrag erhalten hat, insbesondere von anderen im persönlichen Netzwerk

[257] Newberry, C. (2022): How the Facebook Algorithm Works in 2023 and How to Make it Work for You. In: Hootsuite. https://blog.hootsuite.com/facebook-algorithm/ [12.01.2023].
[258] Meta: So funktioniert der Feed. https://www.facebook.com/formedia/tools/feed [12.01.2023].

Zusätzlich zu diesen algorithmisch vorgenommenen Gewichtungen können Nutzerinnen und Nutzer einige Kriterien auch selbst festlegen, wie beispielsweise eine Favoritenliste der Profile und Seiten, denen sie folgen und deren Inhalte sie prioritär sehen möchten.

Insgesamt kommt es so auf Plattformen und in Sozialen Netzwerken zu einer ausgeprägten Verquickung technisch wie menschlich vermittelter Stimuli und Strukturen, die große soziale und psychologische Wirkmacht entfalten und gemeinsam dazu beitragen, Aufmerksamkeit zu erregen und zu fesseln, die Verweildauer von Menschen auf der Plattform zu maximieren und so letztlich die Rezeption von Werbehinhalten zu optimieren.

7.2.3 Moderation von Inhalten

Eine algorithmisch gesteuerte personalisierte Informationsauswahl, in der ökonomische und aufmerksamkeitsbasierte Faktoren derart eng verbunden sind und die sich anhand des Nutzungsverhaltens ständig weiterentwickelt, führt dazu, dass Inhalte, die besonders sensationell erscheinen oder intensive emotionale Reaktionen auslösen, sich überproportional schnell und weit verbreiten. Dies begünstigt unter anderem Falschnachrichten[259] und Beiträge, auf die mit Wut reagiert wird[260], so zum Beispiel Inhalte wie Hassreden, Beleidigungen und Volksverhetzungen. Das stellt Plattformen vor die Herausforderung, wie solche potenziell problematischen und gleichzeitig verbreitungsstarken Inhalte erkannt werden können und wie mit ihnen umzugehen ist.

In Reaktion auf diese Herausforderung bemühen sich Plattformen darum, ihre Inhalte nach verschiedenen Kriterien zu moderieren (Content Moderation). Hierbei sind einerseits Menschen beteiligt – sowohl Angestellte der Firmen selbst als auch Personen, die für externe Dienstleister arbeiten und die Menschen, die das Plattformangebot nutzen und Inhalte melden, die ihnen unangenehm auffallen. Andererseits kommen auch algorithmische Systeme zum Einsatz. In diesem soziotechnischen Zusammenspiel geht es zum einen um die Gestaltung und Umsetzung von Regeln, die bestimmen, welche Inhalte erlaubt sind und welche gelöscht werden müssen, zum anderen aber auch darum, welche Inhalte im Rahmen einer algorithmischen Kuratierung gegebenenfalls lediglich in ihrer Verbreitung und Sichtbarkeit eingeschränkt werden.

---

[259] Vosoughi, S. et al. (2018): The spread of true and false news online. In: Science 359 (6380), 1146-1151 (DOI: 10.1126/science.aap9559).
[260] Fan, R. et al. (2014): Anger Is More Influential Than Joy: Sentiment Correlation in Weibo. In: PLoS ONE 9(10): e110184 (DOI: 10.1371/journal.pone.0110184).

Grundlage für die Moderation sind zunächst rechtliche Vorgaben, die bestimmte Inhalte als unzulässig einstufen, vor allem solche des Strafrechts, aber auch zivilrechtlich geschützte Persönlichkeitsrechte setzen der Kommunikation Grenzen. Auch die Anforderungen für die Meldung, Prüfung und Entfernung solcher Inhalte wie auch für Beschwerden gegen solche Eingriffe ergeben sich zunächst aus den allgemeinen rechtlichen Regelungen für rechtswidrige Äußerungen. Im deutschen Kontext ist darüber hinaus insbesondere auf das Gesetz zur Verbesserung der Rechtsdurchsetzung in Sozialen Netzwerken (Netzwerkdurchsetzungsgesetz) hinzuweisen, welches Plattformbetreibende verpflichtet, offensichtlich illegale Inhalte innerhalb von 24 Stunden zu löschen. Das Gesetz selbst definiert nicht, was illegal ist, sondern verweist auf strafrechtliche Normen. An das Konzept des Netzwerkdurchsetzungsgesetzes knüpft der 2022 von der Europäischen Union erlassene Digital Services Act an, der das Netzwerkdurchsetzungsgesetz wohl jedenfalls teilweise ersetzen wird.[261]

Hinzu kommen in der Regel Kommunikationsregeln, die von Plattformbetreibern selbst entwickelt werden (Community Standards), auf deren Grundlage auch rechtlich zulässige Inhalte gelöscht, gesperrt oder in ihrer Reichweite eingeschränkt werden können. Hierzu gibt es eine kontroverse Diskussion, ob und wann solche Praktiken die Meinungsfreiheit auf unzulässige Weise beschränken.[262]

Moderationsvorgänge können sowohl durch die Meldung anstößiger die Einzelpersonen bei der Nutzung des Plattformangebots auffallen, ausgelöst werden als auch durch Algorithmen, die beispielsweise nach bestimmten Schlüsselwörtern suchen. Bei Inhalten, die als potenziell problematisch markiert werden, muss anschließend geprüft werden, ob sie die aktuell geltenden Regeln verletzen. Auch an diesen Prüfaufgaben sind Menschen und Maschinen gemeinsam beteiligt. Menschliche Moderation erfolgt typischerweise durch Personen, die häufig unter äußerst prekären Arbeitsbedingungen bei Drittanbietern angestellt sind, mit denen eine Plattform vertraglich zusammenarbeitet.[263] Sie sichten gemeldete Inhalte und gleichen sie mit den jeweils geltenden Regeln ab. Dabei werden sie häufig mit extrem belastendem Material wie Tötungen, Kindesmissbrauch, Tierquälerei und Suizid konfrontiert. Zudem müssen sie innerhalb weniger

[261] Vgl. Weiden, H. (2022): Mehr Freiheit und Sicherheit im Netz. Gutachten zum Entwurf des Digital Services Act im Auftrag der Friedrich-Naumann-Stiftung Für die Freiheit. Potsdam-Babelsberg. https://shop.freiheit.org/#!/Publikation/1201 [08.02.2023].
[262] Raue, B. (2018): Meinungsfreiheit in sozialen Netzwerken. In: Juristenzeitung 73, 961–970.
[263] Siehe dazu etwa den Dokumentarfilm „The Cleaners“ aus dem Jahr 2018 (www.thecleaners-film.de) und Perrigo, B. (2023): Meta Fails Attempt to Dodge a Worker Exploitation Lawsuit in Kenya. In: Time. https://time.com/6253180/meta-kenya-lawsuit-motaung/ [28.02.2023]. Die gleiche Problematik zeigt sich auch bei ChatGTP; vgl. Perrigo, B. (2023): Exclusive: OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic. In: TIME. https://time.com/6247678/openai-chatgpt-kenya-workers/ [06.02.2023].

Sekunden sprachlich und kulturell komplexe Nuancen berücksichtigen, von denen die Zulässigkeit eines Beitrags entscheidend abhängen kann. So kann es zum Beispiel einen großen Unterschied machen, ob ein Beitrag als Satire gewertet wird oder nicht oder ob eine anstößige Botschaft im Kontext des Beitrags nur zitiert und gegebenenfalls kritisch reflektiert oder sich zu eigen gemacht wird.

Der Einsatz von Algorithmen zur Moderation wirkt vielversprechend, da er zum einen anstößige, beispielsweise brutale Inhalte herausfiltern kann, ohne dass diese von Menschen angesehen werden müssen, und zum anderen mit der unübersichtlichen Menge an Daten und Inhalten im Netz besser umgehen kann. Allerdings sind automatisierte Methoden jedenfalls bislang häufig unzureichend, um den kulturellen und sozialen Zusammenhang einer Äußerung einzubeziehen und diese damit adäquat zu beurteilen, da aufgrund der beschriebenen komplexen Kontexte und Nuancen sowie durch absichtlich verwendete Stilmittel wie Ironie und Sarkasmus Unschärfen entstehen, die sich algorithmisch noch nicht immer zuverlässig zuordnen lassen[264]. Genau solche kontextuellen Faktoren spielen bei der juristischen sowie moralischen Beurteilung von Äußerungen jedoch eine sehr große Rolle.

So kommt es zum Beispiel darauf an, wer gegenüber wem in welchem Kontext einen Begriff verwendet, um zu beurteilen, ob es sich hier um eine (rassistische) Beleidigung handelt. Im amerikanischen Raum konnte etwa gezeigt werden, dass der Einsatz automatisierter Content-Moderation-Techniken verschiedene gesellschaftliche Gruppen unterschiedlich betrifft, da bestimmte Sprachformen beispielsweise eher markiert werden. Facebooks Ansatz, allgemeingültige und „blinde“ Regeln zu erschaffen, führte hier dazu, dass Gruppen wie *white men* eher vor Hassrede geschützt wurden als vulnerablere Gruppen wie *black children*.[265]

Eine Sorge bei der Content Moderation liegt in der Gefahr, dass systematisch auch Inhalte gelöscht oder unzugänglich gemacht werden, die nicht gegen (vom Staat oder von Plattformen gesetzte) Regeln verstoßen (sogenanntes *Overblocking*). Dass es dazu kommt, ist jedenfalls dann plausibel, wenn der Rechtsrahmen Anreize setzt, Inhalte im Zweifel zu löschen, etwa

---

[264] Algorithmische Methoden zur *Sentiment Analysis*, die auch Stilmittel wie z. B. Ironie und Sarkasmus erkennen können, werden ständig weiterentwickelt, vgl. etwa Sarsam, S. M. et al. (2020): Sarcasm detection using machine learning algorithms in Twitter: A systematic review. In: *International Journal of Market Research 62*, 578-598 (DOI: 10.1177/1470785320921779).

[265] Angwin, J.; Grassegger, H. (2017): Facebook’s Secret Censorship Rules Protect White Men From Hate Speech But Not Black Children. In: ProPublica. https://www.propublica.org/article/facebook-hate-speech-censorship-internal-documents-algorithms [12.01.2023]; Davidson, T. et al. (2019): Racial Bias in Hate Speech and Abusive Language Detection Datasets. In: Proceedings of the Third Workshop on Abusive Language Online. Florenz, 25–35 (DOI: 10.18653/v1/W19-3504).

wenn Sanktionen für den Fall der Nicht-Löschung vorgesehen sind, nicht aber für die fehlerhafte Löschung. Dies wird beispielsweise in kritischen Betrachtungen des Netzwerkdurchsetzungsgesetzes von Kritikern angenommen.[266] Ob es tatsächlich zu Overblocking kommt, ist schwer empirisch zu prüfen, nicht nur weil dazu Datenzugang benötigt wird, sondern auch, weil eine angemessene Prüfung von Kommunikationsinhalten eine Auslegung der Äußerung und eine Berücksichtigung des Kontextes erfordert, was jedenfalls den Aufwand erhöht.

7.2.4 Auswirkungen auf die Erweiterung und Verminderung menschlicher Handlungsfähigkeiten

Die beschriebenen Funktionsweisen von Plattformen und Sozialen Medien sind dadurch gekennzeichnet, dass erstens große Teile der Auswahl und Moderation von Inhalten an Algorithmen delegiert werden, dass es dabei zweitens zu komplexen Wechselwirkungen zwischen diesen technisch vermittelten Prozessen und den in diesen Netzwerken agierenden Menschen kommt und dass hier drittens ganz überwiegend kommerziell motivierte Gestaltungsparameter eingesetzt werden, die dazu dienen, die Aufmerksamkeit und Verweildauer von Menschen auf der jeweiligen Plattform zu maximieren. Mithilfe dieser soziotechnischen Verquickungen können menschliche Handlungsfähigkeiten in unterschiedlicher Weise erweitert oder vermindert werden.

Die Delegation von Kuratierungsprozessen an Algorithmen, die aufgrund datenreicher Profile ein überaus effektives personalisiertes Angebot von Inhalten zusammenstellen, erweitert zunächst die Möglichkeiten von Plattformbetreibern und Werbetreibenden ein großes Publikum erfolgreich zu erreichen. Für einzelne Nutzerinnen und Nutzer ist die Personalisierung des Angebots mit Komfort- und Effizienzgewinnen verbunden, weil gewünschte bzw. zu den eigenen Interessen passende Inhalte schnell auffindbar sind oder proaktiv angeboten werden. Auch dies kann eine Erweiterung von Handlungsmöglichkeiten bedeuten, zum Beispiel wenn persönliche Ziele so besser oder schneller erreicht werden können oder aufgrund der effektiven Delegation der Inhaltsauswahl an Algorithmen Entlastungseffekte auftreten, die Freiräume für andere Aktivitäten schaffen.

Die soziotechnisch hybride Moderation von Inhalten erweitert zudem die Möglichkeiten aller Beteiligten, eine sehr große Inhaltsfülle und -vielfalt verbreiten und rezipieren zu können, ohne

---

[266] Überblick bei Liesching, M. et al. (2021): Das Netzwerkdurchsetzungsgesetz in der praktischen Anwendung – Eine Teilevaluation des Netzwerkdurchsetzungsgesetzes. Berlin. Vgl. Fichtner, L. (2022): Content Moderation and the Quest for Democratic Legitimacy. In: Weizenbaum Journal of the Digital Society 2(2), 1-29 (DOI: 10.34669/wi.wjds/2.2.2).

dabei ein Übermaß problematischen Materials in Kauf nehmen zu müssen. Die beschriebenen, hierbei bestehenden Unzulänglichkeiten und Abwägungsschwierigkeiten können jedoch, wie in Abschnitt 7.4 noch näher beleuchtet wird, auch vermindernd auf die Diskursqualität wirken.

Eine Verminderung menschlicher Handlungsspielräume kann sich ebenso ergeben, wo es Menschen schwerfällt, sich dem Sog von Plattformangeboten zu entziehen und die Nutzung dieser Angebote auf ein für sie gesundes Maß zu beschränken. Die soziotechnische Optimierung von Plattformangeboten zur Maximierung von Aufmerksamkeit und Verweildauer entwickelt hier große Wirkmacht; sogar regelrechte Abhängigkeiten können entstehen.[267] Persönliche Freiheit wird z. B dann vermindert, wenn das selbstbestimmte Verfolgen eigentlich gesetzter Ziele erschwert wird, weil eine Loslösung vom Plattformangebot nicht rechtzeitig hinreichend gelingt.

Auch die Delegation der Inhaltsauswahl an eine algorithmische Kuratierung kann Autorschaft vermindern, insbesondere wenn eine rationale Auseinandersetzung mit Alternativen aufgrund der algorithmischen Vorwegnahme bestimmter Relevanzentscheidungen nur noch eingeschränkt stattfinden kann. Die auf Grundlage solcher Vorwegnahmen entstehenden personalisierten und soziotechnisch komplexen Strukturen, innerhalb derer Inhalte zur Auswahl angeboten werden (*choice architectures*) können ausgeprägte Nudging-Effekte entfalten, die menschliche Entscheidungen wirkmächtig beeinflussen, ohne dass dies für die Betroffenen erkenntlich ist (vgl. Abschnitt 4.5). Dies ist als besonders problematisch einzustufen, wenn die algorithmischen Auswahlkriterien kommerziell motiviert oder von anderen Interessen privater Konzerne bestimmt werden, die von außen kaum nachvollziehbar sind und deren Ausrichtung am Gemeinwohl nicht vorausgesetzt werden kann.[268]

Neben diesen allgemeinen Auswirkungen der Funktionsweisen von Plattformen und Sozialen Medien verändern sich auch die Informations- und Diskursqualität, die wichtige Grundlagen der öffentlichen Meinungsbildung sind – mit potenziell weitreichenden Konsequenzen für Prozesse der politischen Willensbildung. Hier steht eine Reihe von Phänomenen in der Diskussion, die in den nächsten beiden Abschnitten näher betrachtet werden. Wie weit verbreitet und wirkmächtig diese Effekte sind, lässt sich aktuell zwar noch nicht abschließend beurteilen, auch weil die Datenlage mitunter unklar oder widersprüchlich ist. Manche Fragen lassen sich nur schwer

[267] Cheng, C. et al.(2021): Prevalence of social media addiction across 32 nations: Meta-analysis with subgroup analysis of classification schemes and cultural values. In: Addictive Behaviors 117: 106845 (DOI: 10.1016/j.addbeh.2021.106845).
[268] Hildebrandt, M. (2022): The Issue of Proxies and Choice Architectures. Why EU Law Matters for Recommender Systems. In: Frontiers in Artificial Intelligence 5, Artikel 789076. (DOI: 10.3389/frai.2022.789076); Yeung, K. (2016): 'Hypernudge': Big Data as a mode of regulation by design. In: Information, Communication & Society, 20 (1), 118-136 (DOI: 10.1080/1369118X.2016.1186713).

untersuchen, da zum Beispiel Daten zur Plattformnutzung und den dort verwendeten Algorithmen auch für die Forschung nur teilweise oder in einigen Fällen, wie beispielsweise bei verschlüsselten Messenger-Diensten, gar nicht öffentlich zugänglich sind. Ein genauerer Blick auf die postulierten Mechanismen lohnt jedoch schon deswegen, weil die von ihnen berührten Prozesse grundlegend für unsere Demokratie sind.

7.3 Informationsqualität

Mit Blick auf die Informationsqualität ist zunächst auf die vielen positiven Veränderungen zu verweisen, die sich trotz der oben beschriebenen Herausforderungen bei der Auswahl, Kuratierung und Moderation von Plattforminhalten aus dem im Internet verfügbaren Informationsangebot ergeben. Plattformen und Soziale Medien erweitern die Möglichkeiten vieler Menschen, sich zu informieren, sich Gehör zu verschaffen und sich mit anderen auszutauschen und zu vernetzen, ganz erheblich.[269] Zweifelsohne eröffnet die Quellenvielfalt und -fülle auf vielen Plattformen wie auch im Internet insgesamt im Vergleich mit klassischen Informationskanälen wie Massenmedien, Bibliotheken oder dem direkten persönlichen Austausch wesentlich umfangreichere Möglichkeiten, sich schnell und einfach zu informieren. Dies gilt erst recht für Menschen, die offline keinen guten Zugang zu Informationsangeboten haben, wie zum Beispiel in strukturschwachen oder sich erst entwickelnden Regionen, oder für Menschen, die in ihrer Mobilität eingeschränkt sind.

Es ist wie noch nie zuvor möglich, dass Menschen Informationen und Einschätzungen von unterschiedlichsten Seiten und in hoher Detailgenauigkeit einholen können. Dazu tragen zum einen Projekte bei wie die gemeinnützige, umfassende und ständig wachsende Enzyklopädie des Menschheitswissens Wikipedia. Auch wissenschaftliche Erkenntnisse werden durch Open-Access-Modelle leichter zugänglich. Des Weiteren können sich an speziellen Themen Interessierte auf Plattformen, in Gruppen oder Foren oder über die Verwendung relevanter Markierungen von Beiträgen beispielsweise mit Hashtags zielgerichtet austauschen.[270] Es steht online jeder Person frei, die eigene Informationsnutzung eigenständig zu diversifizieren und beispielsweise

---

[269] Ortiz, J. et al. (2019): Giving Voice to the Voiceless: The Use of Digital Technologies by Marginalized Groups. In: Communications of the Association for Information Systems, 45 (2), 20-38 (DOI: 10.17705/1CAIS.04502); Donelan, H. (2016): Social media for professional development and networking opportunities in academia. In: Journal of Further and Higher Education, 40 (5), 706-729 (DOI: 10.1080/0309877X.2015.1014321).

[270] Imani (2022): With Twitter Crumbling, It Feels Like The World Is Collapsing On Disabled People. In: Crutches and Spice. https://crutchesandspice.com/2022/11/16/with-twitter-crumbling-it-feels-like-the-world-is-collapsing-on-disabled-people/ [12.01.2023].

gezielt nach Quellen und Stimmen zu suchen, die den eigenen Horizont erweitern. Gemeinsam eröffnen diese Optionen erhebliche Chancen, menschliche Handlungsspielräume, Teilhabemöglichkeiten und Räume für den gemeinsamen öffentlichen Vernunftgebrauch zu erweitern.[271]

Demgegenüber wird vielfach die Sorge geäußert, dass sich insbesondere aus den derzeit gängigen Kriterien der algorithmischen Kuratierung mit ihrer Priorisierung fesselnder Inhalte auch negative Auswirkungen auf die Informationsqualität ergeben. In der Diskussion stehen hier vor allem die Verbreitung von Falschnachrichten, die Entstehung von Filterblasen und Echokammern und ein Trend zu Inhalten, die negative emotionale und moralische Reaktionen und Interaktionen provozieren und so möglicherweise zu den in Abschnitt 7.4 näher beleuchteten gesellschaftlichen Polarisierungs- oder Fragmentierungstendenzen beitragen.

7.3.1 Falschnachrichten und Verschwörungstheorien[272]

Die bereits erwähnte Tendenz, dass sich unter den aktuellen Kuratierungskriterien vieler Plattformen sensationell wirkende Falschinformationen, insbesondere bei politischen Themen[273], besonders schnell verbreiten, ist mit der Sorge verbunden, dass dies selbst dann nachhaltige Effekte auf die Informationsqualität haben kann, wenn solche Inhalte im Rahmen von Moderationsbemühungen erkannt, entfernt, in ihrer Verbreitung gehemmt oder mit einordnenden Hinweisen versehen werden.

In diesem Zusammenhang wird häufig auch die Annahme vorgebracht, dass populistische Strömungen, Verschwörungstheorien und weitere politisch motivierte Bewegungen, bei denen die Erzeugung und Verstärkung bestimmter Stimmungslagen in der Bevölkerung eine wichtige Rolle spielt, in den vergangenen Jahren zugenommen bzw. an Zulauf und/oder Bedeutung gewonnen haben und dass Falschnachrichten in diesem Rahmen gezielt für politische Zwecke

[271] Waterson, J. et al. (2022): The most effective press watchdog. In: The Guardian. https://www.theguardian.com/technology/2022/nov/30/twitter-journalism-elon-musk [12.01.2023].
[272] Der Begriff der Verschwörungstheorie hat sich eingebürgert, obwohl der Ausdruck „Theorie“ angesichts der Fragwürdigkeit bzw. Widersprüchlichkeit der postulierten Inhalte oftmals nicht angemessen ist.
[273] Vosoughi, S. et al. (2018): The spread of true and false news online. In: Science 359 (6380), 1146-1151 (DOI: 10.1126/science.aap9559).

eingesetzt werden.[274] Dieser Eindruck wird unter anderem dadurch begünstigt, dass populistische und verschwörerische Elemente bei vielen Ereignissen und Themen Hand in Hand gehen[275] und beide in jüngerer Zeit eine große Rolle im öffentlichen Diskurs haben.

Es ist nicht von der Hand zu weisen, dass populistische und von verschwörungstheoretischen Ideen geprägte Inhalte demokratiegefährdend wirken können, da sie häufig ausdrücklich darauf abzielen, Vertrauen in demokratisch legitimierte Prozesse und Institutionen zu untergraben, schuldige Personen ausfindig zu machen, die Welt in Gut und Böse aufzuteilen und vermeintliche Gegenseiten zu identifizieren und zu diskreditieren (vgl. Abschnitt 7.4.1).[276] So gibt es beispielsweise empirische Hinweise darauf, dass Menschen, die mit Verschwörungstheorien konfrontiert werden, danach weniger Vertrauen in etablierte Fakten haben und auch weniger Hilfsbereitschaft zeigen.[277]

Ob der Glaube an Verschwörungstheorien in den letzten Jahren tatsächlich zugenommen hat, ist umstritten, auch weil der Definitionsgegenstand oft schwer zu fassen ist und eindeutige empirische Belege fehlen.[278] Die These, dass die in Abschnitt 7.2 beschriebenen Selektionskriterien auf Plattformen auch zur einfachen und schnellen Verbreitung von Verschwörungstheorien und populistischen Inhalten beitragen, erscheint jedoch zumindest plausibel. Falschnachrichten verbreiten sich beispielsweise unter anderem dann besonders erfolgreich, wenn sie Empörung und Wut schüren.[279] Es lässt sich zudem nachweisen, dass Menschen, die Soziale Medien intensiv als Informationsquelle nutzen, empfänglicher für konspiratives Gedankengut sind.[280] Zudem lässt die wiederholte Konfrontation mit Falschnachrichten jene bereits als glaubwürdig

---

[274] Muirhead, R. et al. (2019): A Lot of People Are Saying: The New Conspiracism and the Assault on Democracy. Princeton.
[275] Castanho S. et al. (2017): The Elite Is Up to Something: Exploring the Relation Between Populism and Belief in Conspiracy Theories. In: Swiss Political Science Review, 423-443 (DOI: 10.1111/spsr.12270).
[276] Berg, K. (2022): Eine Gefahr für Demokratien. https://www.deutschland.de/de/topic/politik/verschwoerungstheorien-und-demokratie-experte-michael-butter [12.01.2023].
[277] Van der Linden, S. (2015): The conspiracy-effect: Exposure to conspiracy theories (about global warming) decreases pro-social behavior and science acceptance. In: Personality and Individual Differences 87, 171-173 (DOI: 10.1016/j.paid.2015.07.045).
[278] Uscinski, J. et al. (2022): Have beliefs in conspiracy theories increased over time? In: PLoS ONE 17 (7): e0270429 (DOI: 10.1371/journal.pone.0270429). Vgl. hierzu auch Expertenstatements im Dossier des Science Media Center Germany unter https://www.sciencemediacenter.de/alle-angebote/research-in-context/details/news/glaube-an-verschwoerungstheorien-nahm-im-laufe-der-zeit-nicht-zu [18.01.2023].
[279] Chuai, Y.; Zhao, J. (2022): Anger can make fake news viral online. In: Frontiers in Physics, 10:970174 (DOI: 10.3389/fphy.2022.970174).
[280] Ender, A. M. et al. (2021): The Relationship Between Social Media Use and Beliefs in Conspiracy Theories and Misinformation. In: Political Behavior (DOI: 10.1007/s11109-021-09734-6).

erscheinen[281] und besonders einprägsam wirken[282], selbst wenn sie später diskreditiert oder als umstritten markiert werden. Dies gilt erst recht, wenn Inhalte innerhalb eines persönlichen Netzwerks verbreitet werden, dessen Quellen aufgrund gemeinsamer Interessen und Überzeugungen von vornherein als besonders vertrauenswürdig gelten (vgl. Abschnitt 7.3.2).[283]

Moderationsversuche können bei der Eindämmung von Falschnachrichten oft kaum mithalten. Inhalte, die aus vertrauten oder als wirkmächtig wahrgenommenen Quellen (z. B. von Influencern) stammen oder bereits weit verbreitet wurden und daher als besonders beliebt erscheinen, werden unabhängig vom Wahrheitsgehalt vorzugsweise konsumiert[284] und geteilt.[285] Die gezielte und massenhafte Verbreitung von Falschinformationen über eigens hierfür eingerichtete gefälschte und teilweise automatisch betriebene Konten (Fake-Accounts, Bots[286]) sorgt derweil dafür, dass fragwürdige Inhalte schnell kritische Verbreitungsschwellen erreichen. Beispielsweise wurden einer Schätzung zufolge 2018 auf Twitter zwei Drittel aller Verweise auf politische Seiten durch Bots verbreitet.[287]

7.3.2 Filterblasen und Echokammern

Filterblasen und Echokammern sind Metaphern, die auf der Hypothese gründen, dass der starke Fokus der algorithmischen Kuratierung auf individuelle Präferenzen dazu führen kann, dass die individuell wahrnehmbare Informationsauswahl zunehmend homogener wird. Das Phänomen der Filterblase beschreibt eine potenzielle Verarmung des Themen- und Meinungsspektrums im personalisierten Angebot.[288] Dahinter steht die Vermutung, dass Empfehlungssoftware, sobald sie persönliche Interessen und Vorlieben erkannt hat, vornehmlich Inhalte präsentiert, die

---

[281] Pennycook., G. et al. (2017): Prior exposure increases perceived accuracy of fake news. In: Journal of Experimental Psychology General 147 (12), 1865-1880 (DOI: 10.1037/xge0000465).
[282] Murphy, G. et al. (2019): False Memories for Fake News During Ireland's Abortion Referendum. In: Psychological Science 30, 1449–1459 (DOI: 10.1177/0956797619864887).
[283] Del Vicario, M. et al. (2016): The spreading of disinformation online. In: Proceedings of the National Academy of Sciences 113 (3) , 554-559 (DOI: 10.1073/pnas.1517441113).
[284] Vgl. z. B. für Jugendliche in den USA: Robb, M. B. (2020): Teens and the News: The Influencers, Celebrities, and Platforms They Say Matter Most. San Francisco. https://www.commonsensemedia.org/sites/default/files/research/report/2020_teensandnews-fullreport_final-release-web.pdf [12.01.2023].
[285] Allem, J. P. (2020): Social media fuels wave of coronavirus misinformation as users focus on popularity, not accuracy. In: The Conversation. https://theconversation.com/social-media-fuels-wave-of-coronavirus-misinformation-as-users-focus-on-popularity-not-accuracy-135179 [12.01.2023].
[286] Unter einem Bot (von englisch robot ‚Roboter') versteht man ein Computerprogramm, das weitgehend automatisch sich wiederholende Aufgaben abarbeitet, ohne dabei auf eine Interaktion mit einem menschlichen Benutzer angewiesen zu sein.
[287] Wojcik, S. et al. (2018): Bots in the Twittersphere. In: Pew Research Center. https://www.pewresearch.org/internet/2018/04/09/bots-in-the-twittersphere/ [12.01.2023].
[288] Pariser, E. (2011): The Filter Bubble: What the Internet Is Hiding from You. New York.

dem Material ähneln, auf das man zuvor bereits mit erhöhter Aufmerksamkeit oder Interaktionen reagiert hat, während andere Inhalte, die weniger offensichtlich in das ermittelte Profil passen, immer stärker ausgeschlossen werden. Eine Untersuchung auf Facebook zeigte beispielsweise, dass von persönlichen Kontakten geteilte Nachrichten, die von der politischen Position einer Person abweichen, nachrangig platziert wurden, sodass sie seltener prominent im Newsfeed erschienen und Inhalte, die von den im persönlichen Netzwerk durchaus vorhandenen Menschen mit abweichender politischer Auffassung somit benachteiligt dargestellt wurden.[289] In einer Untersuchung von Google-Suchergebnissen zu Parteien im Vorfeld der Bundestagswahl 2017 konnten hingegen kaum Personalisierungseffekte bei den Ergebnissen festgestellt werden.[290]

Eng mit der Metapher der Filterblase verbunden ist das Konzept der Echokammer, nach dem die algorithmisch geförderte Homogenisierung des personalisierten Angebots dazu führt, dass Überzeugungen und Meinungen zu kontroversen Themen durch ständige Wiederholungen und gegenseitige Bestätigung innerhalb eines Netzwerks Gleichgesinnter verstärkt und verfestigt werden. Dieser Effekt konnte zum Beispiel in einer Studie auf den algorithmisch kuratierten Plattformen Twitter (zum Thema Schwangerschaftsabbrüche) und Facebook (zum Thema Impfen und beim Nachrichtenkonsum) nachgewiesen werden, nicht jedoch beim Nachrichtenkonsum auf der Plattform Reddit, auf der die Reihenfolge von Inhalten nicht algorithmisch personalisiert wird, sondern davon abhängt, wie viele Stimmen ein Beitrag von anderen Teilnehmenden einer Diskussion erhalten hat.[291]

Das Ausmaß und die Wirkung von Filterblasen und Echokammern sind umstritten und dürften sich zwischen verschiedenen Anwendungen und Regionen stark unterscheiden.[292] Es ist also Vorsicht geboten bei zu schnellen Verallgemeinerungen. Ob und in welchem Ausmaß Filter-

---

[289] Bakshy, E. et al. (2015): Exposure to ideologically diverse news and opinion on Facebook. In: Science 348 (6239), 1130-1132. (DOI: 10.1126/science.aaa1160).
[290] Krafft, T. et al. (2018): Wer sieht was? Personalisierung, Regionalisierung und die Frage nach der Filterblase in Googles Suchmaschine. Kaiserslautern. https://algorithmwatch.org/de/filterblase-geplatzt-kaum-raum-fuer-personalisierung-bei-google-suchen-zur-bundestagswahl-2017/ [12.01.2023].
[291] Cinelli, M. et al. (2021): The echo chamber effect on social media. In: Proceedings of the National Academy of Sciences 118 (9): e2023301118 (DOI: 10.1073/pnas.2023301118).
[292] Stark, B. et al. (2021): Maßlos überschätzt. Ein Überblick über theoretische Annahmen und empirische Befunde zu Filterblasen und Echokammern. In: Eisenegger, M. et al. (2021): Digitaler Strukturwandel der Öffentlichkeit. Wiesbaden, 302-321 (DOI: 10.1007/978-3-658-32133-8_17); Dubois E.; Blank G. (2018): The echo chamber is overstated: The moderating effect of political interest and diverse media. In: Information, Communication & Society 21 (5), 729–745. (DOI: 10.1080/1369118X.2018.1428656); Rau, J. P.; Stier, S. (2019): Die Echokammer-Hypothese: Fragmentierung der Öffentlichkeit und politische Polarisierung durch digitale Medien?. In: Zeitschrift für vergleichende Politikwissenschaft 13, 399-417 (DOI: 10.1007/s12286-019-00429-1).

blasen oder Echokammern entstehen oder verstärkt werden, hängt von der verwendeten Technologie, dem Kontext und den Nutzungsweisen ab. Ebenso trägt vermutlich die Wahl der jeweiligen methodischen Ansätze bei der Untersuchung solcher Phänomene dazu bei, dass es zu unterschiedlichen Einschätzungen kommt. Auch der bereits erwähnte mangelnde Zugang zu Plattformdaten für die Forschung erschwert solche wissenschaftlichen Analysen zusätzlich. Allerdings sind die potenziellen Gefahren solcher informationellen Defizite derart essenziell für demokratische Meinungs- und Willensbildungsprozesse, dass ihnen auch bereits bei unklarer Datenlage entschieden entgegengewirkt werden sollte.

7.3.3 Moralische und emotionale Aufladung

Klarer nachweisen lässt sich die zunehmende moralische und emotionale Aufladung des Informationsangebots auf vielen Plattformen. Sie ist zum einen deshalb zu verzeichnen, weil emotionsbehaftete Inhalte grundsätzlich mehr Aufmerksamkeit und Reaktionen auslösen und sich so nach aktuell üblichen Kuratierungskriterien gut verbreiten. Gerade Ärger und Wut verbreiten sich besonders schnell und weit und beziehen sich bei öffentlichen Plattformen häufig auf politische Inhalte.[293] Auf Twitter wurde beispielsweise für den vor allem vom nordamerikanischen Diskurs geprägten Raum gezeigt, dass Tweets zu kontroversen Themen umso häufiger geteilt wurden, je mehr moralisch-emotionale Wörter wie z. B. „angreifen“, „schlecht“ oder „beschuldigen“ enthalten waren, und schon ein einziges solches Wort die Verbreitung um 20 Prozent erhöhte.[294] Mit Blick auf die Video-Plattform YouTube konnte derweil nachgewiesen werden, dass deren Empfehlungsalgorithmen im Lauf der Zeit zunehmend extreme Inhalte vorschlagen und der Medienkonsum sich auch entsprechend verändert.[295]

Hinzu kommt, dass besonders gefühlsbetonte oder intensive Reaktionen wie beispielsweise ein mehrfaches Hin-und-her-Kommentieren von Algorithmen mitunter auch noch zusätzlich ausdrücklich positiv gewichtet werden. Die 2018 von Facebook eingeführten Neuerungen seiner Gewichtungskriterien, die laut eigenen Angaben dazu gedacht waren, „Gespräche und sinnvolle Interaktionen zwischen Menschen anzuregen“ und „Menschen näher zusammenzubringen“[296],

---

[293] Fan, R. et al. (2014): Anger Is More Influential than Joy: Sentiment Correlation in Weibo. In: PLoS ONE 9(10): e110184 (DOI: 10.1371/journal.pone.0110184).
[294] Brady, W. et al. (2017): Emotion shapes the diffusion of moralized content in social networks. In: Proceedings of the National Academy of Sciences 114 (28), 7313-7318 (DOI: 10.1073/pnas.1618923114), Appendix Tabelle S3. Die 15 wirkungsvollsten moralisch-emotional aufgeladenen Worte waren: *attack, bad, blame, care, destroy, fight, hate, kill, murder, peace, safe, shame, terrorism, war* und *wrong*.
[295] Ribeiro, M. H. et al. (2020): Auditing Radicalization Pathways on YouTube. In: Conference of Fairness, Accountability, and Transparency, 131-141 (DOI: 10.1145/3351095.3372879).
[296] Mosseri, A. (2018): Bringing People Closer Together. In: Meta. https://about.fb.com/news/2018/01/news-feed-fyi-bringing-people-closer-together/ [13.01.2023].

führten beispielsweise dazu, dass fortan Inhalte priorisiert wurden, zu denen es besonders viele emotionale Reaktionen und Kommentare gab und die häufig geteilt wurden. Nach der Einführung der Änderung stieg die Zahl der Reaktionen und Interaktionen auf Facebook schnell erheblich an. Am besten verbreiteten sich allerdings kontroverse Beiträge, auf die mit dem „Wut"-Emoji reagiert wurde oder die hitzige Diskussionen auslösten. Dieser Trend wurde von Personen und Organisationen, die Facebook zur Kommunikation nutzen, schnell erkannt und führte zu einem teilweise bewussten Wechsel hin zur Produktion provokanterer Inhalte, gerade auch im politischen Bereich, da sich nur noch so Reichweite erzielen ließ (vgl. Abschnitt 7.4.1).[297]

7.3.4 Relevanz der beobachteten Effekte

Wie weit verbreitet die hier postulierten Effekte auf die Informationsqualität sind und welche Relevanz dies für politische Prozesse hat, kann derzeit nicht abschließend beantwortet werden. Die Bedeutung von Filterblasen und Echokammern wird, wie bereits angedeutet, kontrovers diskutiert und mit Blick auf Falschinformationen gibt es zahlreiche und teilweise von Plattformen selbst bereitgestellte Werkzeuge und Angebote, um den Wahrheitsgehalt und die Seriosität von Inhalten eigenständig zu prüfen. Dem stehen selbstverständlich zahlreiche positive Auswirkungen der digitalen Transformation für den öffentlichen Meinungsbildungsprozess entgegen, etwa die Möglichkeit, aus einem weitaus größeren und diversen, idealerweise hochwertigen informationellen Angebot zu wählen.

Gleichzeitig erscheint plausibel, dass Falschnachrichten, Filterblasen und Echokammern sowie eine emotional-moralische Zuspitzung vieler Inhalte im regulären Nutzungsalltag auf algorithmisch kuratierten Plattformen negative Auswirkungen auf die Informationsqualität haben können. Dies gilt vor allem in Situationen, in denen Menschen kein spezielles Erkenntnisinteresse verfolgen, sondern vorwiegend die von der Empfehlungssoftware zuvorderst präsentierten Inhalte konsumieren – erst recht, wenn die Kuratierungsmechanismen nicht bekannt sind oder nicht aktiv reflektiert werden.

Hinzu kommt, dass selbst wo ein Bewusstsein für potenziell problematische Effekte vorhanden ist und der Wille besteht, diesen mit dem eigenen Nutzungsverhalten vorzubeugen, die Umsetzung solcher Vorsätze angesichts der psychologischen Wirkmacht vieler Plattformangebote

[297] Hagey, K.; Horwitz, J. (2021): Facebook Tried to Make Its Platform a Healthier Place. It Got Angrier Instead. In: The Wall Street Journal. https://www.wsj.com/articles/facebook-algorithm-change-zuckerberg-11631654215 [13.01.2023].

sehr schwierig sein kann. Die Personalisierung des Informationsangebots, seine soziale Einbettung durch Likes, Follower, Kommentare und sonstigen Reaktionen wie auch die beschriebene Priorisierung von Inhalten, die besonders überraschend oder emotional anregend sind, sorgen für die fortlaufende Bereitstellung höchst attraktiver Stimuli für das menschliche Belohnungssystem[298], denen man sich nur schwer entziehen kann[299] und die unsere Aufmerksamkeit überaus erfolgreich[300] und nachhaltig[301] beanspruchen. Die Freiheit, qualitativ hochwertige Informationen zu finden, wird unter diesen Umständen durch die Wirkmacht der zum Einsatz kommenden Algorithmen praktisch vermindert.

7.4 Diskursqualität

Die beschriebenen Änderungen in der Qualität, Darbietung und Verbreitung algorithmisch vermittelter Informationen verändern auch die Qualität vieler online stattfindender Diskurse in ethisch wie politisch relevanter Hinsicht. Auch hier sind zunächst wieder positive Entwicklungen und Potenziale zu benennen, die sich insbesondere aus den auf Plattformen und in Sozialen Medien wesentlich erhöhten Möglichkeiten zu Teilhabe und direkter Vernetzung ergeben. Angesichts der weltweit zunehmenden Verfügbarkeit von Online-Zugängen und internetfähigen Endgeräten ist die Schwelle zur Beteiligung an öffentlichen Diskursen für viele Menschen deutlich gesunken. Die Kuratierungskriterien auf Plattformen und in Sozialen Medien geben im Gegensatz zu traditionellen Medien auch weniger etablierten Akteuren und Institutionen die Chance, mit gut formulierten Beiträgen sehr viele Personen direkt zu erreichen und – je nach Plattform – auch von einflussreichen Individuen und Organisationen direkt beachtet zu werden. Im Journalismus beispielsweise wurden aufgrund der erweiterten Möglichkeiten zur direkten Kontaktaufnahme mitunter wesentliche Verbesserungen in der Diskursqualität wahrgenommen.[302]

---

[298] Burhan, R.; Moradzadeh, J. (2020): Neurotransmitter Dopamine (DA) and its Role in the Development of Social Media Addiction. In: Journal of Neurology & Neurophysiology 11 (7), 507. https://www.iomcworld.org/open-access/neurotransmitter-dopamine-da-and-its-role-in-the-development-of-social-media-addiction-59222.html [13.01.2023].

[299] He, Q. et al. (2017): Brain anatomy alterations associated with Social Networking Site (SNS) addiction. In: *Scientific Reports 7* (DOI: 10.1038/srep45064).

[300] Ward, A. F. (2017): Brain Drain: The Mere Presence of One's Own Smartphone Reduces Available Cognitive Capacity. In: Journal of the Association for Consumer Research 2 (2), 140-154 (DOI: 10.1086/691462).

[301] Mickes, L. et al. (2013): Major memory for microblogs. In: Memory & Cognition, 41 (4), 481-489 (DOI: 10.3758/s13421-012-0281-6).

[302] Vgl. Waterson, J. et al. (2022): 'The most effective press watchdog': Owen Jones, Arwa Mahdawi and more on how Twitter changed journalism. In: The Guardian. https://www.theguardian.com/technology/2022/nov/30/twitter-journalism-elon-musk [11.01.2023].

Manche Diskurse erhalten durch die auf Plattformen und Sozialen Medien geschaffenen soziotechnischen Infrastrukturen überhaupt erst einen Raum, weil es über diese Online-Angebote möglich wird, sich mit Menschen mit ähnlichen Interessen oder Problemen ortsunabhängig auszutauschen und zu vernetzen, auch zu sehr speziellen Themen, zu denen sich lokal und offline vielleicht kaum Gelegenheiten zum Diskurs ergäben. Durch solche Zusammenschlüsse können zudem Anliegen, die sonst wohl wenig Chancen auf Wahrnehmung hätten, auch in der breiteren Öffentlichkeit Beachtung finden. Auch Potenziale für die Auseinandersetzung mit Argumenten und anderen Positionen werden durch die neuen Diskursmöglichkeiten online zunächst einmal erweitert. Wer dies möchte, kann mit entsprechenden Suchanfragen und Interaktionen auf Plattformen und in Sozialen Medien die eigenen Horizonte erweitern, Potenziale für Perspektivenwechsel entdecken und mehr Verständnis und Empathie für Andersdenkende entwickeln.

Inwieweit solche Möglichkeiten sich in der Praxis entfalten können, wird allerdings kontrovers beurteilt, denn gegenüber den genannten Chancen werden auch mit Blick auf die Diskursqualität negative Entwicklungen diskutiert. Dabei geht es vor allem um drei Themen. Das erste betrifft die Frage, ob und inwieweit Plattformen und Soziale Medien zu einer in den letzten Jahren beobachteten zunehmenden politischen Polarisierung öffentlicher Diskurse beitragen. Das zweite Thema berührt Praktiken der politischen Werbung und Manipulation und wie diese sich aufgrund der neuen, algorithmisch vermittelten Kommunikationsmöglichkeiten verändern. Ein dritter Diskussionsschwerpunkt gilt Kontroversen um Diskursverrohungen mit Hassrede und ähnlichen Integrationsverletzungen auf der einen und überbordenden Eingriffen in die Äußerungs- und Meinungsfreiheit auf der anderen Seite.

Alle drei Themen beschreiben komplexe Phänomene, die es zweifelsohne nicht erst seit Beginn der digitalen Transformation des öffentlichen Diskurses gibt und die sowohl multifaktoriell bedingt als auch in ihrem Facettenreichtum oft schwer einheitlich fassbar sind. Indizien, dass sich aktuelle Trends in der technisch mediierten öffentlichen Kommunikation über Plattformen und Soziale Medien in diesen Bereichen auf die Diskursqualität auswirken, sind vor dem Hintergrund dieser Komplexität zu betrachten und sollten daher nicht zu vorschnellen Schlüssen über ursächliche Zusammenhänge verleiten.

7.4.1 Politische Polarisierung

Es gibt zahlreiche Hinweise, dass der in Abschnitt 7.3.1 beschriebene Umstand, dass emotional und moralisch aufgeladene Inhalte sich auf Plattformen besonders schnell und weit verbreiten, zu Tonfallverschiebungen geführt hat, auch und insbesondere auf vielen Kanälen, die aktiv zur Gestaltung des politischen Diskurses beitragen. Im Zuge der Enthüllungen der Facebook Files

durch das Wall Street Journal wurde im September 2021 eine Reihe von Belegen veröffentlicht, dass Kommunikationsteams in Bereich Politik sich nach den 2018 eingeführten Änderungen im Facebook-Algorithmus auch gezielt zunehmend provokanteren und negativen Inhalten und Kommunikationsstilen zuwandten und dass Facebook sich dessen bewusst war.

Infokasten 8: Facebook Files – Verminderung der Kommunikationsfreiheit durch algorithmische Formatierung

Im September 2021 veröffentlichte das Wall Street Journal in einer Artikelserie die sogenannten Facebook Files, in denen auf Grundlage von Whistleblower-Informationen zahlreiche Details darüber bekannt wurden, in welchem Ausmaß Facebook gesellschaftlich bedenkliche Auswirkungen seiner Produkte bekannt waren. Dabei ging es auch darum, wie die im Jahr 2018 vorgenommenen Änderungen der Facebook-Algorithmen dazu beigetragen haben, den politischen Diskurs zu polarisieren.[303]

Nach der Einführung neuer Gewichtungskriterien, die ursprünglich zur Förderung „bedeutsamer sozialer Interaktion" konzipiert worden waren, stellte sich bald heraus, dass fortan vor allem solche Inhalte erfolgreich waren, auf die Menschen besonders aufgebracht reagierten, etwa weil sie besonders provokant oder schockierend wirkten oder Themen behandelten, zu denen es ausgeprägte emotionale Kontroversen gab (vgl. Abschnitt 7.3.3). Diese Verschiebung fiel bald auch denjenigen auf, die über Facebook einen Großteil ihrer öffentlichen Kommunikation abwickelten und daher darauf angewiesen waren, über die Plattform Reichweite zu erzielen.

Gerade im politischen Bereich gab es laut dem internen Facebook-Bericht „Political Party Response to '18 Algorithm Change" entsprechende Beobachtungen und daran angepasste Änderungen in der Kommunikationsstrategie hin zu „provokativen Inhalten von niedriger Qualität". Die Reichweite positiver und sachlicher Beiträge hatte sich nach Angaben aus den Kommunikationsteams politischer Parteien aus verschiedenen Ländern deutlich reduziert, während hetzerische Inhalte und direkte Angriffe auf die politische Konkurrenz erfolgreich liefen. Das Soziale-Medien-Team einer polnischen Partei beispielsweise schätzte, dass es den Tenor seiner Inhalte von einer zuvor ausgewogenen Balance zwischen positiven und negativen Beiträgen infolge der neuen algorithmischen Auswahlkriterien zu 80 Prozent in Richtung negative Beiträge verändert hatte, um weiterhin Reichweite zu erzielen. Ähnliche Angaben machten dem Bericht zufolge unter anderem Parteien aus Spanien, Taiwan und Indien, verbunden mit Beschwerden aus den Kommunikationsteams, die diese Entwicklung als negativ und demokratiegefährdend einschätzten – sich aber gleichwohl außerstande sahen, sich ihr zu entziehen.

Erwähnenswert ist weiterhin, dass die Geschäftsleitung einschließlich Mark Zuckerberg nach den Belegen der Whistleblower auch nach Offenlegung der bedenklichen Zusammenhänge Maßnahmen zu deren Entschärfung mit Verweis auf die Priorisierung ökonomischer Interessen ablehnte.

Hieran knüpfen sich zahlreiche Fragen zur Verantwortung von Internetfirmen für das Gemeinwohl, die im weiteren Text noch näher beleuchtet werden.

[303] Hagey, K.; Horwitz, J. (2021): Facebook Tried to Make Its Platform a Healthier Place. It Got Angrier Instead. In: The Wall Street Journal. https://www.wsj.com/articles/facebook-algorithm-change-zuckerberg-11631654215 [11.01.2023]. Alle weiteren Angaben in diesem Kasten stammen aus diesem Artikel.

Durch die algorithmische Konfigurierung und die sich daraus ergebenden Wechselwirkungen zwischen Menschen und algorithmischen Systemen ergaben sich hier also für die am Diskurs teilnehmenden Personen und Organisationen verminderte Handlungsspielräume mit negativen Auswirkungen auf die Diskursqualität. Aus den Facebook-Dokumenten geht zudem laut dem Artikel im Wall Street Journal hervor, dass die Auswirkungen von Änderungen in den Algorithmen auch für die intern bei Facebook an den Programmier- und Entscheidungsprozessen Beteiligten mitunter schwer abzuschätzen waren. Dies unterstreicht die Komplexität der online entstandenen soziotechnischen Systeme, die zu überraschenden Rückwirkungen auf menschliches Handeln führen kann.

Die Beobachtung, dass auf Polarisierung ausgelegte Inhalte erfolgreicher sind, sendet deutliche Signale an alle, die politisch kommunizieren. Erhöhte Chancen auf virale Verbreitung durch die Verwendung einer moralisierenden Sprache, die Empörung über die politische Gegenseite auslöst[304], fallen auch denjenigen auf, die für Kommunikationsstrategien verantwortlich sind und den Erfolg ihrer Kampagnen analysieren. Die Möglichkeit, mithilfe der von den Plattformen bereitgestellten Analysetools sehr schnell Rückmeldung zur Effektivität der eigenen Kommunikation zu erhalten, fördert die ständige Optimierung dieser Effektivität nach den Kriterien der Plattform und treibt selbst nach Einschätzung von Personen, die aktiv an solchen Optimierungsversuchen mitwirken, eine ungünstige Polarisierung des politischen Diskurses voran.[305]

Eine postulierte Spaltung der Öffentlichkeit findet angesichts dieser Beobachtungen womöglich weniger deswegen statt, weil mögliche Filterblasen-Effekte Inhalte und Argumente von außerhalb des personalisierten Präferenzenprofils ausschließen und zu fragmentierten Diskursen führen; vielmehr führt nach dieser Deutung die Priorisierung emotional und moralisch aufgeladener Kommunikation gerade zu einer ständigen Konfrontation mit Inhalten und Argumenten von als gegnerisch wahrgenommenen Gruppen, die auf Grundlage des eigenen personalisierten Präferenzenprofils als bedrohlich, verwerflich oder anderweitig abzulehnend präsentiert werden.[306] Eine verschärfte Polarisierung und Radikalisierung von Positionen und Befindlichkeiten geht demnach vor allem auf diese andauernde Reibung zurück, mit dem Er-

[304] Für Twitter vgl. hierzu etwa Brady, W. J. et al. (2017): Emotion shapes the diffusion of moralized content in social networks. In: Psychological and Cognitive Science 114 (28), 7313-7318 (DOI: 10.1073/pnas.1618923114).
[305] Vgl. Center for Humane Technology (2022): Your Undivided Attention Podcast. Episode 53: How Political Language Is Engineered with Drew Westen and Frank Luntz. https://www.humanetech.com/podcast/53-how-political-language-is-engineered [11.01.2023].
[306] Törnberg, P. (2022): How digital media drive affective polarization through partisan sorting. In: Proceedings of the National Academy of Sciences 119 (42): e2207159119. (DOI: 10.1073/pnas.2207159119).

gebnis, dass die Bereitschaft zum offenen, rationalen und wohlwollenden Diskurs stark eingeschränkt wird und die Spielräume für demokratisches Zusammenwirken im Sinne eines gemeinsamen öffentlichen Vernunftgebrauchs sich damit vermindern.

Solche unterschiedlichen Erklärungsansätze können auch auf der empirischen Ebene zu unterschiedlichen und mitunter widersprüchlichen Resultaten führen. Um möglichen Mechanismen genauer auf den Grund zu gehen, erscheint daher auch in diesem Kontext mehr multimethodische Forschung empfehlenswert, für die auch in diesem Fall wiederum ein besserer Zugang zu Plattformen und ihren Daten notwendig wäre.

Vorerst bleibt aufgrund der noch unzureichenden Studienlage umstritten, welche Rolle algorithmisch mediierte Kommunikationsprozesse für eine beobachtete oder vermutete gesellschaftliche Diskursverschärfung spielen. Angesichts vielfältiger weiterer Einfluss- und Änderungsfaktoren, die auf politisch relevante Situationen, Prozesse und Befindlichkeiten wirken können, gibt es sicherlich zahlreiche weitere Ursachen hierfür, die nicht primär an der besonderen Qualität technisch vermittelter Diskurse liegen. Sie können jedoch im Rahmen solcher Diskurse aufgegriffen werden und Verstärkereffekte erfahren.

7.4.2 Politische Werbung und Manipulation

Angesichts der zunehmenden Bedeutung von Sozialen Medien und Plattformen bei politischen Auseinandersetzungen stehen neben den beschriebenen Polarisierungstendenzen auf offiziellen Kanälen auch weitere, mitunter im Verborgenen laufende Aktivitäten zur Beeinflussung politischer Prozesse in der Diskussion. Nun sind Versuche, die öffentliche Meinung durch wahrhaftige und unwahrhaftige, wahre und falsche Behauptungen und Stellungnahmen zu verändern, ein uraltes Phänomen der Politik, das schon für die antiken Vorformen der modernen Demokratie in Griechenland und Rom eine wichtige Rolle spielte. Auch damals war nicht immer klar, wer mit welchen Absichten welche Behauptungen in die Welt gesetzt hatte und warum sich diese und nicht jene zu einer umfassenden Verleumdungskampagne auswuchs. Durch die Digitalisierung der Kommunikation verschärft sich die Problematik allerdings insofern, als die Zahl kommunikativer Akte nicht begrenzt ist und ihre Kosten minimal sind. So erweitern sich einerseits die Handlungs- und Gestaltungsmöglichkeiten für diejenigen, die an politischen Kommunikationskampagnen mitwirken. Kombiniert mit der Funktionsweise Sozialer Medien und Plattformen, die sowohl eine besonders genaue Personalisierung von Inhalten und Überprüfung ihrer Effektivität ermöglicht als auch die Verbreitung polarisierender, persuasiver, emotionsbetonter und moralisierender Botschaften fördert, ergibt sich zudem viel Potenzial für besonders wirkmächtige Kommunikationskampagnen, die eingebettet in den digitalen Alltag

ablaufen, ohne dass die von solchen Kampagnen erreichten Personen sich dessen gewahr werden. Dies vermindert jedoch andererseits die Freiheit der von diesen Kampagnen adressierten Personen, politische Werbung zu erkennen und sich bewusst damit auseinanderzusetzen.

Die in diesem Zusammenhang relevanten Entwicklungen gehen auf die gleichen Mechanismen zurück, die bereits in Abschnitt 7.2 als grundlegend für die Funktionsweise und das Geschäftsmodell Sozialer Medien beschrieben wurden. Auf Basis der reichhaltigen datenbasierten Profile, die sich aus den digitalen Spuren, die man bei der Nutzung von Plattformen hinterlässt, erstellen lassen, kann man nicht nur konsumrelevante Interessen extrahieren, um das Schalten personalisierter kommerzielle Werbung zu ermöglichen, sondern auch psychologische Merkmale und politische Neigungen sehr präzise ableiten.[307] Hier besteht die Gefahr, dass solche Informationen manipulativ eingesetzt werden, um beispielweise zielgenaue politische Werbung zu schalten (*targeted advertisement*) oder um die Wählerschaft, deren Interessen der politischen Konkurrenz zuneigen, strategisch zu desinformieren oder von der Wahl abzuhalten. Unternehmen wie die bis 2018 operierende Datenanalyse- und Beraterfirma Cambridge Analytica verfolgen ausdrücklich das Ziel, aus dem Nutzungsverhalten und sonstigen Datenspuren politische Persönlichkeits- und Interessenprofile zu erstellen und mit deren Hilfe für bestimmte Gruppen oder Individuen maßgeschneiderte Botschaften zu verbreiten (*microtargeting)*, um politische Meinungsbildungsprozesse und Wahlen zu beeinflussen.

Infokasten 9: Cambridge Analytica

Das Unternehmen Cambridge Analytica wurde 2013 gegründet, um politische Kampagnen mit einer Mischung aus datenbasierter Verhaltensforschung und strategischer Kommunikation zu unterstützten. Seine Kundschaft kam vornehmlich aus dem konservativen Spektrum, darunter die US-Präsidentschaftskandidaten Donald Trump und Ted Cruz. Die Firma beanspruchte, mit ihren Algorithmen besonders präzise psychologische Einschätzungen von Einzelpersonen aus deren Nutzungsdaten ableiten zu können und diese für einen psychografisch personalisierten Zuschnitt politischer Botschaften nutzbar zu machen – ein Ansatz, dessen Wirksamkeit nie eindeutig nachgewiesen und von Cambridge Analytica vermutlich überspitzt dargestellt wurde, um das Interesse am Unternehmen zu schüren.

Die Firma verwendete in ihren Kampagnen aber auch Methoden, die ohne die Berücksichtigung aufwendiger psychografischer Profile auskommen, sondern stattdessen schlicht die Wahlmotivation bestimmter ethnischer oder

[307] Vgl. Kosinski, M.; Stillwell, D.; Graepel, T. (2013): Private traits and attributes are predictable from digital records of human behavior. In: Proceedings of the National Academy of Sciences 110 (15), 5802-5805. (DOI: 10.1073/pnas.1218772110); Cadwalladr, C. (2018): „I made Steve Bannon´s psychological warfare tool": meet the data war wistleblower. In: The Guardian. https://www.theguardian.com/news/2018/mar/17/data-war-whistleblower-christopher-wylie-faceook-nix-bannon-trump [11.01.2023]; Matz, S. C.; Appel, R. E.; Kosinski, M. (2020): Privacy in the age of psychological targeting. In: Current Opinion in Psychology 31, 116-121. (DOI: 10.1016/j.copsyc.2019.08.010).

+++ Vorabfassung! Wird durch die lektorierte Fassung ersetzt. +++

demografischer Gruppen zu senken versuchten. Im US-Präsidentschaftswahlkampf 2016 beispielsweise nutzte Cambridge Analytica Informationen aus Facebook-Profilen, um Schwarzen Menschen, von denen auf Grundlage vorheriger Wahlergebnisse zu erwarten war, dass sie ihre Stimme mit hoher Wahrscheinlichkeit den Demokraten geben, gezielt negative Inhalte über Trumps Konkurrentin Hillary Clinton anzuzeigen, beispielsweise ein Videoclip aus dem Jahr 1996, in dem sie kriminelle Schwarze Jugendliche als „Superraubtiere" bezeichnete.[308] Besonders der Umstand, dass diese Kampagnen oft ganz oder teilweise im Verborgenen abliefen, ohne dass Auftraggebende die Beauftragung offenlegten oder Betroffene wussten, dass sie Ziele personalisierter politischer Werbung waren, wird als demokratiegefährdend gewertet.[309]

2018 gab der Whistleblower Chris Wylie in Interviews bekannt, dass große Teile der Datenbasis von Cambridge Analytica missbräuchlich ohne Einwilligung der Nutzenden beschafft worden waren. Hierfür war eine Fragebogen-App für Facebook-Nutzende entwickelt worden, die nicht nur die Daten von Teilnehmenden sammelte, sondern unbekannterweise auch von den Profilen aus ihrem persönlichen Netzwerk.[310] Dies löste eine breite öffentliche Diskussion über Sicherheit und Zugang zu den Daten in Sozialen Netzwerken sowie die Legitimität personalisierter politischer Werbung aus. 2020 folgten weitere Enthüllungen von der ehemaligen Cambridge-Analytica-Mitarbeiterin Brittany Kaiser, die über Twitter zahlreiche Dokumente veröffentlichte, die eine Einflussnahme des Unternehmens auf Wahlen und Politik in 68 Ländern belegten.[311]

Wie erfolgreich personalisierte politische Werbung mit Microtargeting-Ansätzen tatsächlich sein kann, ist zwar umstritten bzw. noch nicht hinreichend erforscht.[312] Es gibt jedoch deutliche Hinweise darauf, dass schon die akzentuierte Konfrontation mit bestimmten Inhalten politische Meinungen beeinflussen kann. In Experimenten konnte beispielsweise gezeigt werden, dass die Reihenfolge, in der Suchmaschinen Ergebnisse präsentieren, wenn politisch noch unentschiedene Personen online Informationen über Kandidatinnen und Kandidaten suchen, entscheidend auf die Meinungsbildung auswirkt, da weiter oben platzierte Ergebnisse als bedeutsamer eingestuft werden.[313] Allein das Wissen darum, dass versucht wird, auf Grundlage sehr persönlicher psychologischer Merkmale politische Präferenzen zu manipulieren, kann zudem plausibel

[308] Rabkin et al. (2020): Revealed: Trump campaign strategy to deter millions of Black Americans from voting in 2016. In: Channel 4. https://www.channel4.com/news/revealed-trump-campaign-strategy-to-deter-millions-of-black-americans-from-voting-in-2016 [11.01.2023].
[309] Kaiser, B. (2020): Die Datendiktatur. Wie Wahlen manipuliert werden. Hamburg.
[310] Cadwalladr, C. (2018): „I made Steve Bannon´s psychological warfare tool": meet the data war wistleblower. In: The Guardian. https://www.theguardian.com/news/2018/mar/17/data-war-whistleblower-christopher-wylie-faceook-nix-bannon-trump [11.01.2023].
[311] Cadwalladr, C. (2020): Fresh Cambridge Analytica leak 'shows global manipulation is out of control'. In: The Guardian. https://www.theguardian.com/uk-news/2020/jan/04/cambridge-analytica-data-leak-global-election-manipulation [16.01.2023].
[312] Bodo, B. et al. (2017): Political micro-targeting: A Manchurian candidate or just a dark horse? Towards the next generation of political micro-targeting research. In: Internet Policy Review 6 (4). (DOI:10.14763/2017.4.776).
[313] Epstein, R.; Robertson, R. E. (2015): The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. In: Proceedings of the National Academy of Science, E4512–E4521 (DOI: 10.1073/pnas.1419828112).

negative Effekte auf den politischen Diskurs und das Vertrauen in politische Meinungsbildungsprozesse entfalten.

Vertrauensschädigend kann sich weiterhin der Umstand auswirken, dass zur strategischen Beeinflussung des öffentlichen politischen Diskurses auch vielfach unechte Profile (Fake Accounts) eingesetzt werden, die teilweise automatisiert betrieben werden (Bots). In einer investigativen Recherche trug die New York Times hierzu vielfältige Daten zusammen, nach denen beispielsweise bis zu 27 Prozent der Beiträge auf Facebook und Twitter im Zusammenhang mit den Wahlen in Mexiko im Jahr 2018 von Bots und Fake-Accounts stammten und Chinas staatliche Nachrichtenagentur Xinhua hunderttausende falsche Profile und Retweets bezahlte, die Inhalte gezielt an westliche Twitterkonten schicken.[314] In einer Anhörung vor dem US-Senat zum Einfluss Russlands auf die Wahlen im Jahr 2016 legten die Vertreter von Facebook, Twitter und Google offen, dass etliche Personen in den USA über künstliche Profile mit Inhalten der russischen "Agentur für Internet-Forschung" Glavset konfrontiert worden waren, darunter ca. 150 Millionen Personen allein über Inhalte auf Facebook und Instagram.[315]

Auch die Wirkmacht von Fake-Accounts und Bots auf politische Prozesse ist umstritten und lässt sich oft schwer präzise fassen. Schon die Identifikation unechter Profile stellt eine sich ständig fortentwickelnde Herausforderung dar.[316] Es gibt jedoch zahlreiche Hinweise, dass Kommunikationskampagnen über Fake-Accounts zumindest das Potenzial für eine hohe Effektivität haben. So demonstrierte eine experimentelle Studie, dass wenige strategisch gut platzierte Bots, die extreme Botschaften verbreiten, nach aktuellen Kuratierungskriterien überdurchschnittlich erfolgreich auf Entscheidungen Einfluss nehmen können.[317] In einer Untersuchung von über 240 Millionen Tweets zur US-Präsidentschaftswahl 2020 wurde deutlich, dass wenige tausend Bots genauso viele Spitzen im Kommunikationsgeschehen zu aktuellen politischen Themen verursachten wie die Konten echter Menschen, und das vorwiegend, indem sie lediglich über Retweets als Verstärker bestimmter Botschaften wirkten und ihnen

[314] Semple, K. & Franco, M. (2018). Bots and Trolls Elbow Into Mexico's Crowded Electoral Field. https://www.nytimes.com/2018/05/01/world/americas/mexico-election-fake-news.html [12.01.2023]. Confessore, N. et al. (2018): The Follower Factory. In: New York Times. https://www.nytimes.com/interactive/2018/01/27/technology/social-media-bots.html [12.01.2023].
[315] Wortprotokoll der Anhörung des Select Committee on Intelligence des US Senat am 1. November 2017. https://www.congress.gov/event/115th-congress/senate-event/LC55602/text [090.3.2023].
[316] Ferrara, E. et al. (2016): The Rise of Social Bots. In: Communication of the ACM 59 (7), 96-104. (DOI:10.1145/2818717).
[317] Stewart, A. J. et al. (2019): Information gerrymandering and undemocratic decisions. In: Nature 573, 117–121 (DOI: 10.1038/s41586-019-1507-6).

somit mehr Überzeugungskraft verliehen.[318] Mit solchen Verstärkereffekten können Diskurse problematisch verzerrt werden, sodass sie nicht mehr das Interesse oder die Ansichten der an ihnen beteiligten Menschen spiegeln. Verunsicherungen und der Verlust des Vertrauens in den Diskursprozess selbst können die Folge sein, wenn die Legitimität der Diskurse und die Identität der Teilnehmenden zunehmend bezweifelt wird.

Dass solche Zweifel berechtigt sein können, zeigen Analysen, die nachweisen, dass mit unechten Profilen im Rahmen der Funktionsweise eines sozialen Netzwerks auch langfristige Strategien zur subtilen Beeinflussung verfolgt werden. Hier werden beispielsweise zunächst über längere Zeiträume Reichweite und Netzwerke mit unpolitischen Inhalten wie humorvollen Memes (Text- und Bildmontagen) zu identitätsstiftenden Themen wie Familie, Mode oder lokale Kultur aufgebaut. Erst später oder nur punktuell werden Botschaften solcher Konten um politische Inhalte angereichert, in oft subtiler Weise und stark aufbauend auf zuvor etablierten Gruppenidentitäten.[319]

7.4.3 Spannungsfeld Diskursverrohung und Äußerungsfreiheit

Im Zusammenhang mit den beschriebenen Beobachtungen zu Verschärfung von Tonlagen auf Plattformen und in Sozialen Medien wird häufig die Annahme vorgebracht, dass dieser Trend auch zu einer Verrohung des politischen Diskurses beitragen kann. Besorgniserregend ist hier insbesondere die Zunahme stark negativ und aggressiv geprägter Kommunikationsstile bis hin zu Hassrede, Drohungen und Aufforderung zu Gewalt. Regelmäßige Untersuchungen der Landesanstalt für Medien NRW zur Hassrede im Internet zeigen, dass die Wahrnehmung von Hasskommentaren seit 2016 zugenommen hat und seit 2018 auf unverändert hohem Niveau verharrt.[320] Ansätze zur Moderation von Inhalten hinken dieser Entwicklung häufig hinterher. Laut den vom Wall Street Journal in den Facebook Files dargestellten Enthüllungen schätzen Face-

---

[318] Ferrara, E. et al. (2020): Characterizing social media manipulation in the 2020 U.S. presidential election. In: *First Monday* 25 (11) (DOI: 10.5210/fm.v25i11.11431).

[319] DiResta, R. et al. (2019): The Tactics & Tropes of the Internet Research Agency. U.S. Senate Documents. https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1003&context=senatedocs [12.01.2023]; Bradshaw, S. et al. (2022): Playing Both Sides: Russian State-Backed Media Coverage of the #BlackLivesMatter Movement. In: The International Journal of Press/Politics, 1-27 (DOI: 10.1177/19401612221082052); Bradshaw, S.; Henle, A. (2021): The Gender Dimensions of Foreign Influence Operations. In: International Journal of Communication, 15 (23), 4596–4618.

[320] Vgl. zur Entwicklung von 2016 bis 2020: Landesanstalt für Medien NRW (2020): Ergebnisbericht. forsa-Befragung zu: Hate Speech 2020. https://www.medienanstalt-nrw.de/fileadmin/user_upload/NeueWebsite_0120/Themen/Hass/forsa_LFMNRW_Hassrede2020_Ergebnisbericht.pdf [12.01.2023].

book-Angestellte beispielsweise den Erfolg ihrer Bemühungen, Hassbotschaften mit algorithmischen Methoden zu identifizieren und einzudämmen, als überaus unzureichend ein, mit Erfolgsquoten im niedrigen einstelligen Bereich.[321]

Die bloße Existenz solcher Inhalte ist bereits für sich genommen problematisch, da Gehässigkeit, Verleumdungen und Einschüchterungen auf Sozialen Medien so viel Unbehagen und Angst schüren können, dass dies Personen davon abhält, ihre Meinung zu äußern, im Internet präsent zu sein oder sich am öffentlichen Diskurs zu beteiligen. Man spricht auch hier von Chilling-Effekten, die diskriminierend auf Betroffene wirken und deren Freiheit und Handlungsmöglichkeiten in der Online-Kommunikation erheblich vermindern können.

Hinzu kommt die Gefahr, dass online verbreitete Hetze in bedrohliche Handlungen in der realen Welt umschlägt. Auch wenn kausale Zusammenhänge im Einzelfall schwer nachweisbar sein mögen, gibt es genügend Korrelationen zwischen solchen Vorfällen, sodass diese Sorge gerechtfertigt erscheint und jedenfalls plausibel zu den genannten Chilling-Effekten beiträgt. So wurde der 2017 an den Rohingya in Myanmar verübte Völkermord durch zahlreiche Hassbotschaften und Gewaltaufrufe auf Facebook befeuert, das zu diesem Zeitpunkt aufgrund seiner kostenlosen Nutzbarkeit über die Mobilfunknetze monopolartig den Zugang der Bevölkerung zum Internet darstellte. Die Bemühungen des Konzerns, diese problematischen Inhalte zu moderieren, wurden auch hier weitgehend als unzureichend eingestuft.[322] Ähnliche Vorwürfe über die Rolle Facebooks bei ethnisch motivierter Gewalt werden beispielsweise im Zusammenhang mit dem Bürgerkrieg in Äthiopien[323]und religiös motivierter Gewalt in Indien[324] diskutiert.

In einer westlichen Demokratie wurde das demokratiegefährdende Potenzial von online verbreiteten Entrüstungsstürmen am deutlichsten augenfällig, als in den USA am 6. Januar 2021 zahlreiche Personen, die die Wahlniederlage des abgewählten Präsidenten Donald Trump nicht anerkennen wollten, nach Befeuerung in den Sozialen Medien – unter anderem durch Trump selbst – das Kapitol stürmten, um den Senat und das Repräsentantenhaus an der förmlichen

---

[321] Seetharaman, D.; Horwitz, J.; Scheck, J. (2021): Facebook Says AI Will Clean Up the Platform. Its Own Engineers Have Doubts. In: The Wall Street Journal. https://www.wsj.com/articles/facebook-ai-enforce-rules-engineers-doubtful-artificial-intelligence-11634338184?mod=article_inline [12.01.2023].
[322] Stecklow, S. (2018): Hatebook. Why Facebook is losing the war on hate speech in Myanmar. In: Reuters. https://www.reuters.com/investigates/special-report/myanmar-facebook-hate [12.01.2023].
[323] Jackson, J.; Kassa, L.; Townsend, M. (2022): Facebook 'lets vigilantes in Ethiopia incite ethnic killing'. In: The Guardian. https://www.theguardian.com/technology/2022/feb/20/facebook-lets-vigilantes-in-ethiopia-incite-ethnic-killing [12.01.2023].
[324] Purnell, N.; Horwitz, J. (2021): Facebook Services Are Used to Spread Religious Hatred in India, Internal Documents Show. In: The Wall Street Journal. https://www.wsj.com/articles/facebook-services-are-used-to-spread-religious-hatred-in-india-internal-documents-show-11635016354?mod=article_inline [12.01.2023].

Bestätigung des Wahlsieges von Joe Biden bei der Präsidentschaftswahl 2020 zu hindern. Eine Reaktion auf dieses Ereignis, bei dem fünf Menschen starben und Hunderte verletzt wurden, war die Deaktivierung von Trumps Accounts auf Facebook und Twitter durch die entsprechenden Plattformbetreibenden. Die internen Entscheidungsprozesse bei Twitter zur Moderation von Inhalten und Konten im Umfeld dieser Ereignisse wurden nach der Übernahme des Konzerns durch Elon Musk im Oktober 2022 im Rahmen der *Twitter Files* veröffentlicht und teils kritisch beurteilt.[325] Solche Reaktionen werfen ihrerseits demokratietheoretische Fragen auf. Müssen und sollen wir es hinnehmen, dass große private Anbieter, die teilweise eine monopolistische Stellung auf den Medienmärkten haben, entscheiden, wer in den Sozialen Medien zu Wort kommen kann und wer nicht? Wollen wir, dass Community Rules, die von privaten Medien intern festgelegt werden, die Formen der Kommunikation auch jenseits rechtlicher Vorgaben bestimmen?

Der Umgang mit solchen Fragen gestaltet sich schwierig, da hier zwei normativ relevante Gesichtspunkte einander gegenüberstehen. Auf der einen Seite können übermäßige Löschungen und Sperrungen –(Overblocking) einen Eingriff in die Meinungs- und Pressefreiheit darstellen. Overblocking kann selbst zu Chilling-Effekten beitragen, nämlich dann, wenn Menschen bestimmte Inhalte gar nicht erst veröffentlichen, weil sie befürchten, dass diese gleich wieder gelöscht werden und ihrem Konto Einschränkungen drohen. Bei solchen Eingriffen wird auch die Frage relevant, *wer* über Löschungen bzw. deren Kriterien entscheidet. Bereits die Löschpraktiken Sozialer Medien sind solche Einschnitte. Meinungs- und Pressefreiheit sind verfassungsrechtlich in erster Linie Abwehrrechte gegenüber dem Staat, denen das Hausrecht privater Plattformen nicht unterworfen ist. Das bedeutet, dass Plattformen durchaus mehr löschen dürfen und können als nur illegale Inhalte. Nach Auffassung des Bundesgerichtshofs sind die Plattformen aber auch bei der Anwendung der eigenen Regeln, die sie über die Vertragsbeziehungen mit ihrer Kundschaft vermitteln, an die Grundrechte der Personen gebunden, die ihr Angebot nutzen, etwa deren Freiheit zur Meinungsäußerung.[326] Seit Einführung des Netzwerkdurchsetzungsgesetzes 2018 sind sie allerdings verpflichtet, solche illegalen Inhalte *nach* Meldung durch Dritte zu löschen.[327] Auf der anderen Seite stellt sich jedoch die Frage, wie man, wenn

[325] Taibbi, M. (2023): Capsule Summaries of all Twitter Files Threads To Date, With Links and a Glossary. In: Racket News. https://www.racket.news/p/capsule-summaries-of-all-twitter [31.01.2023].
[326] BGH, Urteile vom 29. Juli 2021 – III ZR 179/20 und III ZR 192/20.
[327] 2022 hat die EU den Digital Services Act (DSA) erlassen; Anbieter müssen die neuen Reglungen bis zum 17. Februar 2024 schrittweise umsetzen. Derzeit prüft die Bundesregierung, inwieweit das Netzwerkdurchsetzungsgesetz mit Inkrafttreten des DSA unanwendbar wird, und plant, Materien, die weiterhin nationalstaatlich geregelt werden können, in einem neuen Gesetz über Digitale Dienste zu bündeln.

man solche Eingriffe in die Meinungsfreiheit durch Löschungen und Sperrungen ablehnt, verhindern will, dass Hassbotschaften, Aufrufe zu Gewalt, Verleumdungen und sonstige Einschüchterungen die oben beschriebenen Chilling-Effekte aus Angst vor Übergriffen, Förderung von Gewalt und weitere demokratiegefährdende Wirkung entfalten. Die gebotene Pluralismussicherung widerstreitet jedoch ebenso vorschnellen algorithmenbasierten Exklusionsmechanismen unter einseitiger Berufung auf angeblich vertrauenswürdige Quellen. An sich ungewollten Chilling-Effekten, die sich aus u.a. durch gesetzliche Regelungen (v.a. das Netzwerkdurchsetzungsgesetz) bedingten, zu weit reichenden Blockierpraktiken ergeben, ist konsequent entgegenzuwirken.

Ein möglicher Lösungsansatz sieht vor, Soziale Medien und Plattformen, soweit diese rundfunkartige Funktionen übernehmen (z. B. der Newsfeed von Facebook), in diesen Funktionen anderen Rundfunkanbietern gleichzustellen, sie also an die einschlägigen gesetzlichen Bestimmungen für den Rundfunk zu binden. Die Kuratierung von Inhalten würde dann auf Grundlage einer rechtlichen Regelung erfolgen, die sicherstellt, dass das Recht auf Meinungsfreiheit gewahrt wird, ohne die demokratische Zivilkultur zu gefährden. Zudem könnte die Etablierung einer alternativen Infrastruktur digitaler Kommunikation in der Europäischen Union als Alternative zu privaten und ausschließlich kommerzgetriebenen Angeboten einen Beitrag zur demokratischen Zivilkultur in Zeiten der digitalen Transformation leisten.

7.4.4 Erweiternde und vermindernde Rückwirkungen auf den öffentlichen Vernunftgebrauch

Die genaue Wirkmacht der hier vorgestellten Veränderungen in der Diskursqualität, die sich aus den spezifischen soziotechnischen Merkmalen und Möglichkeiten öffentlicher Kommunikation auf Plattformen und in Sozialen Medien ergeben, ist aktuell noch nicht vollständig abschätzbar. Die vielfältig erweiterten Möglichkeiten zur Teilhabe und Vernetzung sowie grenzüberschreitend wirksamer Kommunikation eröffnen zahlreiche Chancen zur Verbesserung der Diskursqualität und zur Stärkung demokratischer Prozesse.

Die Entfaltung dieser Potenziale scheint in der Praxis jedoch oft schwer planbar zu sein oder wird durch negative Aspekte der Online-Kommunikation behindert. Aufgrund der komplexen Vernetzung von Kommunikation im Internet sind viele politische Diskursprozesse von Einzel-

+++ Vorabfassung! Wird durch die lektorierte Fassung ersetzt. +++

nen, aber auch von großen Gruppen nur begrenzt beeinflussbar. Manipulationen und Gegenmanipulationen, die strategischen Formen der Kommunikation[328], schaukeln sich zu chaotischen Prozessen hoch. Diese lassen der politischen Deliberation, der Abwägung von Gründen pro und contra, womöglich nur noch wenig Spielraum und vermindern so die Möglichkeiten des öffentlichen Vernunftgebrauchs. Extremere und emotionalere Aussagen haben unabhängig von ihrem Wahrheitsgehalt gute Chancen, sich gegenüber abgewogenen Argumenten durchzusetzen oder diese sogar zu marginalisieren.

Speziell mit Blick auf Fake-Accounts und Bots kommt hinzu, dass die hier eingesetzte kalkulierte Verzerrung und bewusste Täuschung selbst Rückwirkungen auf die Diskursqualität und Prozesse politischer Willensbildung haben kann, indem sie zu Vertrauensverlusten führt. Wenn gefälschten Identitäten in die menschliche Interaktion eintreten, agieren sie auch als Substitute für menschliche Subjekte. Die Zurechenbarkeit einer Botschaft zu einem personalen Gegenüber ist in der Folge nicht mehr gegeben und insbesondere die Delegation kommunikativer Akte an Software kann zu Verunsicherungen führen, wenn man nicht unterscheiden kann, ob man mit Menschen oder Bots kommuniziert. Sowohl das Vertrauen in den Diskurs und die mögliche Einigung auf Kompromisse als auch das Vertrauen in demokratische Prozesse insgesamt, die eigene Urteilskraft sowie die individuellen Wirkmöglichkeiten können auf diese Weise vermindert werden.

7.5 Fazit und Empfehlungen

Die hier aufgezeigten Phänomene und Entwicklungen, die sich in den jungen soziotechnischen Infrastrukturen digitaler Netzwerke vollziehen, haben erhebliche Auswirkungen auf Prozesse der öffentlichen Kommunikation sowie der politischen Meinungs- und Willensbildung, auch und vielleicht insbesondere in demokratischen Gesellschaften. Demokratie als kollektive Selbstbestimmung der Freien und Gleichen setzt individuelle Autonomie voraus. Sie gründet auf dem in Kapitel 3 entwickelten Gedanken, dass Freiheit und Vernunft eng verbunden sind. Praktische Vernunft, die Fähigkeit, das eigene Leben, seine Einstellungen und Handlungen von

[328] Vgl. dazu Jürgen Habermas, der in seiner *Theorie des kommunikativen Handelns* drei Typen rationalen Handelns vorstellt: instrumentelles Handeln, strategisches Handeln und kommunikatives Handeln (Habermas, J. (1981): Theorie des kommunikativen Handelns. Frankfurt a. M.) sowie das Gibbard-Satterthwaite-Theorem (Gibbard, A. (1973): Manipulation of Voting Schemes. A General Result. In: Econometrica, 41 (4), 587-601 (DOI: 10.2307/1914083); Satterthwaite, M. (1975): Strategy-proofness and Arrow's Conditions. Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions. In: Journal of Economic Theory 10, 187-217; Nida-Rümelin, J. (2020): Die gefährdete Rationalität der Demokratie. Ein politischer Traktat. Hamburg, Kap. 14 und 15.

Gründen leiten zu lassen, ist eine Voraussetzung von Autorschaft und Voraussetzung für gelingende demokratische Prozesse, die auf rationalen Austausch mit anderen und die freie Auseinandersetzung unterschiedlicher Auffassungen setzen, um auf dieser Grundlage gute Entscheidungen gemeinsam und wohlbegründet treffen zu können.

Diese Diskurs- und Entscheidungskultur hat im modernen öffentlichen Raum der Plattformen und Sozialen Medien vielfältige Entwicklungen erfahren. Durch den enorm verbesserten Zugang zu vielfältigen und auch qualitativ hochwertigen Informationen wie auch durch die stark erhöhten Vernetzungsmöglichkeiten haben sich viele erweiternde Potenziale für einen umfassenderen, sachlich gestützten und dynamischen Austausch ergeben. Menschen, die vorher nur wenig oder keinen Zugriff auf Informationen und Expertise hatten und kaum Möglichkeiten, sich über ihr lokales Umfeld hinaus Gehör zu verschaffen, können mit den von Plattformen und Sozialen Netzwerken angebotenen Infrastrukturen und Mitteln gleichberechtigt am Informationsaustausch teilnehmen, Sichtbarkeit für ihre Anliegen erreichen und Zusammenschluss mit Gleichgesinnten finden.

Trotz dieser auf den ersten Blick erheblich erweiterten Möglichkeiten zu Informationsgewinnung, Teilhabe an Kommunikations- und Entscheidungsprozessen und sozialer Vernetzung – auch mit Menschen, die anderer Auffassung sind –, werden diese Potenziale unter den aktuellen Bedingungen häufig nicht realisiert oder verdeckt von Prozessen, die menschliche Handlungsspielräume und Möglichkeiten für den öffentlichen Vernunftgebrauch und demokratische Verständigungsprozesse eher vermindern. So werden beispielsweise viele Prozesse der Online-Kommunikation durch die beschriebenen algorithmischen Formatierungen immer wieder eingeschränkt und verzerrt.

Die auf eine Maximierung der Aufmerksamkeit und Verweildauer ausgerichtete Kuratierung erhöhen so zwar einerseits den Komfort für Nutzende, wenn diese automatisch zu ihren Präferenzen passende und unterhaltsam bzw. spannende Inhalte erhalten oder in Online-Suchen schnell und leicht die gewünschten Informationen finden. Andererseits laufen aber gerade die nach aktuellen Kuratierungskriterien geförderten schnelllebigen und verbreitungsstarken, polarisierenden, emotional und moralisch aufgeladenen Inhalte und Reaktionen in vielerlei Hinsicht den Grundgedanken demokratischer Kommunikation und Willensbildung zuwider. Insbesondere eine stark von negativen Emotionen geprägte Polarisierung vermindert die Handlungsspielräume für den freien und sachlichen Austausch von Gründen und Argumenten. Wo alternativ oder zusätzlich Effekte wie Filterblasen und Echokammern hinzukommen – Phänomene,

deren Verbreitung nach wie vor unklar ist –, vermindern sich die Möglichkeiten für konstruktive Auseinandersetzungen gegebenenfalls noch zusätzlich, da man dann konträren Auffassungen und Argumenten im Rahmen eines personalisierten Feeds mitunter kaum noch begegnet.

Der Einsatz von persuasiver personalisierter politischer Werbung, oft jenseits klar kenntlich gemachter Kampagnen und unter Einsatz von Fake-Accounts und Bots, gefährdet demokratische Prozesse nicht nur, indem sie den individuellen Vernunftgebrauch mit manipulativen Techniken erschwert, sondern auch, indem durch den täuschenden und verzerrenden Charakter solcher Aktionen das Vertrauen in den demokratischen Prozess selbst untergraben wird. Derartige Vertrauensverluste wiederum begünstigen gemeinsam mit der bereits erwähnten algorithmischen Favorisierung aufmerksamkeitserregender und negativer emotional wie moralisch aufgeladener Inhalte Verschwörungsmythen und andere kollektive Erregungen, die wiederum den öffentlichen Vernunftgebrauch erschweren, der für eine funktionierende Demokratie unerlässlich ist.

Hinzu kommen die problematische Konzentration digitaler Kommunikationsinfrastrukturen in den Händen weniger Konzerne mit vielfach als unzureichend betrachteten Verantwortlichkeiten, Rechenschaftspflichten und öffentlichen Kontrollmöglichkeiten ebenso wie die aktuell zu beobachtenden Defizite in der Moderation problematischer Inhalte.[329] Letztere betreffen sowohl die technische Umsetzung als auch die Frage, wie eine angemessene Balance zwischen zu wenig und zu viel Kontrolle in das digitale Kommunikationsgeschehen gelingen kann. Die aktuelle Aufstellung, in der die auf kommerziellen Plattformen zum Einsatz kommenden soziotechnischen Mechanismen für die Öffentlichkeit weitgehend opak bleiben und selbst Forschung keinen Zugang zu den proprietären Algorithmen hat, geht für die Gesellschaft mit stark verminderten Handlungs- sowie Kontrollmöglichkeiten einher.

In diesem Rahmen stellt sich auch die Frage, ob eine digitale Kommunikationsinfrastruktur in öffentlicher Verantwortung Lösungsansätze für diese Probleme bieten könnte. Das Ziel wäre nicht nur, Menschen eine Alternative zu kommerziellen Plattformen anzubieten, sondern darüber hinaus durch Meinungsvielfalt, kommunikatives Ethos, Minderheitenschutz und Seriosität auch die privaten Anbieter zu einem höheren Maß an Demokratieverträglichkeit zu veranlassen. Der Demokratieentwicklung unter den Bedingungen einer dynamischen digitalen Transformation käme das zugute.

---

[329] Diesen Problemen soll und wird teilweise auch schon mit europäischen und nationalen Gesetzen entgegengewirkt werden, z. B. das Netzwerkdurchsetzungsgesetz (s.o.), Digital Service Act, Digital Market Act etc. Auch der Medienstaatsvertrag sieht solche Regelungen vor.

Online-Plattformen und Soziale Medien bilden die zentralen Infrastrukturen für Information und Kommunikation und sind somit von entscheidender Bedeutung für die öffentliche Meinungsbildung. Gleichzeitig liegen sie jedoch in der Hand weniger globaler Akteure, die primär ökonomische Zwecke verfolgen, wodurch deren sozio-technische Praktiken der Selektion und Moderation einerseits erheblichen Einfluss auf Prozesse der Information, Kommunikation und der öffentlichen Meinungsbildung haben, andererseits sich jedoch effektiver Kontrolle der Rechtmäßigkeit häufig entziehen.

Empfehlungen

- *Empfehlung Kommunikation 1:* Regulierung Sozialer Medien: Es bedarf klarer rechtlicher Vorgaben, in welcher Form und in welchem Ausmaß Soziale Medien und Plattformen über ihre Funktions- und Vorgehensweisen zur Kuratierung und Moderation von Inhalten informieren müssen und wie dies auf der Grundlage institutioneller Regelungen umgesetzt wird. Dies muss durch externe Kontrollen überprüfbar sein; rein freiwillige Ansätze privater Handelnder, insbesondere die unverbindliche Überprüfung durch von diesen selbst besetzen Aufsichtsgremien, sind nicht ausreichend. Hier gibt es auf Ebene der Europäischen Union im Digital Services Act bereits Ansätze, die aber noch nicht weit genug gehen.
- *Empfehlung Kommunikation 2:* Transparenz über Moderations- und Kuratierungspraktiken: Anstelle allgemeiner Moderations- und Löschungsrichtlinien und wenig aussagekräftigen Zahlen über Löschungen muss für externe Kontrollen nachvollziehbar sein, wie, unter welchen Umständen und anhand welcher Kriterien solche Entscheidungen gefällt und umgesetzt werden und welche Rolle hierbei Algorithmen bzw. menschliche Moderierende übernommen haben. Darüber hinaus, müssen auch die grundlegenden Funktionsweisen der Kuratierung von Inhalten Sozialer Medien und Plattformen in dem Ausmaß offengelegt werden, das nötig ist, um systemische Verzerrungen und möglicherweise resultierende informationelle Dysfunktionen erkennen zu können. Die Berichtspflichten und Transparenzvorgaben im Medienstaatsvertrag, im Netzwerkdurchsetzungsgesetz und im Digital Services Act stellen dies noch nicht hinreichend sicher. Die datenschutzrechtlichen Auskunftspflichten gemäß Art. 12 ff. DSGVO sind zum Teil auf nationalstaatliche Ebene beschränkt worden und erfassen oftmals diese weitergehenden Aspekte nicht.
- *Empfehlung Kommunikation 3:* Zugriff auf wissenschaftsrelevante Daten von Plattformen: Um die Wirkungsweisen von Plattformen und Sozialen Medien, ihren Einfluss auf öffent-

liche Diskurse, aber auch weitere Themen von hoher gesellschaftlicher Relevanz zu untersuchen, sollte sichergestellt werden, dass unabhängigen Forschenden der Zugriff auf wissenschaftsrelevante Daten von Plattformen nicht mit dem pauschalen Verweis auf Betriebs- oder Geschäftsgeheimnisse verweigert werden kann. Für den Zugang müssen sichere, datenschutzkonforme sowie forschungsethisch integre Wege gefunden werden. Netzwerkdurchsetzungsgesetz und Digital Services Act enthalten bereits Reglungen zum Datenzugang, die aber in ihrem Anwendungsbereich sehr begrenzt sind; auch der Data Act sieht vergleichbare Regelungen vor.

- *Empfehlung Kommunikation 4:* Berücksichtigung von Sicherheit, Datenschutz und Geheimhaltungsinteressen: Anforderungen an Offenlegungen und Datenzugang müssen kontextsensitiv spezifiziert werden, wobei Anforderungen an Sicherheit und Schutz vor Missbrauch, Datenschutz sowie dem Schutz von intellektuellem Eigentum und Geschäftsgeheimnissen angemessen Rechnung zu tragen ist. Je nach Kontext muss zwischen unterschiedlich klar definierten Zeitpunkten der Prüfung und Graden der Offenlegung unterschieden werden.

- *Empfehlung Kommunikation 5:* Personalisierte Werbung, Profiling und Microtargeting: Personalisierte Werbung ist das zentrale Geschäftsmodell Sozialer Medien und Plattformen. Die Praktiken des Profiling und Microtargeting können jedoch problematische Auswirkungen auf öffentliche Kommunikation und Meinungsbildung entfalten, insbesondere im Kontext politischer Werbung. Um solche negativen Auswirkungen durch effektive Regelungen zu verhindern, ist es zunächst notwendig, die Bedingungen für eine Erforschung und Überprüfung der Zusammenhänge zwischen Geschäftsmodellen und Praktiken algorithmischer Kuratierung in ihren Wirkungsweisen und Effekten zu schaffen. Der auf Ebene der Europäischen Union diskutierte Vorschlag für eine Verordnung über die Transparenz und das Targeting politischer Werbung adressiert diesen Bedarf. Hierbei zeigen sich allerdings auch die Herausforderungen, Regeln so zuzuschneiden, dass sie einerseits wirksam sind, andererseits aber die Freiheit der politischen Kommunikation nicht übermäßig beschränken.

- *Empfehlung Kommunikation 6:* Bessere Regulierung von Online-Marketing und Datenhandel: Ursache vieler der in diesem Kapitel beschriebenen informationellen und kommunikativen Dysfunktionen haben ihre Ursache im Online-Marketing, welches das grundlegende Geschäftsmodell vieler Sozialer Medien und Plattformen ist und auf der Sammlung, Analyse und dem Verkauf vielfältiger Daten über die Personen, die diese Angebote nutzen, beruht. Das Problem ist hierbei nicht die Werbefinanzierung per se, sondern der invasive

Umgang mit diesen Daten. Hier gilt es einerseits, die Auswirkungen dieses Geschäftsmodells auf öffentliche Diskurse besser zu erforschen. Andererseits bedarf es besserer gesetzlicher Regelungen, um sowohl Individuen in ihren Grundrechten online effektiver schützen als auch negative systemische Effekte auf den öffentlichen Diskurs zu minimieren. In diese Richtung gehende Vorschläge hat der Deutsche Ethikrat unter dem Stichwort Datensouveränität in seiner Stellungnahme Big Data und Gesundheit vorgestellt. Europäische Regelungen wie der Digital Markets Act adressieren das Problem der Datenmacht großer Plattformen, aber – schon aus Gründen der Regelungskompetenz – nicht mit Blick auf die Folgen für den öffentlichen Diskurs.

- *Empfehlung Kommunikation 7:* Machtbeschränkung und Kontrolle: Unternehmen, die im Bereich der öffentlichen Vorstellung von Daten bzw. Tatsachen de facto monopolartige Machtmöglichkeiten haben, sind durch rechtliche Vorgaben und entsprechende Kontrolle auf Pluralismus, Minderheiten- und Diskriminierungsschutz zu verpflichten. Ein Teil der Mitglieder des Deutschen Ethikrates ist der Auffassung, dass medienrechtliche Regelungen zur Sicherung von Pluralität, Neutralität und Objektivität generell auf Nachrichtenfunktionen von Sozialen Medien und Plattformen, ausgedehnt werden sollten, sofern sie denen traditioneller Medien ähneln.
- *Empfehlung Kommunikation 8:* Erweiterung der Nutzerautonomie: Plattformen und Soziale Medien sollten ihre Inhalte auch ohne eine personalisierte Kuratierung verfügbar machen. Darüber hinaus sollten sie für die Kriterien, nach denen Inhalte auf Plattformen und in Sozialen Medien algorithmisch ausgewählt und prioritär präsentiert werden, weitere Wahlmöglichkeiten anbieten. Dazu sollte auch die Möglichkeit gehören, bewusst Gegenpositionen angezeigt zu bekommen, die den bisher geäußerten eigenen Präferenzen zuwiderlaufen. Solche Wahlmöglichkeiten sollten gut sichtbar und leicht zugänglich sein.
- *Empfehlung Kommunikation 9:* Förderung kritischer Rezeption von Inhalten: Zur Eindämmung unreflektierter Verbreitung fragwürdiger Inhalte sollten diverse Hinweisfunktionen entwickelt und eingesetzt werden, die eine kritische Auseinandersetzung mit Material fördern, bevor man sich dafür entscheidet, es zu teilen oder öffentlich darauf zu reagieren. Dies könnten etwa Rückfragen sein, ob Texte gelesen und Videos geschaut wurden, bevor man sie teilt, oder Angaben zur Seriosität von Quellen.
- *Empfehlung Kommunikation 10:* Alternative Informations- und Kommunikationsinfrastruktur: Zu erwägen wäre, den privaten Social-Media-Angeboten im europäischen Rahmen eine digitale Kommunikationsinfrastruktur in öffentlich-rechtlicher Verantwortung zur

+++ Vorabfassung! Wird durch die lektorierte Fassung ersetzt. +++

Seite zu stellen, deren Betrieb sich nicht am Unternehmensinteresse eines möglichst langen Verweilens von Menschen auf der Plattform oder an anderen kommerziellen Interessen orientiert. Damit sollte nicht etwa der öffentlich-rechtliche Rundfunk (TV und Radio) auf eine weitere digitale Plattform ausgedehnt, sondern eine digitale Infrastruktur bereitgestellt werden, die eine Alternative zu den kommerzbetriebenen, stark oligopolartigen Angeboten bietet. Um eine hinreichende Staatsferne zu garantieren, könnte auch an eine Trägerschaft in Gestalt einer öffentlichen Stiftung gedacht werden.

+++ Vorabfassung! Wird durch die lektorierte Fassung ersetzt. +++

8 Öffentliche Verwaltung

8.1 Einleitung

Die öffentliche Verwaltung stellt die exekutive Ebene der Umsetzung der in demokratischen Meinungsbildungs- und Entscheidungsverfahren gefassten Beschlüsse dar. Sie versteht sich als *vollziehende Gewalt des Staates.*[330] und lässt sich systematisieren in die Ordnungsverwaltung (Vollzug und Kontrolle von Gesetzen), die Dienstleistungsverwaltung (Umsetzung der gesetzlich begründeten, technischen oder persönlichen Dienstleistungsansprüche aller Bürgerinnen und Bürger), die Wirtschaftsverwaltung (Bewirtschaftung aller Einnahmen, Ausgaben und Vermögen der öffentlichen Hand), die Organisationsverwaltung (Personalwesen einschließlich von Personalrekrutierung sowie Fort- und Weiterbildung) sowie die politische Verwaltung (insbesondere Zuarbeit für die politische Führung sowie Entscheidungstragende in der Legislative).[331]

Für Menschen, aber auch viele Organisationen stellt die Öffentliche Verwaltung, so etwa im Finanz-, Steuer-, Melde- und Sozialwesen und in der Straffälligen- und Jugendgerichtshilfe die unmittelbar erfahrbare Staatsgewalt dar. Ihre Exekutivfunktion darf dabei nicht als bloß ausführende Vollstreckung bereits anderweitig getroffener Entscheidungen verstanden werden. Zwar müssen übergeordnete und direkt demokratisch legitimierte Entscheidungen, etwa im Sozialrecht oder Steuerrecht eingehalten und umgesetzt werden; dennoch verbleiben auf der Ebene der Verwaltung weitreichende Beurteilungs- und Entscheidungsspielräume in der einzelfallbezogenen Adaptation allgemein formulierter Gesetze und Verordnungen, so zum Beispiel unter dem Begriff der Verhältnismäßigkeit, bei unbestimmten Rechtsbegriffen oder Risikoentscheidungen.

Zudem sind der Öffentlichen Verwaltung von Gesetzes wegen mitunter unmittelbare Entscheidungsbefugnisse zugewiesen – auch in der Dienstleistungsverwaltung etwa des Gesundheits- und Sozialwesens. Gerade hier unterliegen die Mitarbeitenden der Öffentlichen Verwaltung einer spezifischen Professionsethik.[332] Funktionierende, transparente, als legitim anerkannte

[330] Schmidt-Assmann, E. (1998): Das allgemeine Verwaltungsrecht als Ordnungsidee. Berlin, 148ff.
[331] Schmidt, M. G. (2010): Art, Verwaltung (Öffentliche) In: ders.: Wörterbuch zur Politik. 3., überar. und akt. Auflage. Hamburg. 859f.; Hesse, J.; Ellwein, T. (1992): Das Regierungssystem der Bundesrepublik Deutschland. 7., völlig neubearb. und erw. Aufl. Opladen, 308ff.
[332] Lob-Hüdepohl, A. (2002): Verantwortung im Verwaltungshandeln. In: Deutsche Verwaltungspraxis. Fachzeitschrift für die öffentliche Verwaltung 53, 45-52; Trappe, T. (2013): Ethik der öffentlichen Verwaltung. Eine Skizze. In: Büsch, D.; Kutscha, M. (Hg.): Recht, Lehre, und Ethik der öffentlichen Verwaltung. Baden-Baden, 145-162.

und bürgernahe Verwaltung ist daher für ein funktionierendes Gemeinwesen und die Akzeptanz von Demokratie und Staat wesentlich.

8.2. Ethische Fragen algorithmischer Automatisierung im Verwaltungshandeln

Seit den 1970er-Jahren werden Konzepte für eine gleichzeitig effizientere wie auch bürgernähere öffentliche Verwaltung im Rahmen von Digitalisierungsstrategien entwickelt, erprobt und teils umgesetzt.[333] Damit verbinden sich unter anderem Hoffnungen auf eine Rationalisierung und Beschleunigung staatlichen Verwaltungshandelns, eine effektivere und kohärentere Datennutzung sowie eine Ausweitung der Einbeziehung wissenschaftlichen und bürgerschaftlichen Sachverstandes. Dem steht die Schreckensvision einer sogenannten „Algokratie" gegenüber, in der autonome Softwaresysteme die staatliche Herrschaft über Menschen ausüben, Bürgerinnen und Bürgern durchgehend Entscheidungen unterworfen sind, deren Algorithmen intransparent sind und keinen Widerspruch dulden.[334] Werden in den positiven Stimmen die Potenziale der Bürgernähe und der niedrigschwelligen digitalen Erreichbarkeit von Verwaltungsleistungen vom Wohnzimmer aus betont, so sehen kritische Stimmen digitale Technologien als weiteren Schritt zu einer technokratischen Bürokratie, in der Kommunikation hinter anonymen Datenmengen und standardisierten, noch dazu schwer verständlichen Benutzeroberflächen verschwindet.

Tatsächlich lässt sich in den vergangenen Jahren in vielen Ländern in und außerhalb Europas ein zunehmender Einsatz von automatisierten Entscheidungssystemen in der öffentlichen Verwaltung beobachten. Beispiele reichen von der Bewertung von Arbeitsmarktchancen Jobsuchender in Österreich[335] und Polen[336] über die Verwendung von Software für die Prüfung und

---

[333] Vgl. Simitis, S., Hornung, D., Spiekter gen Döhmann, I (Hg.) (2019): Datenschutzrecht. DSGVO mit BDSG. Baden-Baden. Einleitung Rn. 6; zu den veränderten Vorstellungen von Verwaltung in diesem Zuge Rn. 7.
[334] Der Begriff der Algokratie geht zurück auf Aneesh, A. (2009): Global Labor: Algocratic Modes of Organization. In: Sociological Theory 27 (4), 347-370; Danaher, J. (2016): The Threat of Algocracy: Reality, Resistance and Accommodation. In: Philosophy and Technology 29 (3), 245-268.
[335] Szigetvari, I. (2020): Gericht macht Weg für umstrittenen AMS-Algorithmus frei. In: Der Standard. https://www.derstandard.at/story/2000122684131/gericht-macht-weg-fuer-umstrittenen-ams-algorithmus-frei [22.02.2023].
[336] Niklas, J.; Sztandar-Sztanderska, K.; Szymielewicz, K. (2015): Profiling the Unemployed in Poland: Social and Political Implications of Algorithmic Decison Making. Herausgegeben von Fundacja Panoptykon. Warschau. https://panoptykon.org/biblio/profiling-unemployed-poland-social-and-political-implications-algorithmic-decison-making [30.01.2023].

Vergabe von Sozialleistungen in England[337] , Frankreich[338] und den Niederlanden[339] bis hin zu prädiktiven Analysen im Betreuungs- und Fürsorgebereich in Finnland[340] und Spanien[341] , aber auch im Bereich der Polizei. Auch wenn in Deutschland der Einsatz von automatisierten Entscheidungssystemen in der öffentlichen Verwaltung noch selten ist[342] , finden sich auch hier einzelne Projekte wie etwa die automatische Berechnung des Arbeitslosengelds durch das IT-System ALLEGRO der Bundesagentur für Arbeit[343] oder das in Hamburg zum Einsatz kommende JUS-IT-System für die Koordination und Abrechnung von Sozialdiensten[344] .

Vielen Vorteilen solcher Systeme, wie der Steigerung der Effizienz von Verwaltungsvorgängen und der besseren Absicherung von Entscheidungen angesichts häufig komplexer Datenlagen, stehen Risiken an anderen Stellen gegenüber. Hier ist zunächst einmal die Qualität der verwendeten Systeme zu bewerten und die Frage zu stellen, ob und in welchem Umfang die verwendeten Systeme Diagnosen und Prognosen tatsächlich verbessern. Es stellt sich in diesem Kontext auch die Frage, ob die Genauigkeit für verschiedene Anwendungskontexte oder für verschiedene Personengruppen gleich ist, oder ob es möglicherweise systematische Verzerrungen oder Diskriminierungen gibt (sogenannter *algorithmic bias*). In ethischer Hinsicht sind hierbei insbesondere zwei Themenkomplexe von besonderer Bedeutung, die einerseits Fragen von Autonomie, Autorschaft und Verantwortung sowie andererseits Fragen der Gerechtigkeit berühren.

Im ersten Fall geht es um die Frage, ob und wie die Nutzung von KI zur Unterstützung von Entscheidungen oder gar das vollständige Delegieren von Entscheidungen an KI-Systeme menschliche Handlungsfähigkeiten und Autorschaft beeinflusst. Bereits in den Bezeichnungen für solche Systeme wird diese Problematik deutlich. Die häufig verwendete Bezeichnung

---

[337] Booth, Robert (2019): Benefits system automation could plunge claimants deeper into poverty. In: The Guardian. https://www.theguardian.com/technology/2019/oct/14/fears-rise-in-benefits-system-automation-could-plunge-claimants-deeper-into-poverty [22.02.2023].
[338] Inland, L. (2021): How French welfare service are creating 'robo-debt'. In: AlgorithmWatch. https://algorithmwatch.org/en/robo-debt-france/ [22.02.2023].
[339] Braun, I. (2018): Risikobürger. In: AlgorithmWatch. https://algorithmwatch.org/de/risikobuerger/ [22.02.2023].
[340] Ruckenstein, M.; Lehtiniemi, T. (2020): Automating Society Report 2020. Finland. In: AlgorithmWatch. https://automatingsociety.algorithmwatch.org/report2020/finland/ [22.02.2023].
[341] Peiró, K. (2019): Automating Society Report 2019. Spain. In: AlgorithmWatch. https://algorithmwatch.org/en/automating-society-2019/spain/ [22.02.2023].
[342] Vgl. Kamps, L. (2020): Automatisierung schreitet auch in Deutschland voran. In: Netzpolitik.org. https://netzpolitik.org/2020/automating-society-report-2020-automatisierung-schreitet-auch-in-deutschland-voran/ [22.02.2023].
[343] Well, L. (2020): Unsere Untersuchung der Hartz-IV-Algorithmen zeigt: Hier diskriminiert der Mensch und nicht die Maschine. In: Algorithm Watch. https://algorithmwatch.org/de/hartz-iv-algorithmen-diskriminierung/ [30.01.2023].
[344] JUS-IT Hamburg. https://www.hamburg.de/jus-it/ [30.01.2023].

ADM-Systeme wird teils in *algorithmic decision-making* teils in *automated decision-making* aufgelöst. In beiden Fällen wird nahegelegt, dass die Entscheidungen vollständig automatisiert erfolgen oder aber Entscheidungen vollständig an Maschinen delegiert werden. Dem gegenüber steht beispielsweise der Begriff *decision support system*, der nahelegt, dass KI-Systeme menschliche Entscheidungen lediglich unterstützen sollen. Diese Unterscheidung kommt auch in der EU-Datenschutz-Grundverordnung zum Tragen, wenn in Artikel 22 postuliert wird, dass niemand zum Objekt einer allein auf Algorithmen basierenden Bewertung gemacht werden sollte, sondern belastende Wertungsentscheidungen von einem Menschen verantwortet werden sollten. Mit Blick auf den sogenannten Automation Bias, das heißt die menschliche Tendenz, sich maschinellen Empfehlungen vorbehaltslos anzuschließen, stellt sich allerdings die Frage, ob dieses Problem nur bei Entscheidungen ohne jedes menschliche Eingreifen und nicht bereits im Fall der Entscheidungsunterstützung zum Tragen kommt.

Schon bei der Nutzung von Software zur Entscheidungsunterstützung, beispielsweise mithilfe von Risikoscores, ändert sich die Rolle der menschlichen Verantwortungs- und Entscheidungsträger grundlegend. Nach einer Phase der Bewährung dieser Systeme kann es leicht zur Dominanz einer Default-Praxis des Verwaltungshandelns in diesem Bereich kommen: Es treten Gewöhnung und Routine ein. Die maschinelle Interpretation der Daten gibt im Regelfall eine Entscheidung vor, von der nur in besonderen Einzelfällen, für die es dann Gründe geben muss, abgewichen wird. Die Entscheidung gegen eine maschinell vorbereitete Empfehlung, wenn der Softwareeinsatz etabliert und im Ganzen bewährt ist, legt der verantwortlichen Person Recherche- und Begründungspflichten auf, die bei komplexen Sachverhalten einen erheblichen Aufwand nach sich ziehen. Daher können auch bei formaler Beschränkung von ADM-Systemen auf Entscheidungsvorbereitung und -unterstützung die Verantwortungsspielräume von Verwaltungsmitarbeitern de facto eingeschränkt werden. Umgekehrt und angesichts der Tatsache, dass auch das traditionelle Verwaltungshandeln von Routinen und Mustern geprägt ist, was im Einzelfall immer wieder zu Fehlentscheidungen führt und der notwendigen Flexibilität bei komplexeren Sachverhalten entgegensteht, kann die Ausweitung der Entscheidungsunterstützung durch ADM-Systeme jedoch auch helfen, starre Routinen aufzubrechen, komplexere Entscheidungssituationen rechtzeitig zu diagnostizieren und eine angemessene Reaktion erst möglich zu.

Wenn Entscheidungen von Softwaresystemen unterstützt oder gar vollständig an diese delegiert werden, stellen sich auch Fragen der Verantwortungszuschreibung. Auch beim vollautomatisierten Verwaltungshandeln muss das Recht der Bürgerschaft gewahrt werden, sich mit den in der Verwaltung Verantwortlichen auseinanderzusetzen, Einspruch zu erheben und letztlich,

wenn keine Einigung erzielt werden kann, zu versuchen, ihr Recht vor dem Verwaltungsgericht durchzusetzen. Daher muss es eine klare und für sie erkennbare Verantwortungszuweisung an eine dafür zuständige und somit verantwortliche Behörde geben.

Das zweite Problem betrifft Fragen der Gerechtigkeit und der Verhinderung von Diskriminierung. Auch hier wird der Begriff Bias verwendet, jedoch mit einer völlig anderen Bedeutung: Im Gegensatz zum zuvor geschilderten Fall des Automation Bias geht es hier nicht um kognitive Verzerrungen menschlicher Akteure, sondern um systematische Verzerrungen im Output der KI-Systeme, die zuweilen als *technical bias*[345] oder Algorithmic Bias beschrieben werden. Ein viel diskutiertes Beispiel hierfür ist die Software COMPAS, die in den USA zur Ermittlung von Risikoprofilen für ehemalige straffällige Personen eingesetzt wird. Hier konnte nachgewiesen werden, dass die Prognosen insofern diskriminierend waren, als dunkelhäutige Menschen deutlich häufiger eine fälschlicherweise zu negative Prognose erhielten als hellhäutige Menschen, die im Gegensatz dazu fälschlicherweise zu positive Prognosen bekamen. Häufige Ursache solch diskriminierender Systeme sind Verzerrungen in den Trainingsdaten, aber auch verschiedene methodische Entscheidungen[346]. Das Resultat ist, dass insbesondere datenbasierte Systeme gesellschaftliche Stereotypen, aber auch gesellschaftliche Ungleichheit in scheinbar neutrale Systeme übersetzen und verschleiern und somit existierende gesellschaftliche Ungleichheit weiter in die Zukunft fortschreiben. Dies ist jedoch keine zwangsläufige Entwicklung. Im Gegenteil können datenbasierte Systeme solche historischen Ungerechtigkeiten auch zuvorderst aufdecken und sie damit Gegenmaßnahmen zugänglich zu machen.

---

[345] Friedman, B.; Nissenbaum, H. (1996): Bias in Computer Systems. In: ACM Transactions on Information Systems, 14 (3), 330-347 (DOI: 10.1145/230538.230561).

[346] Barocas und Selbst (2016) liefern hier eine ausgesprochen hilfreiche Übersicht möglicher Quellen von Bias und Diskriminierung in datenbasierten Systemen. Neben der Auswahl von Zielvariablen, Labels oder relevanten Merkmalen, sind hier vor allem die verwendeten Daten von entscheidender Bedeutung. Hierbei kann es nicht nur durch Fehlerhaftigkeit der Daten selbst, sondern auch durch Über- und Unterrepräsentativität zu Problemen kommen, d.h. wenn zum Beispiel Personen in Datensätzen überproportional oder unterproportional repräsentiert sind. Darüber hinaus besteht das Probleme sogenannter Proxy-Variablen: sensible Kategorien wie beispielsweise Geschlecht, religiöse Zugehörigkeit oder sexuelle Orientierung lassen sich teilweise aus anderen Datenpunkten, wie etwa Hobbies, Einkaufsverhalten, Wohnort, etc. erschließen. Aufgrund dieser sogenannten redundanten Enkodierung können Personen aufgrund ihres Geschlechts oder Ihrer sexuellen oder religiösen Orientierung durch Software auch dann systematisch diskriminiert werden, wenn diese Angaben gar nicht erhoben wurden: eben weil diese Kategorien aus anderen Daten ableitbar sind. Die meisten der zuvor geschilderten Quellen von Bias und Diskriminierung sind nicht beabsichtigt, sondern das Resultat methodischer Entscheidungen, bzw. Fehler. Es ist jedoch prinzipiell auch möglich absichtlich diskriminierende Effekt in Softwaresysteme einzubauen, ein Problem welches Barocas und Selbst Verschleierung (Masking) nennen. Vgl. Barocas, S.; Selbst, D. (2016): Big Data's Disparate Impact. In: California Law Review 104 (3), 671-732.

Eine grundsätzliche Grenze für die Anwendung von ADM-Systemen liegt in nicht eliminierbaren normativen Ziel- oder Regelkonflikten im deutschen, deontologisch verfassten Rechtssystem. Folgenabwägung allein bestimmt in der deontologisch orientierten Ordnung in Deutschland nicht das Rechtmäßige. Sie kommt zwar zum Einsatz, zum Beispiel im Zusammenhang mit der Verhältnismäßigkeit, wenn es darum geht, das mildeste Mittel im Sinne eines Eingriffs in Grundrechte zu bestimmen, das dennoch für die Zielerreichung effektiv ist. Aber in der Bestimmung des deontologisch Rechtmäßigen geht es nicht um das mehr oder weniger Vorteilhafte und nicht um Nutzenoptimierung in Bezug auf die Folgen, sondern um die Frage, wer welche Rechte hat und wie diese im Konfliktfall so abzuwägen und zu gewichten sind, dass unbedingte Ansprüche auf Schutz der Person gewahrt und ihre Verrechenbarkeit (im Sinne einer reinen Nutzenoptimierung) verhindert werden kann.[347] Insoweit sind der quantitativen Berechnung Grenzen gesetzt; vielmehr es geht um eine nur menschlicher Urteilskraft zugängliche Bewertung. So müssen in menschlicher Deliberation Argumente und Gegenargumente vorgebracht werden, auf deren Grundlage am Ende entschieden wird (vgl. Abschnitt 3.2.2). Hier zeigt sich eine ultimative Grenze der Algorithmisierung ethischer und rechtlicher Entscheidungsprozesse.

Diese grundsätzlichen Überlegungen sprechen nicht dagegen, Teilbereiche der Entscheidungsprozesse zu algorithmisieren und die rechtliche und moralische Deliberation mit Softwaresystemen zu unterstützen. Im Gegenteil geht es darum, auch im Verwaltungshandeln die Vorteile von digitalen und insbesondere KI-gestützten Systemen zu nutzen, ohne jedoch in die genannten technokratischen Fallen zu tappen. Die Gestaltung entsprechender Systeme kann proaktiv von Beginn an so angelegt werden, dass ethische Werte und Normen berücksichtigt werden, wie dies beispielsweise im Ansatz des Value Sensitive Design angestrebt wird.[348] Dies kann der Unterstützung und Erweiterung menschlicher Autorschaft dienen.

Im Folgenden werden diese Fragen anhand von Beispielen aus den Bereichen des Sozialwesens (vgl. Abschnitt 8.3) und des Polizeiwesens (vgl. Abschnitt 8.4) exemplarisch illustriert und vertieft. Die beiden Felder wurden ausgewählt, weil in ihnen häufig Entscheidungen von großer Tragweite und möglicherweise großer Eingriffstiefe vorgenommen werden. Daher erscheint es einerseits geboten, Technologien zu nutzen, welche die Qualität dieser Entscheidungen verbessern können, indem sie beispielsweise genauere Vorhersagen liefern und Fehlentscheidungen

[347] Ross, D. (1930): The Right and the Good. Oxford.
[348] Friedman, B.; Hendry, D. G. (2019): Value Sensitive Design: Shaping Technology with Moral Imagination. Cambridge (MA), London.

verringern. Ob KI-Technologien die Entscheidungsqualität jedoch verbessern, hängt von deren Qualität ab. So gilt es einerseits zu prüfen, ob und in welchem Ausmaß es durch den Einsatz von Software zur Entscheidungsunterstützung tatsächlich zu weniger Fehlern in beide Richtungen, das heißt weniger falsch-positiven und weniger falsch-negativen Resultaten kommt. Zum anderen ist zu prüfen, ob sich Genauigkeit und Fehleranfälligkeit für unterschiedliche Personengruppen unterscheiden, um mögliche diskriminierende Effekte zu vermeiden.

8.3 Automatische Entscheidungssysteme am Beispiel des Sozialwesens

Im Sozialwesen sind auf der Grundlage einschlägiger (sozial-)rechtlicher Gesetze beziehungsweise Ausführungsbestimmungen Entscheidungen mit weitreichenden Folgen für die Betroffenen zu treffen, so etwa über die Gewährung von Hilfen, bei Maßnahmen im Kontext einer Kindeswohlgefährdung oder bei der Abschätzung von Gefährdungspotenzialen straffälliger Personen in der Bewährungshilfe. Solche Entscheidungen werden von den involvierten Fachkräften der Sozialverwaltungen für gewöhnlich unter enger Einbindung der Betroffenen und Leistungsberechtigten getroffen. So umfasst der Schutzauftrag des Jugendamtes bei Kindeswohlgefährdung nicht nur die Erhebung des Gefährdungsrisikos unter Einbeziehung aller relevanten Fachkräfte, die mit dem betroffenen Kind und/oder den Personensorgeberechtigten beruflich in Kontakt stehen (§ 8a Abs. 1 SGB VIII) , sondern auch die unmittelbare Einbindung der Erziehungsberechtigten sowie des betroffenen Kindes (§ 8a Abs. 4 Nr. 3 SGB VIII). Oftmals münden Entscheidungen der Öffentlichen Verwaltung in konkrete Hilfeplanungen, die auf der Basis von Hilfeplanbesprechungen oder sogar förmlichen Hilfeplankonferenzen gemeinsam mit den Betroffenen erstellt und verbindlich vereinbart werden, nachdem ebenfalls gemeinsam auf der Basis einer Ressourcen- und Potenzialanalyse der entsprechende Hilfebedarf erhoben wurde. Im Zentrum der Entscheidungsprozesse stehen deshalb Gespräche, in denen die lebensweltliche Expertise der Betroffenen bzw. der Leistungsberechtigten mit der professionellen Expertise der Fachkräfte verknüpft werden.[349] Dies dient nicht nur der besseren Ergebnis-

[349] Schwabe, M. (2008): Methoden der Hilfeplanung. Zielentwicklung, Moderation und Aushandlung. Frankfurt a. M.; Moch, M. (2018): Hilfen zur Erziehung. In: Otto, H-U. et al. (Hg.): Handbuch Soziale Arbeit, 6. Auflage. München, 632-645. Vgl. zur (verdrängten) Fehlerhaftigkeit intuitionsgestützten Entscheidens und der dringenden Notwendigkeit verbesserter Entscheidungsgrundlagen ausführlich auch Schwabe, M. (2022): Die „dunklen Seiten“ der Sozialpädagogik. Über den Umgang mit Fehler, Unvermögen, Ungewissheit, Ambivalenzen, Idealen und Destruktivität. Weinheim, Basel.

sicherung der professionellen Entscheidung, sondern unmittelbar auch der Steigerung der Lebensführungskompetenz der Betroffenen, insofern sie durch die gemeinsame Problemdiagnose auch der eigenen Ressourcen gewahr werden können.

Die Implementierung algorithmenbasierter Entscheidungshilfen in die unterschiedlichen Arbeitsbereiche des Sozialwesens[350] steht in Deutschland erst am Anfang.[351] Andere europäische oder transatlantische Länder weisen bereits ein erheblich höheres Implementierungsniveau auf. Gleichwohl befinden sich auch in Deutschland eine Reihe von Systemen in der Einführungs-, Erprobungs-, oder Vorbereitungsphase.[352] Eine ethische Einordnung, die auch die internationale Entwicklung im Blick hat, ist deshalb auch für Deutschland von großem Interesse. Diese soll in den nächsten Schritten entlang der Unterscheidung Erweiterung versus Verminderung professioneller Handlungsoptionen erfolgen, wobei es gerade im Sozialwesen von Bedeutung ist, die Erweiterungs- bzw Verminderungsdimensionen gerade auch aus der Perspektive der von den Entscheidungen jeweils betroffenen Personen in den Blick zu nehmen.

8.3.1 Erweiterung professioneller Handlungskompetenz

Unbeschadet der fachlichen Bedeutsamkeit je spezifischer Personenorientierung im Rahmen von Entscheidungsprozessen bieten Verfahren der Entscheidungsfindung, die auf einem gewissen Maß an Standardisierung – auch und gerade unter Einbeziehung algorithmenbasierter Erfassung und Steuerung – beruhen, gegenüber herkömmlichen Verfahren wichtige Vorteile. Häufig hängen Entscheidungen über die Gewährung von Hilfen und anderen Interventionen im Sozialwesen im Wesentlichen von intuitiven Einschätzungen der Fachkräfte ab, selbst da, wo eine stärkere Standardisierung von Entscheidungsfindungen möglich wäre. Solche intuitiven Einschätzungen profitieren zwar von der beruflichen Erfahrung und dem in den relevanten Bereichen des Sozialwesens durchaus stark verbreiteten innerkollegialen Austausch. Wie eine

[350] Schneider, D. (2020): Decision Support Systeme in der Sozialen Arbeit – Herausforderungen an die Rolle der TA in Innovationsprozessen. In: Nierling, L.; Torgersen, H (Hg.): Die neutrale Normativität der Technikfolgenabschätzung.: Konzeptionelle Auseinandersetzung und praktischer Umgang. Baden-Baden, 117-138.

[351] Vgl. entsprechende Beiträge in Kutscher, N. et al. (2020): Handbuch Soziale Arbeit und Digitalisierung. Weinheim, Basel. Insbesondere Teil VI Digitalisierung in Handlungsfeldern der Sozialen Arbeit, 439-599.

[352] Chiusi, F. et al. (2020): Report Automation Society 2020. Herausgegeben von Algorithm Watch und der Bertelsmann Stiftung. https://automatingsociety.algorithmwatch.org/wp-content/uploads/2020/12/Automating-Society-Report-2020.pdf [12.01.2023],114 ff.

hohe Zahl empirischer Studien gleichwohl belegt, bleiben sie aber hinsichtlich ihrer Sachangemessenheit unter dem Qualitätsniveau „statistisch-aktuarischer“[353] oder sogar „mustererkennend-statistischer“[354] Ansätze mittlerweile deutlich zurück – möglicherweise zum Nachteil der von den Entscheidungen Betroffenen. Dies zeigt sich besonders bei der Generierung prognostischen Wissens, also bei der Einschätzung über Gefährdungs- oder Entwicklungspotenziale hilfebedürftiger Personen. Für soziale Professionen, bei denen eine gute Wirkung von Maßnahmen im Mittelpunkt stehen muss, dürfen die Vorteile evidenzbasierter Instrumente daher nicht vernachlässigt werden.[355] In dieser Hinsicht kann die Verwendung von Softwaresystemen zur Entscheidungsunterstützung grundsätzlich zur Erweiterung professioneller Handlungskompetenz beitragen. Verbesserte Ergebnisse professioneller Entscheidungen führen zudem in der Regel zu erweiterten Gestaltungsspielräumen der professionell unterstützten Hilfeempfangenden und dies ist der entscheidende Maßstab zur Bewertung der Güte sozialprofessioneller Interventionen.

Eine Verbesserung der Entscheidungsgrundlage ist besonders wichtig bei der Abschätzung von Gefährdungspotenzialen, denen entweder besonders vulnerable Personen (Stichwort Kindeswohlgefährdung) ausgesetzt sind oder die etwa von straffällig gewordenen Personen für Dritte beziehungsweise für die Allgemeinheit insgesamt ausgehen.[356] Internationale Studien belegen, dass die Treffsicherheit von Prognosen zur Eintrittswahrscheinlichkeit, die auf der Basis standardisiert-versicherungsmathematischer Berechnungen vorgenommen werden, deutlich höher ist als jene, die sich mehr oder minder allein auf die zwar fachlich geschulte, gleichwohl we-

[353] Rettenbach, M. (2018): Intuitive, klinisch-ideographische und statistische Kriminalprognosen im Vergleich – die Überlegenheit wissenschaftlich strukturierten Vorgehens. In: Forensische Psychiatrie, Psychologie, Kriminologie 12; 28-36 (DOI: 10.1007/s11757-017-0463-y).

[354] Sowohl der aktuarialistische als auch der mustererkennende Ansatz ist ein mechanisches Prognoseverfahren und beruht auf speziellen Berechnungen. Während das aktuarialistische Verfahren statistische Modelle verwendet, die stets durch Studien validiert werden müssen, sind die mustererkennenden Modelle zur Risikoeinschätzung dynamisch angelegt. Vgl. Schrödter, M., Bastian, P. (2020): Risikodiagnostik und Big Data Analytics. In: Kutscher, N. et al. (2020): Handbuch Soziale Arbeit und Digitalisierung. Weinheim, Basel, 255-264.

[355] Lob-Hüdepohl, A. (2021): Messen welcher Wirkung? Normativ-handlungstheoretische Vorbemerkungen zur Wirkungsmessung sozialprofessioneller Interventionen. In: Eurich, J.; Lob-Hüdepohl, A. (Hg.): Gute Assistenz für Menschen in Behinderungen. Wirkungskontrolle und die Frage nach dem gelingenden Leben. Stuttgart, 70-86.

[356] Butz, F. et al. (2021): Automatisierte Risikoprognosen im Kontext von Bewährungsentscheidungen. In: Bewährungshilfe, 68 (3), 241–259.

sentlich intuitivere Einschätzung von Fachkräften selbst mit hoher beruflicher Erfahrung stützen.[357] Aus diesem Grund bilden besonders in den angelsächsischen Ländern prädiktive Risikomodelle mittlerweile einen festen Bestandteil in der Entscheidung von Fachkräften bei der Frage, ob in einer bestimmten Fallkonstellation die Familiensituation zum gegenwärtigen Zeitpunkt oder in Zukunft die Gefährdungslage für ein Kind ein solches Ausmaß angenommen hat oder haben wird, dass das gefährdete Kind in staatliche Obhut genommen werden sollte oder aber noch in der Familie verbleiben kann.[358]

Eine vorschnelle präventive Inobhutnahme verbietet sich nicht nur deshalb, weil damit elterliche Sorgerechte und damit Kinderrechte insgesamt berührt sind, sondern weil auch die Herausnahme eines Kindes aus der Familie durch den Verlust von sozialen Alltagsbindungen und so weiter sein Wohl belastet. Damit stehen sich in diesen Entscheidungskonflikten in erster Linie nicht die moralischen Ansprüche zweier unterschiedlicher Personen (Eltern gegen Kind) gegenüber, wohl aber zwei moralisch relevante Güter des Kindes (beispielsweise der Schutz vor übergriffigem, gewaltförmigem Verhalten der Eltern und der Schutz der gewachsenen emotionalen Bindungen innerhalb der Familie), die beide zum Kindeswohl gehören, in der konkreten Situation aber kollidieren. Bestrebungen gehen nun dahin, diese ethische Güterabwägung mithilfe algorithmenbasierter Entscheidungsunterstützung auf eine bessere Grundlage zu stellen.[359]

Infokasten 10: Kindeswohlgefährdung

ADM-Systeme im Kontext der Entscheidungsfindung bei Kindeswohlgefährdung sind mittlerweile auf nahezu allen Kontinenten wenigstens in ersten Schritten eingeführt. Manche dieser Anwendungen sind hoch umstritten, etwa aufgrund von Ungenauigkeit, Fehleranfälligkeit oder diskriminierenden Effekten.

In Deutschland befindet sich die Einführung von ADM-Systemen im Bereich der Kinder- und Jugendhilfe zur Gefährdungsabschätzung des Kindeswohls in der Planungs- und Diskussionsphase. Bereits etablierter Software-Einsatz im Bereich des Kindesschutzes schafft dazu die nötigen Voraussetzungen. Das OK.JUS[360], das Bestandteil

[357] Bastian, P. (2012): Die Überlegenheit statistischer Urteilsbildung im Kinderschutz. Plädoyer für einen Perspektivwechsel hin zu einer angemessenen Form sozialpädagogischer Diagnosen. In: Bode, M. et al. (Hg.): Rationalitäten des Kinderschutzes. Wiesbaden, 249-267; Bastian, P. et al. (2017): Risiko und Sicherheit als Orientierung im Kinderschutz. Deutschland und USA im Vergleich. In: Soziale Passagen 9, 245-261; Bastian, P. (2018): Bauchgefühle in der Sozialen Arbeit. In: Kommission Sozialpädagogik: Wa(h)re Gefühle. Sozialpädagogische Emotionsarbeit im wohlfahrstaatlichen Kontext. Weinheim, Basel, 128-140; Gillingham, P. (2021): Big Data, prädiktive Analytik und Soziale Arbeit. Ein Überblick. In: Sozial Extra 1, 31-35. (DOI: 10.1007/s12054-020-00348-6).

[358] Gillingham, P. (2021): Big Data, prädiktive Analytik und Soziale Arbeit. In: Sozial Extra 1, 31-35. (DOI: 10.1007/s12054-020-00348-6).

[359] Vgl. etwa das Forschungsprojekt „KAIMo – Kann ein Algorithmus im Konflikt moralisch kalkulieren?“ (www.kaimo.bayern); Gutwald, R. et al. (2021): Soziale Konflikte und Digitalisierung. Chancen und Risiken digitaler Technologien bei der Einschätzung von Kindeswohlgefährdungen. In: Ethik Journal 7 (2).

[360] https://www.akdb.de/loesungen/oksoziales/okjus [16.01.2023].

+++ Vorabfassung! Wird durch die lektorierte Fassung ersetzt. +++

einer breiten Palette von Digitalanwendungen in der Öffentlichen Verwaltung ist, unterstützt Jugendämter bzw. die fallführenden Professionellen in der Wahrnehmung ihres Schutzauftrags bei Kindeswohlgefährdung gemäß § 8a SGB VIII. Dieser Schutzauftrag besteht darin, dass im Falle von gemeldeten Hinweisen auf eine Kindeswohlgefährdung das Jugendamt auf der Basis bisheriger Erkenntnisse zuverlässig und angemessen reagiert. OK.JUS dokumentiert nach vordefinierten Standards die verschiedenen Arbeitsabläufe ab dem Meldeeingang bis zu etwaigen Interventionen sowie die Bewertung von Meldungen einschließlich die zur Entscheidung notwendigen Unterlagen. Damit entstehen E-Akten, die in Verbindung mit OK.Jus CAP, einer Controlling- und Analyseplattform mit einer Vielzahl von Einzelberichten der fallführenden Fachkraft eine fundiertere Einschätzung der Kindeswohlgefährdung ermöglichen. Zugleich entsteht ein umfangreicher Datenpool, auf den das ADM-System zukünftig zugreifen kann.

Deutlich weiterentwickelt sind ADM-Systeme in anderen europäischen Ländern bzw. in Übersee.[361] In Dänemark etwa war das „Gladsaxe-model" – benannt nach einem Ort im Umkreis Kopenhagens – eines der ersten Profiling-Systeme, das verschiedene Risiko-Indikatoren kombiniert, um frühzeitig gefährdete Kinder in ihren Familien zu erkennen. Es verbindet und gewichtet elternbezogene Sachverhalte wie Erwerbslosigkeit, seelische Gesundheit, nicht eingehaltene medizinische bzw. zahnmedizinische (Termin-)Vereinbarungen oder auch Scheidungen. Die endgültige Einführung wurde nach ersten Pilotphasen auch aufgrund erheblicher Proteste zurückgestellt; das System wurde aber in Forschungskontexten weiterverwendet.[362]

In den Niederlanden starteten bereits Ende der 2000er-Jahre Versuche, das in den USA entwickelte California Family Risk Assessment für den niederländischen Kinderschutz nutzbar zu machen und auf seine prognostische Validität hinsichtlich des Risikos für Gewalt oder Vernachlässigung von Kindern in problemaffinen Familien hin zu überprüfen.[363] Dieses algorithmenbasierte Assessment-Tool zielt auf die Stärkung standardisierter bzw. strukturierter Fallführung (structured-decision-making). Es kombiniert zehn verschiedene Items, die das Risiko einer zukünftigen Vernachlässigung indizieren, mit weiteren zehn, die das Risiko eines zukünftigen Missbrauchs anzeigen sollen. Neben Aspekten wie unsichere Wohnverhältnisse oder früher eingetretene Gefährdungssituationen beziehen sich viele Items auf die (elterlichen) Sorgeberechtigten (seelische Gesundheit, Suchtprobleme, Rechtfertigung von Missbrauch durch den Sorgeberechtigten, Dominanz und Strenge und so weiter). Studien ergaben eine beachtliche Treffsicherheit dieses Prognoseinstruments für ein frühzeitiges Erkennen von Gefährdungslagen.[364]

Demgegenüber legen Studien aus den USA und Neuseeland nahe, dass sich die Erwartungen an ADM-Systeme mit Blick auf eine verbesserte prädiktive Analytik nicht im gewünschten Umfang erfüllen.[365] Das prädiktive Risikomodell CARE (Neuseeland) oder Family Screening Tool (Pennsylvania/USA) erweisen sich etwa aufgrund der

---

[361] Vgl. für eine vergleichende Studie: Drake, B. et al. (2020): A Practical Framework for Considering the Use of Predictive Risk Modeling in Child Welfare. In: The ANNALS of the American Academy of Political and Social Science 692 (1), 162-181 (DOI: 10.1177/0002716220978200).
[362] Chiusi, F. et al. (2020): Report Automation Society 2020. Herausgegeben von Algorithm Watch und der Bertelsmann Stiftung. https://automatingsociety.algorithmwatch.org/wp-content/uploads/2020/12/Automating-Society-Report-2020.pdf [16.01.2023], 52.
[363] van der Put, C. E. et al. (2016): Detection of unsafety in families with parental and/or developmental problems at the start of family support. In: BMC Psychiatry 16 (15), 1-13. (DOI: 10.1186/s12888-016-0715-y).
[364] van der Put, C. E. et al. (2016): Detection of unsafety in families with parental and/or developmental problems at the start of family support. In: BMC Psychiatry 16 (15), 1-13. (DOI: 10.1186/s12888-016-0715-y).
[365] Gillingham, P. (2016): Predictive Risk Modelling to Prevent Child Maltreatment and Other Adverse Outcomes for Service Users: Inside the 'Black Box' of Machine Learning. In: The British Journal of Social

Fehlerhaftigkeit der genutzten bzw. algorithmisch verarbeiteten Items als zu wenig präzise oder sogar als offensichtlich verfälschend und diskriminierend, was die Spielräume für die Lebensgestaltung der Betroffenen erheblich vermindert.

Eine ähnliche prädiktive Risikodiagnostik erfolgt auch in der Bewährungshilfe.[366] Dort sind Fachkräfte gehalten, Art und Weise ihrer Bewährungsführung dem Gefährdungspotenzial, die von der unter Bewährungsaufsicht stehenden Person nach ihrer Haftverbüßung für andere ausgehen, anzupassen und entsprechend engmaschig oder weitläufig anzulegen. Auch hier ist die Entscheidung keinesfalls trivial: Eine engmaschige, also mit vielen kontrollierenden Auflagen und Maßnahmen belegte Bewährungsführung geht automatisch zulasten der Freiheit der bewährungspflichtigen Person. Umgekehrt könnten bei einer weitläufigen Bewährungsführung unzumutbare Risiken für Dritte ausgehen. Deshalb versucht etwa der in den USA entwickelte und zwischenzeitlich auch in Deutschland in einigen Bundesländern etablierte Risk-Need-Responsivity-Ansatz[367] mithilfe von algorithmenbasierten statistisch-aktuaristischen Prognoseinstrumenten das individuelle Rückfallrisiko mehr oder minder valide zu ermitteln[368] und damit – wie bei Entscheidungen im Rahmen der Kindeswohlgefährdung – gerichtsfest zu machen.

Infokasten 11: Bewährungshilfe

In Deutschland orientiert sich die Bewährungshilfe in der Abschätzung von Gefährdungsrisiken überwiegend am Risk-Need-Responsivity-Ansatz. Dieser bemisst die Leistungsgestaltung der Bewährungshilfe am jeweiligen Gefährdungsrisiko, den daraus sich ableitenden Bedarfen sowie der spezifischen Ansprechbarkeit der leistungsempfangenden Person. Bei der individuellen Rückfallrisikoabschätzung werden acht Risikofaktoren (kriminelle Vorbelastung, prokriminelle Einstellungen, prokriminelle Kontakte, antisoziale/dissoziale Persönlichkeitsbezüge, Bindungen im Bereich Ehe/Familie, Bindungen im Bereich Schule/Arbeit, Substanzmittelmissbrauch sowie Freizeitaktivitäten) einbezogen.[369] Diese Risikofaktoren werden unter Zuhilfenahme von statistisch-aktuarischen

Work 46 (4), 1044–1058 (DOI:10.1093/bjsw/bcv031); Gillingham, P. (2021): Big Data, prädiktive Analytik und Soziale Arbeit. Ein Überblick. In: Sozial Extra 1, 31-35 (DOI: 10.1007/s12054-020-00348-6).

[366] In ähnlicher Weise gehen Einschätzungen über die Legalprognose bei richterlichen Entscheidungen über das Aussetzen einer Freiheitsstrafe gemäß §§ 56, 57, 57a StGB ein. Auch hier eröffnet sich wie bei der Strafzumessung insgesamt ein breites Spektrum denkbarer Einsatzfelder (vgl. Kaspar, J.; Höffler, K.; Harrendorf, S. (2020): Datenbanken, Online-Votings und künstliche Intelligenz: Perspektiven evidenzbasierter Strafzumessung im Zeitalter von „Legal tech". In: Neue Kriminologie 32 (1), 35-56) Diese Einsatzmöglichkeiten werden im Folgenden nicht weiter thematisiert, da sie nicht den Bereich der Öffentlichen Verwaltung fallen.

[367] Polascheck, D. L. L. (2012): An appraisal of the Risk-Need-Responsivity (RNR) model of offender rehabilitation and its application in correctional treatment. In: Legal and Criminological Psychology 17 (1), 1-17 (DOI: 10.1111/j.2044-8333.2011.02038.x).

[368] Cornel, H.; Pruin, I. (2021): Die Implementierung der Risikoorientierung in den Bundesländern. In: Cornel, H.; Kawamura-Reindl, G. (Hg.): Bewährungshilfe. Theorie und Praxis eines Handlungsfeldes Sozialer Arbeit. Weinheim, Basel, 105-118.

[369] Bonta, J.; Andrews, D. A. (2010): The psychology of criminal conduct. Abingdon, 44 ff.

Prognoseinstrumenten gewichtet. Sie bilden die Grundlage für die von der jeweils zuständigen Fachkraft durchzuführende Risikoabschätzung, die in den einzelnen Bundesländern allerdings unterschiedlich strukturiert und teilweise standardisiert ist.[370]

In der Schweiz sind ADM-Systeme im Rahmen des Risikoorientierten Sanktionenvollzugs (ROS) in Vollzugs- wie Bewährungsdiensten bereits etabliert.[371] Das ROS sieht die Abklärung des individuellen Rückfallrisikos einer gewalttätigen Person im Rahmen eines vierstufigen Prozesses vor, das erstens aus einer Triage, zweitens aus einer genaueren Abklärung des Falles, drittens aus der Planung etwaiger Interventionsinstrumente und viertens aus dem kontinuierlich ausgewerteten Verlauf besteht. Kern der ersten Stufe, in der die Dringlichkeit (Triage) einer genaueren Risikoabschätzung abgeklärt wird, bildet ein Fall-Screening-Tool (FaST), das als ADM-System die jeweiligen Fälle drei Kategorien zuordnet: Die Fälle der Kategorie A mit einem geringen Rückfallrisiko benötigen keinerlei weitere Risikoprognosen; die Fälle der Kategorie B mit einem erhöhten Rückfallrisiko benötigen als weiteren Risikoprognoseschritt eine Kurzabklärung, die die fallführende Fachkraft mittels einer Checkliste aus verschiedenen Quellen (Strafregister, Gutachten und so weiter) erstellt (Fallresümee); die Fälle der Kategorie C mit einem hohen Rückfallrisiko bedürfen der präzisen Beurteilung aller verfügbaren Informationen durch eine forensisch geschulte Fachkraft (Psychologin oder Psychologe) (Risikoabklärung) im Sinne einer strukturierten Urteilsbildung. Das FaST selbst kombiniert und gewichtet unterschiedliche Items und klassifiziert den jeweils anstehenden Fall auf der Basis statistisch-aktuarischer Verfahren.[372]

Bereits vor mehr als einem Jahrzehnt wurde in Wisconsin/USA das aufgrund diskriminierender Prognosen sehr umstrittene COMPAS-System eingeführt *(*Correctional Offender Management Profiling for Alternative Sanctions).[373] Es kann sowohl bei der richterlichen Strafzumessung wie beim nachfolgenden Strafvollzug oder beim späteren Bewährungsdienst zur Anwendung kommen. Es wurde im Bemühen einer evidenzbasierten Entscheidungsfindung eingeführt, um in den verschiedenen Phasen eines Fallverlaufes sowohl aufseiten der entscheidenden Person das Ungewissheitspotenzial als auch aufseiten der Bevölkerung das Gefährdungspotenzial durch möglicherweise rückfällige gewalttätige Personen zu verringern. COMPAS selbst kommt sowohl im Kontext der Untersuchungshaft wie auch während und nach Haftzeit (Bewährung) zur Anwendung. Es kombiniert 137 Items und differenziert damit drei Grobklassifizierungen[374]. COMPAS umfasst neben einer Risikoprognose auch ein darauf aufbauendes needs-assessment, das in die weitere Interventionsplanung eingeht. COMPAS gilt in der internationalen Debatte über die Einführung von ADM-Systemen in Vollzugs- und Bewährungsdienste trotz – oder

[370] Vgl. Cornel/Pruin 2021 Cornel, H.; Pruin, I. (2021): Die Implementierung der Risikoorientierung in den Bundesländern. In: Cornel, H.; Kawamura-Reindl, G. (Hg.): Bewährungshilfe. Theorie und Praxis eines Handlungsfeldes Sozialer Arbeit. Weinheim, Basel, 105-118.
[371] Treuthardt, D. et al. (2018): Der Risikoorientierte Sanktionenvollzug (ROS) – aktuelle Entwicklungen. In: Schweizer Zeitschrift für Kriminologie 2, 24-32.
[372] Treuthartdt, D.; Kröger, M. (2019): Der Risikoorientierte Sanktionsvollzug (ROS) – empirische Überprüfung des Fall-Screening-Tools (FaST). In: Schweizer Zeitschrift für Kriminologie 1, 76-85.
[373] Angwin, J. et al. (2016): Machine Bias. In: ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [10.02.2023].
[374] Hier kommt mit PROXY ein weiteres ADM-Tool zur Anwendung, das die Risikoabschätzung auf der Basis lediglich dreier Indikatoren (Alter, Alter bei einer ersten Haft, Anzahl früherer Inhaftierungen) vornimmt und zur Klassifizierung von „low risk“, „medium risk“ und „high risk of reoffending“ führt. Hartmann, K.; Wenzelburger, G. (2021): Uncertainty, risk and the use of algorithms in policy decisions: a case study on criminal justice in the USA. In: Policy Science 54, 269–287 (DOI: 10.1007/s11077-020-09414-y).

+++ Vorabfassung! Wird durch die lektorierte Fassung ersetzt. +++

gerade wegen – der erheblichen Vorbehalte, die ihm national und international gegenüberstehen, als zentraler Bezugspunkt.[375]

8.3.2 Verminderung von Entscheidungskompetenz, Handlungsoptionen und Autorschaft

Nachdem vorausgehend insbesondere die Erweiterung von Handlungskompetenz im Mittelpunkt stand, ist demgegenüber zu konstatieren, dass menschliche Autorschaft unter Zuhilfenahme sachdienlicher Ergebnisse von KI-Algorithmen allerdings auch vermindert werden kann. Auf der professionellen Seite kann dies beispielsweise dann der Fall sein, wenn sie unbesehen zu einer bloßen Übernahme algorithmisch vorgeschlagener Ergebnisse führt – aus welchen Gründen auch immer (z. B. Zeitdruck, allmähliche Gewöhnung oder um im Falle einer juristischen Auseinandersetzung auf der vermeintlich sicheren Seite zu sein). Auch für die von den Ergebnissen betroffenen Personen sind negative Effekte möglich, etwa wenn ihnen aufgrund algorithmisch unterstützter Entscheidungen Handlungs- oder Entwicklungsmöglichkeiten genommen werden.

So dokumentieren die hier vorgestellten Einsatzmöglichkeiten prädiktiver Diagnosen von Gefährdungspotenzialen in den Bereichen Kindeswohlgefährdung und Bewährungshilfe auf der einen Seite einen gewissen Zugewinn an Sachangemessenheit und Wirksamkeit von sozialprofessionellen Interventionen, wenn die Entscheidung der Fachkräfte durch algorithmenbasierte, statistisch-aktuarische Eintrittswahrscheinlichkeitsberechnungen unterstützt werden. Auf der anderen Seite offenbaren sich gleichwohl auch Schwachstellen und Risiken für eine sachgemäße und nicht zuletzt personenorientierte Entscheidung, die den individuellen Hilfebedarfen der Betroffenen Rechnung zu tragen hat.

Hier treten wieder die beiden zuvor genannten Problemfelder auf: Einerseits können sich bei datenbasierten Systemen diskriminierende Effekte zeigen (Algorithmic Bias) So kann bereits die Datenlage, die den standardisierten Entscheidungen und mehr noch den digitalisierten algorithmischen Prozessen der Mustererkennung zugrunde liegen, die Sachlage verfälschen und gefährliche Stereotypisierungen und gesellschaftliche Ungleichheit aus der Vergangenheit in die Gegenwart und Zukunft verlängern. Wenn beispielsweise bei der prognostischen Einschätzung der Kindeswohlgefährdung auf Prädiktoren zurückgegriffen wird, die – wie das Beziehen von Sozialhilfe, alleinerziehende Elternschaft oder früherer Kontakt mit einer Jugendschutzbehörde – ohne Beachtung des konkreten Kontextes höchst zweifelhaft sind, dann werden sie

[375] Angwin, J. et al. (2016): Machine Bias. In: ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [17.01.2023].

gleichsam automatisch zukünftige Entscheidungen fehlleiten und bestehende Ungleichheiten (Beziehen von Sozialhilfe), Vorurteilsstrukturen (alleinerziehende Elternschaft, vorurteilsbelastete oder sogar bösartige Fehlanzeigen bei Kinderschutzbehörden) unbesehen fortschreiben.[376]

Andererseits steigt das Risiko eines Automation Bias mit Auswirkungen auf Fragen der Autonomie, Autorschaft und Verantwortung.[377] Gerade bei Entscheidungen, die mit einer großen prognostischen Unsicherheit konfrontiert sind und zugleich gravierende Auswirkungen haben, besteht die latente Tendenz, automatisierten Entscheidungsprozeduren mehr zu vertrauen als menschlichen Entscheidungen und damit die Verantwortung – zumindest unbewusst – von sich auf diese „Quasi-Akteure" zu delegieren. Bisweilen wird versucht, dieser Gefahr vorzubeugen, indem bei Verwendung eines Entscheidungsunterstützungstools ein entsprechender Warnhinweis gegeben wird.[378] Eine weitere denkbare Vorkehrung wäre die Verpflichtung der entscheidenden Fachkräfte, die etwaige Übernahme des algorithmischen Entscheidungsvorschlags – etwa mit Verweis auf die eigene erfahrungsbezogen intuitive oder kollegial erörterte Prognose – ausdrücklich zu begründen.

Mit Blick auf die Bewährungshilfe birgt die Nutzung statistisch-aktuarischer Risikoprofile darüber hinaus die Gefahr, dass die eigentlich primäre Zielsetzung der Bewährungshilfe, nämlich die subjektiven wie die objektiven Ressourcen der bewährungspflichtigen Person für Rehabilitation und Wiedereingliederung in die Gesellschaft zu aktivieren, zurückgedrängt wird, zugunsten der Risikominimierung im Bereich der Gefährdung für andere.[379] Diese Verschiebung geht eindeutig zulasten der bewährungspflichtigen Person und vermindert im Falle der Nichtaktivierung ihrer Ressourcen ihre Lebensführungskompetenz.[380]

Insgesamt zeigt sich eine ambivalente Entwicklung für die Entscheidungshoheit der eingesetzten Fachkräfte: Algorithmenbasierte Entscheidungshilfen sollen lediglich fachlich weniger angemessene, intuitive Entscheidungsgrundlagen ersetzen und die Basis für qualitativ bessere

---

[376] Gillingham P. (2021): Big Data, prädiktive Analytik und Soziale Arbeit. Ein Überblick. In: Sozial Extra 1, 31-35 (DOI: 10.1007/s12054-020-00348-6), 32.
[377] Safdar, N. et al. (2020): Ethical considerations in artificial intelligence. In: European Journal of Radiology 122: 108768. (DOI:10.1016/j.ejrad.2019.108768).
[378] In den USA wird bei Verwendung des COMPAS-Algorithmus zur Risikoprognose von (ehemaligen) Strafgefangenen ein entsprechender Warnhinweis gegeben. Über die Wirksamkeit dieses Warnhinweises liegen bislang aber keine Erkenntnisse vor. Vgl. Butz, F. et al. (2021): Automatisierte Risikoprognosen im Kontext von Bewährungsentscheidungen. In: Bewährungshilfe 86 (3), 241-259, 252.
[379] Ghanem, C. (2021): Bewährungshilfe zwischen Risiko- und Ressourcenorientierung. In: Cornel, H.; Kawamura-Reindl, G. (Hg.): Bewährungshilfe. Theorie und Praxis eines Handlungsfeldes Sozialer Arbeit, Weinheim, Basel, 84-91.
[380] Butz, F. et al. (2021): Automatisierte Risikoprognosen im Kontext von Bewährungsentscheidungen. In: Bewährungshilfe 86 (3), 241-259, 254 ff.

Entscheidungen der Fachkräfte liefern, welche jedoch die letztverantwortliche Entscheidungsbefugnis beibehalten. Faktisch aber können die algorithmenbasierten Entscheidungshilfen eine präfigurative Wirkmacht entfalten, die die letztendscheidenden Fachkräfte schon in ihrer Wahrnehmung von Hilfebedarfen in vorgegebene Deutungs- und Entscheidungsmuster lenkt. Dieser Effekt wird verstärkt, wenn sich Fachkräfte für eine von den digital vorgeschlagenen Optionen abweichende Entscheidung rechtfertigen müssen – etwa vor Vorgesetzten oder gar vor Gerichten. Damit kann es im faktischen Verhalten zu einer Verminderung der Entscheidungskompetenz der Fachkräfte und mithin zu ihrer schleichenden Verdrängung oder sogar Ersetzung kommen. Die Entscheidungskompetenz der Fachkräfte würde prekär: formal und prima facie gegeben, faktisch aber unterwandert durch die Wirkmacht algorithmenbasierter Entscheidungsvorgaben.

Auch in anderen Bereichen der öffentlichen Sozialverwaltung kommt es zu ähnlichen Entwicklungen, wenn algorithmische Systeme zur Prognose des mutmaßlichen Entwicklungspotenzials von Hilfsbedürftigen eingesetzt werden, beispielsweise bei Entscheidungen über die Gewährung von Sach-, Geld- und Dienstleistungen (z. B. Sozialhilfe oder Weiterbildungsmaßnahmen für Erwerbslose). Algorithmenbasierte Prognosen und Klassifizierungen können hier unmittelbare Folgen für die Art und das Ausmaß von Unterstützungsmaßnahmen haben und damit entscheidend für die Gewährung oder auch Verweigerung von Lebenschancen sein. Dies kann zur Verminderung von Entwicklungsmöglichkeiten und Autorschaft der betroffenen Personen führen.

Die professionell ausgestaltete dialogisch interaktive Beziehungsarbeit bildet den zentralen Rahmen zur Identifikation eines individuellen Hilfebedarfs. Der Einsatz standardisierter Erfassung von Hilfebedarfen birgt demgegenüber das Risiko der Entkopplung aus einer dialogischen Beziehungsarbeit, jedenfalls dann, wenn nicht explizit dialogische Konzepte weiterhin vorgesehen bleiben. Der Hilfebedarf erschließt sich nicht allein aus objektiven Tatbestandsmerkmalen, sondern muss aus fachlichen Gründen die subjektive Perspektive der hilfebedürftigen Person auf die eigene Lebenslage einbeziehen. Eine Einbindung der betroffenen Person ist auch entscheidend für die Erfahrung von Selbstwirksamkeit[381], ohne die positive Effekte selbst materieller Hilfeleistungen schnell verpuffen und damit kaum nachhaltig wirken. Die Gefahr, dass

[381] Beck, I.; Greving, H. (2012): Lebenswelt, Lebenslage. In: dies. (Hg.): Lebenslage und Lebensbewältigung. Stuttgart, 15-59; Grunwald, K.; Thiersch, H. (2017): Praxishandbuch Lebensweltorientierte Soziale Arbeit. Weinheim, Basel.

die Individualität von Klientinnen und Klienten ignoriert und damit die Wirksamkeit der angebotenen Unterstützung verfehlt wird, steigt im Zuge algorithmenbasierter Informatisierung des Sozialwesens erheblich an.[382]

Ein eindrückliches Beispiel für die letztgenannte Problematik algorithmischer Chancenprognose ist der sogenannte Arbeitsmarktchancen-Assistenz-System-Algorithmus (AMAS-Algorithmus) in Österreich.[383] Er kommt in der Sozialverwaltung dann zum Einsatz, wenn bei einer arbeitssuchenden Person die Chancen ermittelt werden, innerhalb eines bestimmten Zeitfensters erfolgreich wieder in ein reguläres Erwerbsarbeitsverhältnis zu gelangen. Das System klassiert alle Arbeitssuchenden in die Klassen mit niedriger, mittlerer oder hoher Erfolgsprognose. Dementsprechend werden ihnen bestimmte Service-, Beratungs- oder Betreuungsleistungen angeboten oder aber auch nicht angeboten. Zwar ist die Entscheidung der Fachkräfte formal vom Vorschlag des Systems abgetrennt. Personen, die mit ihrer Entscheidung von der algorithemenbasierten Einstufung abweichen und eine Umstufung vornehmen, müssen dies aber ausdrücklich vermerken und begründen. Der österreichische AMAS-Algorithmus – so eine interdisziplinäre Studie der Österreichischen Akademie der Wissenschaften und des dortigen Instituts für Technikfolgenabschätzung – ist ganz offensichtlich an den Werten, Normen und Zielen einer restriktiven Fiskalpolitik ausgerichtet. Ein solches System ist mitnichten neutral. Vielmehr läuft es den Zielen eines personenorientiertes Hilfesystems, das individuelle Hilfebedarfe betroffener Personen fokussieren muss, diametral zuwider.[384] Mit dem Einsatz solcher Systeme ist oftmals die Abkehr von einem zentralen Prinzip des Sozial- und Wohlfahrtswesens verbunden: der Ermittlung des individuellen Hilfebedarfs und der auf diesem Hilfebedarf aufruhenden Angebote, neben sächlicher und finanzieller Unterstützung besonders durch Dienstleistungen wie Weiterbildungsangebote jeder hilfebedürftigen beziehungsweise leistungsberechtigten Person eine Lebensführung zu ermöglichen, die „der Würde des Menschen entspricht“ (§ 1 Abs. 1 SGB II).

---

[382] Ley, T., Reichmann, U. (2020): Digitale Kommunikation in Organisationen der Sozialen Arbeit. In: Kutscher, N. et al. (Hg.): Handbuch Soziale Arbeit und Digitalisierung. Weinheim, Basel, 241-254.; Görder, B. (2021): Die Macht der Muster. Die Ethik der Sozialen Arbeit vor professionsbezogenen und gesellschaftlichen Herausforderungen durch ‚künstliche Intelligenz‘. In: Ethik Journal 7 (2). https://www.ethikjournal.de/fileadmin/user_upload/ethikjournal/Texte_Ausgabe_2021_2/Goerder_Ethikjournal_2.2021.pdf [08.02.2023].

[383] Mager, A. et al. (2020): Der AMS Algorithmus – Eine Soziotechnische Analyse des Arbeitsmarktchancen-Assistenz-Systems (AMAS). Endbericht. Wien. https://epub.oeaw.ac.at/ita/ita-projektberichte/2020-02.pdf [08.02.2023].

[384] Allhutter, D.; Cech, F.; Fischer, F.; Grill, G.; Mager, A. (2020): Algorithmic Profiling of Job Seekers in Austria: How Austerity Politics Are Made Effective. In: Frontiers in Big Data, 3(5), 1-17 (DOI: 10.3389/fdata.2020.00005).

Angesichts der zuvor geschilderten Problematik ist es nicht verwunderlich, dass die Automatisierung der öffentlichen Verwaltung im Sozialwesen häufig kritisch gesehen wird. So wird in einem vielbeachteten Sonderbericht der Vereinten Nationen vor der Gefahr einer „digitalen Wohlfahrtsdystopie", das heißt vor der Schaffung einer menschenrechtsfreien Zone gewarnt, die insbesondere dann problematisch sei, wenn private Unternehmen für die Entwicklung und Implementierung von IT-Lösungen für den Sozialstaat verantwortlich zeichnen.[385] Diese Kritik ist keine Einzelmeinung. In den letzten Jahren haben sich immer mehr journalistische, wissenschaftliche und zivilgesellschaftliche Arbeiten und Initiativen kritisch mit den Auswirkungen staatlicher Automatisierungsbemühungen im Sozialwesen befasst. So kommt etwa Virginia Eubanks zu dem Schluss, dass automatisierte Entscheidungssysteme das soziale Sicherheitsnetz in den Vereinigten Staaten unterminieren, mit einem Vertrauensverlust dem Staat gegenüber einhergehen und somit Armut und Ungleichheit und Ungerechtigkeit weiter verstärken würden.[386] Zu einem ähnlichen Verdikt kommt Monique Mann im Zuge ihrer Analyse des australischen RoboDebt-Sozialhilfeprogramms, das gefährdete und schutzbedürftige Bevölkerungsgruppen unrechtmäßig mit Schulden belaste, anstatt Armut und Ungleichheit zu verringern.[387]

Während Automatisierungsprojekte von staatlicher Seite meist als Mittel zur Kostenreduktion und erhöhter Effizienz gepriesen werden, zeigen Analysen wie die oben genannten mögliche Gefahrenpotenziale und weisen auf die Notwendigkeit einer breiten öffentlichen Debatte insbesondere über die ethischen Fragen hin. Die Verminderung menschlicher Autorschaft hat viele Facetten. Aufseiten der verantwortlichen menschlichen Entscheiderinnen und Entscheider kann sie in einer Herabstufung der Betrachtung individueller Fälle zugunsten einer statistik- und datengetriebenen Klassifikation bestehen oder auch in der schleichenden Gewöhnung an ADM-Verfahren. Dadurch kann auch die Schwelle sinken, trotz eigener anderslautender Einschätzung der Empfehlung der ADM zu folgen und damit die Grenze von einem *decision support* zu einem *automatic decision making* zu überschreiten, in dem der Maschine die Letztentscheidungsmacht überlassen wird. Aufseiten der von den Entscheidungen jeweils Betroffenen zeigt sich die vermindernde Wirkung in der Nichtbeachtung von Aktivierungspotenzialen für eine eigenständige Lebensführung bis hin zur Verweigerung eigentlich offenstehender und nutzbarer Entwicklungschancen.

---

[385] Alston, J. (2019): Report of the Special Rapporteur on Extreme Poverty and Human Rights. A/74/493. https://undocs.org/A/74/493 [22.02.2023], 2.
[386] Eubanks, V. (2018): Automating Inequality. How High-Tech Tools profile, police, and punish the Poor. New York.
[387] Mann, M. (2020): Technological Politics of Automated Welfare Surveillance: Social (and Data) Justice through Critical Qualitative Inquiry. In: Global Perspectives, 1 (1), 1-12 (DOI: 10.1525/gp.2020.12991).

8.4 Predictive Policing – KI in der Kriminalitätsbekämpfung

In der Kriminalitätsbekämpfung können algorithmenbasierte Risikoanalysen nicht nur – wie im Falle der Bewährungshilfe – überführte straffällige Personen betreffen, sondern auch zur Identifikation potenziell straffälliger Personen im Rahmen präventiver Polizeiarbeit eingesetzt werden. Üblicherweise werden unter dem Begriff des *Predictive Policing* analytisch-technische, häufig algorithmenbasierte Anwendungen verstanden, die präventive Polizeiarbeit unterstützen und mittels Prognosen künftiger Straftaten, straffälliger Personen und Tatorte der Verhinderung von Verbrechen dienen.[388]

Die entsprechenden Programme kommen mit dem Versprechen daher, große Mengen an mitunter heterogenen und verstreuten Daten schneller miteinander zu verknüpfen und so zur Verbrechensbekämpfung auswerten zu können. Verwendet werden sowohl raum- als auch personenbezogene Verfahren. Raumbezogene Methoden identifizieren durch Verknüpfung diverser Datenbestände Orte, an denen mit einer bestimmten Wahrscheinlichkeit eine Straftat begangen wird, während personenbezogene Methoden mit Täterprofilen und Opfermerkmalen arbeiten, um Personen, die mit hoher Wahrscheinlichkeit straffällig werden könnten, vorab zu erkennen. In den USA kommen sie seit 2005 zum Einsatz.

Infokasten 12: Predictive Policing

In Deutschland kommen bislang in erster Linie raumbezogene Verfahren des Predictive Policing zum Einsatz, die durch die Ausweisung prädiktiver Risikogebiete mit zeitlicher Präferenz gekennzeichnet sind.[389] In der Hauptsache richten sich die Verfahren auf die Vorhersage raumzeitlicher Parameter bei Wohnungseinbruchsdiebstählen.[390] Dabei kommen folgende Programme zum Einsatz: PRECOBS, SKALA, KrimPro, PreMAP und KLB-operativ.[391] PRECOBS, PreMAP und KLB-operativ beruhen auf der Near-Repeat-Methode[392], basierend auf dem Umstand,

[388] Härtel, I. (2019): Digitalisierung im Lichte des Verfassungsrechts – Algorithmen, Predictive Policing, autonomes Fahren. In: Landes- und Kommunalverwaltung, 49-60. , 54f.; Rademacher, T. (2017): Predictive Policing im deutschen Polizeirecht. In: Archiv des öffentlichen Rechts 142 (3), 366-416 (DOI: 10.1628/000389117X15054009148798) (368 ff.). Auch wenn Bewährungsentscheidungen nicht durch Polizeibehörden, sondern Gerichte getroffen werden, lassen sich auch dort Verfahren ähnlich dem Predictive Policing verwenden, weil diese Entscheidungen einem gefahrenabwehrrechtlichen Zweck verschrieben sind. Der Sache nach geht es um eine Gefährlichkeitsprognose in Bezug auf eine Person, die bereits eine Straftat begangen hat anlässlich dieser konkreten Tat.

[389] Povalej, R.; Volkmann, D. (2021): Predictive Policing. In: Informatik Spektrum 44, 57-61 (DOI: 10.1007/s00287-021-01332-4), 58.

[390] Egbert, S. (2018): Predictive Policing in Deutschland. In: ders. (Hg.): Räume der Unfreiheit. Berlin, 241-265, 244.

[391] Egbert, S. (2020): Predictive Policing als Treiber rechtlicher Innovation? In: Zeitschrift für Rechtssoziologie 40, 26-51 (DOI: 10.1515/zfrs-2021-0002), 33.

[392] Knobloch, T. (2018): Vor die Lage kommen: Predictive Policing in Deutschland. Chancen und Gefahren datenanalytischer Prognosetechnik und Empfehlungen für den Einsatz in der Polizeiarbeit. Herausgegeben von der Stiftung Neue Verantwortung e.V. und der Bertelsmann Stiftung. https://www.stiftung-nv.de/sites/default/files/predictive.policing.pdf [10.02.2023], 13.

+++ Vorabfassung! Wird durch die lektorierte Fassung ersetzt. +++

dass nach einer Ausgangstat in räumlich-zeitlicher Nähe häufig weitere Taten begangen werden[393], also dass das Begehen einer Straftat weitere Taten in unmittelbarer Nähe und innerhalb eines bestimmten Zeitraums nach sich ziehen werde[394]. Dabei werden in der Regel polizeilich ermittelte Daten zugrunde gelegt.[395] SKALA hingegen legt Kriminalitätstheorien zugrunde, die nach Auffassung des nordrhein-westfälischen Landeskriminalamtes geeignet sind, ein möglichst realistisches Bild von Wohnungseinbruchsdiebstählen zu zeichnen.[396] Dazu gehören Rational-Choice-Theorien, wonach Personen, die einen Einbruch planen, vor der Tat eine Kosten-Nutzen-Abwägung vornehmen[397], dass also zum Beispiel die erwartete Höhe des Diebesgutwerts die Wahrscheinlichkeit von Wohnungseinbruchsdiebstählen in einem Gebiet erhöht[398]. Auch KrimPro beruht auf einem kriminologisch komplexeren Ansatz und stützt seine Prognose auf verschiedene Kriminalitätstheorien.[399]

Personenbezogene Verfahren des Predictive Policing[400] stützen die Prognose auf Täter- bzw. Opfermerkmale[401]. Zu nennen ist beispielhaft das in Großbritannien eingesetzte Programm HART[402] sowie die sogenannte Strategic Subject List der Polizei von Chicago[403]. In Deutschland wird seit 2017 vom Bundeskriminalamt das Risikobewertungsinstrument RADAR-iTE eingesetzt. Es ordnet Personen aus dem militant-salafistischen Spektrum auf Basis

---

[393] Gluba, A. (2017): Der Modus Operandi bei Fällen der Near Repeat Victimisation. In: Kriminalistik 71 (6), 369-375, 369.
[394] Trute, H.-H.; Kuhlmann, S. (2021): Predictive Policing als Formen polizeilicher Wissensgenerierung. In: Zeitschrift für das Gesamte Sicherheitsrecht 3, 103-110 (105).
[395] Beck, S. et al. (Hg.) (2020): Digitalisierung, Automatisierung, KI und Recht. Baden-Baden, 519 (522).
[396] Landeskriminalamt NRW (2018): Abschlussbericht Projekt SKALA. Düsseldorf. https://polizei.nrw/sites/default/files/2018-07/180628_Abschlussbericht_SKALA.PDF [17.01.2023], 10 ff.; Beck, S. et al. (Hg.) (2020): Digitalisierung, Automatisierung, KI und Recht. Baden-Baden, 519 (523).
[397] Lüdemann, C.; Ohlemacher,T. (2002): Soziologie der Kriminalität. Theoretische und empirische Perspektiven. Weinheim, München, 54.
[398] Landeskriminalamt NRW (2018) Abschlussbericht Projekt SKALA. Düsseldorf. https://polizei.nrw/sites/default/files/2018-07/180628_Abschlussbericht_SKALA.PDF [17.01.2023], 12.
[399] Knobloch, T. (2018): Vor die Lage kommen: Predictive Policing in Deutschland. Chancen und Gefahren datenanalytischer Prognosetechnik und Empfehlungen für den Einsatz in der Polizeiarbeit. Herausgegeben von der Stiftung Neue Verantwortung e.V. und der Bertelsmann Stiftung. https://www.stiftung-nv.de/sites/default/files/predictive.policing.pdf [10.02.2023], 13; Beck, S. et al. (Hg.) (2020): Digitalisierung, Automatisierung, KI und Recht. Baden-Baden, 524.
[400] Auch Technologien, die im Kontext der Entscheidung über die Strafaussetzung zur Bewährung zum Einsatz kommen (beispielsweise die Software COMPAS) lassen sich als Verfahren des Predictive Policing einstufen, insoweit diese ebenfalls auf die individuelle Gefährlichkeit des Straftäters abstellen. Sofern letzterem nämlich eine negative Sozialprognose erteilt wird, steht dies einer Strafaussetzung zur Bewährung entgegen. Damit ist die Bewährungsentscheidung der Sache nach eine gefahrenabwehrrechtliche. Im deutschen Recht muss gleichwohl berücksichtigt werden, dass die Tätigkeit von Gerichten (die über die Strafaussetzung zur Bewährung entscheiden) klassischerweise von der Polizeiarbeit getrennt wird. Es ist daher weniger üblich, die Bewährungsentscheidung als Polizeiliche Tätigkeit einzuordnen, was zumindest formell gegen die Einstufung als Predictive Policing spricht.
[401] Povalej, R.; Volkmann, D. (2021): Predictive Policing. In: Informatik Spektrum 44, 57-61 (DOI: 10.1007/s00287-021-01332-4), 58. Die Differenzierung zwischen raum- und personenbezogenen Verfahren des Predictive Policing ist alles andere als trennscharf. Überschneidungen ergeben sich bereits daraus, dass raumzeitliche Parameter nicht selten auf die Person des jeweiligen Gefährders schließen lassen. Gleichwohl handelt es sich um eine etablierte Differenzierung, die hier ebenfalls herangezogen wird.
[402] Oswald, M. et al (2018): Algorithmic risk assessment policing models: lessons from the Durham HART model and 'Experimental' proportionality. In: Information & Communications Technology Law, 27 (2), 223-250 (DOI:10.1080/13600834.2018.1458455), 223 ff.
[403] Sommerer, L. (2020): Personenbezogenes Predictive Policing – Kriminalwissenschaftliche Untersuchung über die Automatisierung der Kriminalprognose. Baden-Baden, 80 ff.

einer Verhaltensanalyse als „Gefährder“ oder „Relevante Personen“ auf einer Risikoskala ein.[404] Seit 2017 wird in Hessen das Programm hessenDATA, basierend auf der Gotham-Software des amerikanischen Unternehmens Palantir im Rahmen der Terrorismusbekämpfung auf der Grundlage von § 25a Abs. 1 Alt. 1 des Hessischen Gesetzes über die öffentliche Sicherheit und Ordnung (HSOG) eingesetzt.[405] Es nutzt Informationen aus polizeilichen Datenbanken, Verkehrsdaten aus der Telekommunikationsüberwachung und von Telekommunikationsanbietern zur Verfügung gestellte Daten. Einbezogen werden außerdem sogenannte forensische Extrakte wie zum Beispiel die Ergebnisse der Beschlagnahme eines Mobiltelefons und Informationen aus sozialen Netzwerken.[406] Eine Rechtsgrundlage für personenbezogenes Predictive Policing liefert zudem § 4 Fluggastdatengesetz (FlugDaG). Darin wird das Bundeskriminalamt ermächtigt, Fluggastdaten[407] automatisiert nach bestimmten Mustern abzugleichen. Dies dient der Identifikation von Personen, bei denen ein gewisses Risiko für die Begehung einer terroristischen Straftat oder einer Straftat der schweren Kriminalität besteht.[408] Nachdem das Bundesverfassungsgericht unter anderem die hessische Regelung in § 25a Abs. 1 Alt. 1 HSOG für verfassungswidrig erklärt und eine Neuregelung verlangt hat, weil sie keine ausreichende Eingriffschwelle enthalte,[409] müssen die einschlägigen Ermächtigungsgrundlagen auch in anderen Gesetzen an die vom Bundesverfassungsgericht formulierten Vorgaben angepasst werden.

Der Einsatz digitaler Technologien des Predictive Policing – vor allem im Hinblick auf personenbezogene Verfahren – wird kontrovers diskutiert. Einerseits geht damit die Hoffnung einer Verbesserung der polizeilichen Arbeit und damit der bessere Schutz möglicher Opfer einher (vgl. das Beispiel zum Online-Grooming weiter unten).[410] Andererseits ist algorithmenbasierte

[404] Bundeskriminalamt (2017): Presseinformation: Neues Instrument zur Risikobewertung von potenziellen Gewaltstraftaten. https://www.bka.de/DE/Presse/Listenseite_Pressemitteilungen/2017/Presse2017/170202_Radar.html [17.01.2023].
[405] Sommerer, L. (2020): Personenbezogenes Predictive Policing – Kriminalwissenschaftliche Untersuchung über die Automatisierung der Kriminalprognose. Baden-Baden, 90.
[406] Hessischer Landtag (2019): Drucksache 19/6864. https://starweb.hessen.de/cache/DRS/19/4/06864.pdf [01.03.2023], 18 f.
[407] Fluggastdaten i.S.v. § 2 Abs. 2 FlugDaG sind unter anderem Anschrift und Kontaktangaben des Reisenden (Nr. 5), vollständige Gepäckangaben (Nr. 7), Informationen über nicht angetretene Flüge (Nr. 14) und Angaben über Mitreisende (Nr. 19). Gemäß § 4 Abs. 3 FlugDaG werden die Muster für den Abgleich von der Fluggastdatenzentralstelle des Bundeskriminalamts unter Einbeziehung des Datenschutzbeauftragten erstellt. Allerdings ist nach wie vor ungeklärt, welche Daten in die Mustergewinnung einbezogen werden (Sommerer, L. (2020): Personenbezogenes Predictive Policing – Kriminalwissenschaftliche Untersuchung über die Automatisierung der Kriminalprognose. Baden-Baden). Als Beispiele für Muster, die über Fluggäste gewonnen werden, sind folgende Fälle zu nennen: die Buchung des Fluges erfolgte kurzfristig; es wurde in bar gezahlt; ein bisher nicht alleine reisender Minderjähriger reist nunmehr alleine (vgl. Münch, in: Deutscher Bundestag (2017): Wortprotokoll der 114. Sitzung des Innenausschusses. Protokoll-Nr. 18/114. https://www.bundestag.de/resource/blob/515144/87bdd46c572ebeb4863d139812f3fc4a/Protokoll-114-data.pdf [03.03.2023], 26).
[408] Sommerer, L. (2020): Personenbezogenes Predictive Policing – Kriminalwissenschaftliche Untersuchung über die Automatisierung der Kriminalprognose. Baden-Baden, 96.
[409] Bundesverfassungsgericht (2023): Urteil vom 16. Februar 2023, 1 BvR 1547/19, 1 BvR 2634/20, https://www.bundesverfassungsgericht.de/SharedDocs/Downloads/DE/2023/02/rs20230216_1bvr154719.pdf?__blob=publicationFile&v=3 [20.02.2023].
[410] Ob eine solche tatsächlich gelingt, ist umstritten. Beck, S. et al. (Hg.) (2020): Digitalisierung, Automatisierung, KI und Recht. Baden-Baden, 531 f.

Verbrechensbekämpfung bzw. -verhinderung mit verschiedenen Risiken und Problemen verbunden. Diese betreffen zunächst Fragen der Genauigkeit bzw. Fehlerhaftigkeit von Vorhersagen durch KI-Systeme. Damit verknüpft ist die Frage, ob solche Vorhersagen für verschiedene Personengruppen gleichermaßen zuverlässig funktionieren. Darüber hinaus besteht die Gefahr, dass Personen aufgrund ihrer Herkunft, ihres Wohnorts oder ihrer Vorgeschichte diskriminiert werden.[411] Das tatsächliche Vorliegen einer ungerechtfertigten Ungleichbehandlung hängt maßgeblich von den verwendeten Parametern, deren Gewichtung innerhalb der Prognoseentscheidung, den verfügbaren Daten und deren Charakteristik ab.[412] Von nicht nur technischer, sondern auch ethischer Relevanz sind hierbei auch methodische Entscheidungen, die das Ausmaß und Verhältnis falsch-positiver und falsch-negativer Ergebnisse vorbestimmen. Auch wenn Fehler selbstverständlich auch bei menschlichen Urteilen auftreten, so besteht bei Entscheidungsunterstützung durch Software eine noch größere Gefahr, dass Fehler und Verzerrungen systembedingt besondere Breitenwirkung entfalten. Gerade bei polizeilichen Maßnahmen wirken beide Arten von Fehlern folgenschwer – sowohl wenn fälschlicherweise eine polizeiliche Maßnahme gegen eine Person ergriffen wird als auch wenn ein notwendiger polizeilicher Zugriff auf eine Person unterbleibt. Gesellschaftlich muss also entschieden werden, in welchem Umfang entsprechende Risiken hinnehmbar sind, und auch, ob es einen Unterschied macht, wenn der Fehler auf der Technik oder auf menschlichem Versagen beruht.

Eine weitere Problematik, die Gegenstand intensiver Debatten ist, ist die Frage des Schutzes der Privatsphäre im Kontext von Predictive Policing im Allgemeinen und in Bezug auf Chatkontrollen im Besonderen (siehe Kasten Online-Grooming).[413] Die für die Polizeiarbeit herangezogenen Daten sind in aller Regel besonders sensibel. Bei sogenannten Chatkontrollen zur Prävention und Bekämpfung des sexuellen Missbrauchs von Kindern, zu denen die Kommission der Europäischen Union im Mai 2022 einen Verordnungsvorschlag[414] vorgelegt hat, geht

---

[411] Zu diskriminierenden und stigmatisierenden Effekten: Egbert, S.; Leese, M. (2020): Criminal Futures: Predicitve Policing and Everyday Police Work. London, New York, 186 ff.
[412] Beispielsweise lässt sich darüber streiten, ob es per se diskriminierend ist, im Rahmen einer kriminalrechtlichen Prognose den Wohnort oder den sozio-ökonomischen Status heranzuziehen, wenn diese Umstände zugleich empirisch belegt kriminalitätssteigernde Faktoren darstellen, vgl. hierzu ausführlich – auch mit Blick auf predictive policing – Kischel, in: BeckOK GG, 52. Edition, Stand: 15.08.2022, Art. 3 Rn. 218d.
[413] Singelnstein, T. (2018): Predictive Policing: Algorithmenbasierte Straftatprognosen zur vorausschauenden Kriminalintervention. In: Neue Zeitschrift für Strafrecht 1, 1-8.
[414] Europäische Kommission (2022): Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zur Festlegung von Vorschriften zur Prävention und Bekämpfung des sexuellen Missbrauchs von Kindern. Brüssel. https://eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=CELEX:52022PC0209&from=EN [17.01.2023].

es auch um die Frage, ob eine anlasslose und flächendeckende Überwachung privater Kommunikation gerechtfertigt werden kann, gerade auch in Anbetracht vorhandener Kritik bezüglich der Effektivität solcher Kontrollen für den Schutz der Kinder und der generellen Kritik an Maßnahmen der Vorratsdatenspeicherung. Nach Einschätzung des Bundesdatenschutzbeauftragten stellt der Verordnungsentwurf einen unverhältnismäßig intensiven Eingriff in die Grundrechte dar.[415]

Infokasten 13: Online-Grooming

In den Sozialen Medien und dem Internet sind Kinder und Jugendliche dem Risiko des sogenannten Online-Groomings ausgesetzt, bei dem Erwachsene eine emotionale Beziehung zu Minderjährigen mit dem Ziel des sexuellen Missbrauchs aufbauen.[416] Die Kommission der Europäischen Union hat im Mai 2022 einen Verordnungsentwurf zur Prävention und Bekämpfung des sexuellen Missbrauchs von Kindern vorgelegt. Dieser sieht die systematische Identifizierung entsprechender Anbahnungen in Chats als präventives Mittel vor, um potenzielle Straffällige noch vor der Tat zu entdecken und die Minderjährigen rechtzeitig vor der Sexualstraftat zu schützen.[417] Um Grooming früh und genau zu identifizieren, sollen Chats online bereits mittels automatisierter und spezifisch für diese Problematik trainierter KI-Algorithmen analysiert werden. Die von den Algorithmen zu lösende Aufgabe ist komplex, da die Missbrauchsstrategie sich erst über einen längeren Zeitraum hinweg entfaltet, vom Aufbau von Vertrauen und dem Austausch persönlicher Informationen bis hin zur verharmlosenden Kontextualisierung sexueller Handlungen und schließlich der Vereinbarung persönlicher Treffen.

Die Technologie ist stark umstritten. Aufseiten der Chancen wird insbesondere die Verhinderung von Gewalttaten und sexuellen Übergriffen an Kindern und somit der Schutz einer besonders vulnerablen Bevölkerungsgruppe in den Vordergrund gerückt. In der Kritik stehen hingegen die Risiken einer anlasslosen und flächendeckenden Überwachung privater Kommunikation, und es gibt Zweifel an der Genauigkeit und Unverzerrtheit der Analysen sowie der Effektivität der Maßnahmen in Bezug auf den Schutz von Kindern.

Zu beachten ist auch der Zielkonflikt zwischen der Genauigkeit und der Schnelligkeit einer Vorhersage, dessen ethische Implikationen situationsadäquat abgewogen werden müssen. Sowohl falsch-positive als auch falsch-negative Vorhersagen, bei denen also Personen fälschlicherweise als Straffällige klassifiziert werden bzw. tatsächlich straffällige Personen nicht detektiert werden, können gravierende Auswirkungen für die fälschlich Verdächtigten oder für die fälschlich nicht geschützten Kinder haben. Eine KI-basierte Detektion kann daher zunächst nur als Alarmzeichen dienen, das die Überprüfung der Chatinhalte durch menschliche Sachverständige erzwingt. Im Fall

[415] https://www.bfdi.bund.de/SharedDocs/Pressemitteilungen/DE/2022/09_Chatkontrolle.html?nn=252104 [02.03.2023]. Vgl. auch Entwurf einer Stellungnahme der Bundesregierung vom 17. Februar 2023 zum Vorschlag einer CSA-Verordnung der EU-Kommission (veröffentlicht auf netzpolitik.org: https://netzpolitik.org/2023/positionspapier-innenministerium-macht-wenig-zugestaendnisse-bei-chatkontrolle/#2023-02-17_BMI_Chatkontrolle_Entwurf [02.03.2023].

[416] Wachs, S; Wolf, K. D.; Pan, C.-C. (2012): Cybergrooming: risk factors, coping strategies and associations with cyberbullying. In: Psicothema, 24 (4), 628–633.

[417] Vogt, M., Leser, U., Akbik, A. (2021): Early Detection of Sexual Predators in Chats. In: Zong, C. et al. (Hg.): Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, 4985–4999.

+++ Vorabfassung! Wird durch die lektorierte Fassung ersetzt. +++

einer überprüften Auffälligkeit muss vom Anbieter der Chatplattform sichergestellt werden, dass die Informationen auch an die Strafverfolgungsbehörden gemeldet werden.

Der Erfolg von KI-Algorithmen bei der rechtzeitigen Identifizierung Krimineller hängt auch vom Zugang zu Trainingsdatensätzen und deren Qualität ab, ebenso wie von der Effizienz der Algorithmen bei limitierter Rechenkapazität, da oftmals eine zeitnahe, lokale Analyse auf mobilen Endgeräten (wie etwa Smartphones) erforderlich ist.[418] Es ist zu erwarten, dass Algorithmen fortwährend einer Aktualisierung auf der Grundlage neuer Trainingsdaten bedürfen, wenn diejenigen, die solche Sexualstraftaten planen und begehen, Umgehungsstrategien (wie etwa adaptierte sprachliche Äußerungen) entwickeln.

Personenbezogene Verfahren des Predictive Policing können, wie in anderen Feldern auch, menschliche Beurteilung und Entscheidung unterstützen und dadurch Handlungsmöglichkeiten insbesondere der Strafverfolgungsbehörden erweitern. Andererseits können sie die Handlungsmöglichkeiten der verschiedenen Beteiligten aber auch vermindern oder – im Falle einer vollständigen Übertragung von Entscheidungen an algorithmische Systeme – gar eliminieren. Wenn bei der Verhinderung einer Straftat hohe Anforderungen an Schnelligkeit bestehen, kann es zu einem Konflikt mit dem gleichzeitigen Gebot von möglichst großer Sorgfalt und Heranziehung menschlicher Beurteilung kommen. Nicht zuletzt wird die Sorge geäußert, dass mit algorithmengesteuerter Polizeiarbeit das Risiko der Verfestigung eines mechanischen Menschenbildes einhergehen könne, das den einzelnen Menschen verobjektiviere, seine Individualität auf eine datengetriebene Klassifikation reduziere, jedoch die gesamtgesellschaftlichen Ursachen von Kriminalität unberücksichtigt lasse.[419]

Der Einsatz von algorithmischen Verfahren des Predictive Policing kann die Versuchung bzw. Erwartung einer möglichst weitgehenden, gar lückenlosen Vermeidung von Straftaten im Vorfeld ihrer Begehung wecken. Für eine weiter voranschreitende Durchdringung des Lebensalltags der Menschen mit gefahrenabwehrrechtlichen Maßnahmen sind im Zeitalter der Digitalisierung die technischen Voraussetzungen erfüllt. Immer wieder wird dann auch befürchtet, die freiheitliche Ordnung würde allmählich, mit dem Argument von mehr Sicherheit im Rücken, einem Überwachungsstaat weichen, in dem lückenlose staatliche Überwachung wenig bis keinen Raum für individuelle Freiheit lassen würde. Ist diese Sorge zwar grundsätzlich verständlich, weil der Preis für die erzielte Sicherheit und Gefahrenabwehr zu hoch wäre[420], darf sie

[418] Vogt, M., Leser, U., Akbik, A. (2021): Early Detection of Sexual Predators in Chats. In: Zong, C. et al. (Hg.): Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, 4985–4999.
[419] Zu weiteren Bedenken siehe Egbert, S. (2018): Predicitve Policing in Deutschland. In: ders. (Hg.): Räume der Unfreiheit. Berlin, 241-265, 256.
[420] Rostalski, F. (2019): Brave New World. Vom (Alp-)Traum lückenloser Straftatenahndung in Zeiten der Digitalisierung. In: Goltdammer ´s Archiv für Strafrecht 166 (8), 481-488, 484 f.

jedoch nicht dazu führen, Pauschalurteile gegen Optionen wie das Predictive Policing zu fällen. Wie in diesem Kapitel an vielen Beispielen gezeigt wurde, hängt es vom Design der technischen Systeme und vor allem von ihrer organisatorischen Einbettung in menschliche Entscheidungsverfahren ab, ob die Autorschaft der verschiedenen Beteiligten oder Betroffenen erweitert oder vermindert wird. Statt eines Pauschalurteils bedarf es also der am Ideal menschlicher Urteilsfähigkeit und Verantwortung ausgerichteten Gestaltung der Systeme und ihres Einsatzes in der polizeilichen Praxis.

Dies entspricht dem Bild menschlicher Autorschaft, die prinzipiell dazu in der Lage ist, sich auf der Basis von Einsicht in die Richtigkeit eines bestimmten Verhaltens, beispielsweise zur Wahrung des Körperverletzungsverbots, selbstständig und eigenverantwortlich für die Befolgung des Rechts zu entscheiden. Würde diese Entscheidungsfreiheit in der Folge umfassender staatlicher Sicherungsinstrumente abhandenkommen, würden sich Einzelne nicht länger als verantwortlich für das Funktionieren des freiheitlichen Miteinanders begreifen – mit all den Negativfolgen, die es für ein Gemeinwesen hat, wenn sich dessen Mitglieder selbst nicht mehr in der Pflicht sehen, in der Gestaltung des Rechts und seiner Wahrung eine aktive Rolle einzunehmen.

8.5 Fazit und Empfehlungen

Die Digitalisierung der öffentlichen Verwaltung wird seit über zwanzig Jahren zumeist unter Aspekten von Bürgerfreundlichkeit, Modernität sowie Effizienz und Effektivität thematisiert. KI und dadurch ermöglichte automatisierte Entscheidungsverfahren führen zu neuen Möglichkeiten und Herausforderungen, die erheblich weitreichende ethische und demokratietheoretische Fragen aufwerfen. Denn da staatliches Handeln von vielen Menschen zu einem großen Teil über die öffentliche Verwaltung erfahren wird, beispielsweise in Finanzverwaltung, Meldebehörden oder Sozialwesen, ist mit der Einführung beispielsweise automatisierter Entscheidungsverfahren unmittelbar das Verhältnis von Bürgerschaft und Staat betroffen. Dies gilt etwa in Bezug auf Nachvollziehbarkeit, Erklärbarkeit und Vertrauenswürdigkeit im Verwaltungshandeln, aber auch in Bezug auf Sorgen um Diskriminierung und Technokratie, in der menschliche Kommunikation und Abwägung hinter anonymen Datenmengen und standardisierten Benutzungsoberflächen verschwindet. Eine in diesen Hinsichten als legitim anerkannte Verwaltung ist für ein funktionierendes Gemeinwesen und die Akzeptanz von Demokratie und Staat wesentlich. Die Erfüllung dieser auch für den Einsatz von KI in der öffentlichen Verwaltung geltenden Anforderung hängt mit den durch Automatisierung veränderten Möglichkeiten

menschlicher Autorschaft im Verwaltungshandeln zusammen und führt auf die ethischen Fragen nach ihrer Erweiterung oder Verminderung durch die Einführung KI-gestützter Systeme in diesem Feld.

Seit einigen Jahren kommen in vielen Staaten Systeme zur Entscheidungsunterstützung in der öffentlichen Verwaltung zusehends zum Einsatz, so etwa zur Bewertung von Arbeitsmarktchancen, zur Prüfung und Vergabe von Sozialleistungen und bei Sicherheitsorganen. Aufgrund des dadurch ermöglichten Rückgriffs auf große Datenmengen und ihre zielgerichtete Auswertung in kurzer Zeit können für die erforderlichen Entscheidungen bessere Grundlagen geschaffen und kann die menschliche Autorschaft unterstützt werden. Insofern also die letztliche Entscheidung, etwa über die Bemessung von Sozialleistungen oder die Rückfallwahrscheinlichkeit einer straffälligen Person, bei den Personen, die für diese Entscheidung zuständig sind, verbleibt, ist der Einsatz der ADM-Systeme in ethischer Hinsicht grundsätzlich zu begrüßen. Entscheidend in ethischer Hinsicht ist erstens, dass die Zwecke ihrer Einführung in Übereinstimmung mit dem Ziel der Erweiterung menschlicher Autorschaft stehen und nicht etwa bloßer Effizienzsteigerung des behördlichen Funktionierens oder der Einsparung von Personal dienen.

Jedoch kann es durch Gewöhnung, Überlastung und Entscheidungsdruck zu einer allmählichen Verschiebung kommen, in deren Verlauf die Ergebnisse der ADM-Systeme von den Menschen, die eine Entscheidung treffen, zunehmend ohne weitere Prüfung und Reflexion übernommen würden (Automation Bias). Auf diesem Weg würde menschliche Autorschaft allmählich und fast unmerklich verschwinden. Zurückbleiben würde ein automatisiertes Geschehen, in dem technische Systeme für Betroffene weitreichende, teils existenzielle Festlegungen treffen. Dies ist nicht bloß eine abstrakte Befürchtung, sondern spiegelt Erfahrungen im alltäglichen Verwaltungshandeln wider. Es kommt also zu Konflikten zwischen Erwartungen an die Erweiterung menschlicher Autorschaft einerseits und der realen Umsetzung, in der ihre Verminderung oder gar Ersetzung nicht fernliegen. Ein empirisches Monitoring der realen Adaptation von ADM-Systemen und ihren Folgen ist notwendig, um Fehlentwicklungen in ethischer Hinsicht frühzeitig erkennen und Gegenmaßnahmen einleiten zu können.

Ferner ist an die Forderung nach Diskriminierungsfreiheit des öffentlichen Verwaltungshandelns und seiner möglichst weitgehenden Umsetzung zu erinnern. Automatisiertes Entscheiden, etwa im Sozialwesen oder im Predictive Policing, kann durch die verwendeten Datensätze und

zur Auswertung herangezogene Algorithmen diskriminierend für bestimmte Bevölkerungsgruppen wirken. Dieses mittlerweile gut bekannte Risiko[421] stellt kein Pauschalargument gegen ADM-Systeme in der öffentlichen Verwaltung dar, zumal auch menschliche Entscheider vor Diskriminierung nicht gefeit sind, sondern hat zur Folge, dass sowohl das Design der Systeme als auch die Trainingsdaten und dann auch ihre realen Operationen daraufhin kritisch beobachtet werden müssen.

Schließlich ist festzuhalten, dass die von „Entscheidungen" der ADM-Systeme Betroffenen, etwa durch Entscheidungen des Jugendamtes oder in der Steuerbemessung, die gleichen Rechte etwa auf Erläuterung und Widerspruch haben wie in Institutionen ohne diese Systeme.

Pauschale Aussagen für oder gegen KI bzw. ADM-Systeme in der öffentlichen Verwaltung sind also nicht sinnvoll. Es muss kontextbezogen im Detail eingeschätzt und abgewogen werden, welche Auswirkungen eine entsprechende Maßnahme auf die Autorschaft unterschiedlichster Beteiligter und Betroffener hat, welche Konflikte auftreten und wie mit ihnen umgegangen werden kann oder soll.

Empfehlungen

- *Empfehlung Verwaltung 1:* Die mit automatisierten Entscheidungshilfen (ADM-Systemen) einhergehende verstärkte Standardisierung und pauschale Kategorisierung von Einzelfällen muss umso stärker hinterfragt und um spezifisch einzelfallbezogene Erwägungen ergänzt werden, je intensiver die betroffene Entscheidung individuelle Rechtspositionen berührt.
- *Empfehlung Verwaltung 2:* Es müssen geeignete technische und organisatorische Instrumente zur Vorkehrung gegen die manifeste Gefahr eines Automation Bias bereitgestellt werden, die es den Fachkräften erschweren, selbst bei einer Letztentscheidungskompetenz der algorithmischen Entscheidungsempfehlung unbesehen zu folgen. Es ist zu prüfen, ob eine Umkehrung der Begründungspflicht (nicht eine Abweichung, sondern ein Befolgen ist zu rechtfertigen) hier eine geeignete Vorkehrung sein kann.
- *Empfehlung Verwaltung 3:* Aufgrund ihrer Grundrechtsbindung sind an staatliche Einrichtungen bei der Entwicklung und Nutzung algorithmischer Systeme hohe Anforderungen in

[421] Orwat, C. (2019): Diskriminierungsrisiken durch Verwendung von Algorithmen. Herausgegeben von der Antidiskriminierungsstelle des Bundes. Berlin. https://www.antidiskriminierungsstelle.de/SharedDocs/downloads/DE/publikationen/Expertisen/studie_diskriminierungsrisiken_durch_verwendung_von_algorithmen.pdf?blob=publicationFile&v=3 [31.01.2023].

Bezug auf Transparenz und Nachvollziehbarkeit zu stellen, um den Schutz vor Diskriminierung zu gewährleisten sowie Begründungspflichten erfüllen zu können.

- *Empfehlung Verwaltung 4:* Für Softwaresysteme in der Öffentlichen Verwaltung müssen Qualitätskriterien verbindlich und transparent festgelegt werden (in Bezug auf Genauigkeit, Fehlervermeidung, Unverzerrtheit und so weiter). Ebenso bedarf es einer Dokumentation der jeweils eingesetzten Methoden. Diesbezüglich sollten auch aktuelle Beschaffungspraktiken, in deren Verlauf staatliche Behörden Softwarelösungen kaufen, einer kritischen Prüfung unterzogen werden.
- *Empfehlung Verwaltung 5:* Überall dort, wo algorithmische Systeme Einsatz in der Öffentlichen Verwaltung finden, gilt es, Sorge zu tragen, dass die Personen, die diese Systeme anwenden, über die erforderlichen Kompetenzen im Umgang damit verfügen. Dazu gehört neben Kenntnis der Verwendungsweisen auch das Wissen um die Limitationen und möglichen Verzerrungen, um Systeme angemessen einsetzen zu können.
- *Empfehlung Verwaltung 6:* Die Einsichts- und Einspruchsrechte Betroffener müssen auch beim Einsatz algorithmischer Systeme effektiv gewährleistet werden. Dazu bedarf es gegebenenfalls weiterer wirksamer Verfahren und Institutionen.
- *Empfehlung Verwaltung 7:* In Öffentlichkeit, Politik und Verwaltung sollte eine Sensibilisierung gegenüber möglichen Gefahren von Automatisierungssystemen, wie etwa Verletzungen der Privatsphäre oder Formen systematisierter Diskriminierung, erfolgen. Dazu gehört eine öffentliche Debatte darüber, ob es in bestimmten Kontexten überhaupt einer technischen Lösung bedarf.
- *Empfehlung Verwaltung 8:* Im Bereich des Sozialwesens ist sicherzustellen, dass ADM-Systeme elementare fachliche Standards von sozialprofessionellen Interaktionen (z. B. gemeinsame Sozialdiagnose oder Hilfeplanung als *Teil* therapeutischer bzw. unterstützender Hilfeleistung) nicht unterlaufen oder verdrängen. Dies beinhaltet insbesondere Maßnahmen, die Vergröberungen individueller Fallkonstellationen und -prognosen durch die ADM-induzierte grobklassifizierende Einteilung von Fall- und/oder Leistungsberechtigtengruppen verhindern. Dabei ist Sorge zu tragen, dass die Feststellung individueller Hilfebedarfe nicht erschwert wird und es zu keiner schleichenden Aushöhlung der sozialrechtlich gebotenen Identifizierung individueller Hilfebedarfe zugunsten einseitiger externer Interessen an Gefahrenminimierung oder Kostenersparnis kommt.

- *Empfehlung Verwaltung 9:* Die Arbeit von Gefahrenabwehrbehörden einschließlich der Polizei betrifft besonders grundrechtssensible Bereiche. Dies wirkt sich auf die Reichweite eines zulässigen Einsatzes von algorithmischen Systemen in der prädiktiven Polizeiarbeit aus. Risiken wie Verletzungen der Privatsphäre oder potenziell unzulässige Diskriminierungen der von dem Einsatz betroffenen Personen müssen mit Chancen auf erhebliche Verbesserungen der staatlichen Gefahrenabwehr sorgfältig abgewogen und in ein angemessenes Verhältnis gebracht werden. Hierfür erforderliche gesellschaftliche Aushandlungsprozesse sollten umfangreich geführt werden. Dabei ist der diffizilen Bestimmung des Verhältnisses von Freiheit und Sicherheit Rechnung zu tragen. Jegliche Gesetzesübertretung zu verhindern, wäre mit rechtsstaatlichen Mitteln nicht möglich.

+++ Vorabfassung! Wird durch die lektorierte Fassung ersetzt. +++

TEIL III: QUERSCHNITTSTHEMEN UND ÜBERGREIFENDE EMPFEHLUNGEN

+++ Vorabfassung! Wird durch die lektorierte Fassung ersetzt. +++

9 Zusammenfassung der bisherigen Analyse

9.1 Anthropologische und ethische Grundorientierung

Der Begriff der Künstlichen Intelligenz hat in der öffentlichen Debatte zunehmend an Aufmerksamkeit gewonnen und wird mit teils überzogenen Hoffnungen aber auch mit teilweise fehlgeleiteten Befürchtungen verknüpft. Insbesondere in den Debatten rund um die sogenannte starke KI wird zuweilen sogar nahegelegt, maschinelle Akteure könnten normativ menschlichen Akteuren ähnlich oder gleichgestellt sein – was etwa zu Diskussionen um „Menschenrechte für Roboter“ führt.

Der Deutsche Ethikrat geht in seiner Stellungnahme hingegen von einem normativ grundlegenden Unterschied zwischen Mensch und Maschine aus.[422] Softwaresysteme, auch solche, die der KI zugerechnet werden, verfügen weder über theoretische noch über praktische Vernunft. Sie können keine Verantwortung für ihr Handeln übernehmen, sie sind kein personales Gegenüber, auch dann nicht, wenn sie Anteilnahme, Kooperationsbereitschaft oder Einsichtsfähigkeit simulieren.

Die oben angesprochenen anthropomorphen Missverständnisse sind möglicherweise Folge der Rede von Künstlicher Intelligenz. Der Begriff Intelligenz wird bislang ganz überwiegend für die Beschreibung menschlicher kognitiver Leistungen verwendet. Der Begriff der menschlichen Intelligenz verweist auf ein Ensemble von Leistungen, die eine Reihe von Primärfaktoren umfassen, wie induktives Schließen, räumliches Vorstellungsvermögen, Wahrnehmungsgeschwindigkeit, Rechenfähigkeit, verbales Verständnis, assoziatives Gedächtnis und Wortflüssigkeit – und die sich, jedenfalls auf dem heutigen Stand der Softwareentwicklung, nicht alle auf KI oder zumindest nicht vollständig auf diese übertragen lassen. Dies gilt erst recht für Konzepte wie die der sozialen oder emotionalen Intelligenz. Insofern enthält der Ausdruck „Künstliche Intelligenz“ eine metaphorische Verwendung von „Intelligenz“, die das Missverständnis einer wenigstens weitgehenden Ähnlichkeit oder sogar Gleichheit menschlicher und Künstlicher Intelligenz fördert. Entsprechend wäre es in vielen Fällen hilfreicher, den Begriff der Künstlichen Intelligenz nicht inflationär zu verwenden, sondern genauer zu spezifizieren,

[422] Diese Feststellung lässt zwei Deutungen zu: Sie kann einmal im Sinne eines kategorialen Unterschieds verstanden werden, der auch durch weiteren technischen Fortschritt nicht überwunden werden kann; dieser Lesart schließt sich die Mehrheit des Deutschen Ethikrates an. Sie kann aber auch lediglich im Sinne einer Feststellung hinsichtlich des gegenwärtigen und allenfalls für die nahe Zukunft überblickbaren technischen Entwicklungsstandes verstanden werden; dieser Lesart schließt sich eine Minderheit des Deutschen Ethikrates an.

was eine Software ist, beispielsweise eine Software zur Entscheidungsunterstützung, oder was sie tut, beispielsweise basierend auf statistischen Analysen diverser Daten Prognosen erstellen.

Noch deutlicher wird ein kategorialer Unterschied zwischen Mensch und Maschine, wenn man zum Begriff der Vernunft übergeht. Die *theoretische Vernunft* richtet sich auf den Erkenntnisgewinn, um zu wahren empirischen oder apriorischen Urteilen zu gelangen, während die *praktische Vernunft* auf ein kohärentes, verantwortliches Handeln abzielt, um ein gutes Leben zu ermöglichen. Dabei ist menschliche Vernunft als verleiblicht zu begreifen. Das heißt, die Vernunft des Menschen ist nicht hinreichend beschrieben, wenn sie als abstrakte allgemeine Menschenvernunft aufgefasst wird. Sie ist immer zugleich eingebunden in die konkrete soziale Mit- und Umwelt. Nur so ist zu erklären, dass sie handlungswirksam wird. Als solche bestimmt sie die Sozialität und Kulturalität des Menschen. „Vernünftig" handelt der einzelne Mensch als Teil einer sozialen Mitwelt und einer kulturellen Umgebung. Schon deshalb kann den in dieser Stellungnahme besprochenen Softwaresystemen weder theoretische noch praktische Vernunft zugeschrieben werden (siehe ausführliche Darstellung und Begründung in Kapitel 3).

Menschen entwickeln digitale Technik und nutzen sie als Mittel zu menschlichen Zwecken. Jedoch wirken diese Technologien zurück auf menschliche Handlungsmöglichkeiten: Dadurch können sich einerseits neue Optionen eröffnen und im günstigen Fall Freiheitsgrade vergrößern. Andererseits kann aber auch die Handlungsfähigkeit von Menschen eingeschränkt und können Anpassungen erforderlich werden, die nicht wünschenswert sind. Bei der Verwendung hoch entwickelter vernetzter Softwaresysteme haben Menschen teilweise eine Subjekt-, teilweise aber auch eine Objekt-Rolle inne. Es kann dabei bewusst auf den Subjekt-Status verzichtet werden, um Handlungsmöglichkeiten zu erweitern oder Routine-Aufgaben an Maschinen zu delegieren. Diese instrumentelle Funktion ist jedoch von der Selbstzwecklichkeit menschlicher Akteure zu unterscheiden. Die Fähigkeit, Urheber eigener Handlungen zu sein, die die Grundlage autonomer personaler Praxis ist, bleibt Menschen vorbehalten.

Menschen wirken zweckgerichtet und kontrolliert auf sich selbst und ihre Umwelt ein, um damit bestimmte Veränderungen zu verursachen. Dabei spielt die Möglichkeit, zwischen unterschiedlichen Handlungsoptionen wählen zu können, eine wesentliche Rolle. Auch hochentwickelte Softwaresysteme oder KI-gestützte Roboter sind nicht in der Lage, in diesem anspruchsvollen Sinne zu „handeln". Algorithmische Systeme, die etwa bei der Auswahl von geeigneten Personen für eine Stelle eingesetzt werden, handeln oder entscheiden nicht selbst und können keine Verantwortung übernehmen. Allerdings beeinflussen sie, mitunter in hohem

Maße, die Bedingungen, Möglichkeiten und Grenzen für menschliches Handeln und menschliche Verantwortungsübernahme. In den vorangegangenen Kapiteln wurde dies für die Anwendungsbereiche der Medizin, der schulischen Bildung, der öffentlichen Kommunikation und Meinungsbildung sowie der öffentlichen Verwaltung exemplarisch aufgefächert. Entscheidungsunterstützungssysteme beeinflussen menschliche Handlungsmöglichkeiten in Medizin und Verwaltung ebenso wie KI-gestützte Lernsoftwaresysteme dies im Bildungsbereich tun oder Suchmaschinen und Empfehlungssoftware im Bereich der öffentlichen Meinungsbildung. Auch wenn Maschinen also nicht selbst handeln, so verändern sie die Handlungsfähigkeit von Menschen tiefgreifend und können Handlungsmöglichkeiten erheblich erweitern oder vermindern (vgl. Kapitel 4).

Ziel der Delegation von Handlungssegmenten an Maschinen sollte prinzipiell die *Erweiterung* menschlicher Handlungsfähigkeit und Autorschaft sein. Umgekehrt gilt es, die *Verminderung* menschlicher Handlungsfähigkeit und Autorschaft sowie eine *Diffusion* oder *Evasion von Verantwortung* zu verhindern. Entsprechend ist erforderlich, dass die Übertragung menschlicher Tätigkeiten auf KI-Systeme gegenüber den Betroffenen transparent erfolgt. Es darf nicht zu einer Verantwortungsdiffusion kommen, die dazu führt, dass niemand mehr für Fehlentscheidungen verantwortlich ist – etwa, weil wichtige Entscheidungselemente, -parameter oder -bedingungen nicht mehr nachvollziehbar sind.

Eine zentrale Erkenntnis der hier vorgelegten Analysen lautet, dass Ausmaß und Art sinnvoller und guter Delegation nicht allgemein festgelegt werden können, sondern jeweils kontextspezifisch bestimmt werden müssen. So kann die Verminderung menschlicher Handlungsfähigkeit in bestimmten Kontexten durchaus Vorteile mit sich bringen, wenn beispielsweise durch den Verlass auf teilautomatisierte Diagnostikverfahren in der Medizin Fehldiagnosen reduziert werden. Um über Wert und Nutzen der Delegation vormals menschlichen Handelns an Maschinen zu befinden, müssen allerdings auch die langfristigen Auswirkungen dieser Delegation im Sinne einer Erweiterung oder Verminderung menschlicher Handlungsfähigkeit berücksichtigt werden. So kann sich der Verlust menschlicher Expertise, der entsteht, wenn sich bei ärztlichen Fachkräften durch den Einsatz von algorithmischen Systemen die eigene Fähigkeit zur Diagnose verringert, langfristig als Nachteil erweisen, zum Beispiel in der Ausbildung oder in Situationen, wo die entsprechende Technik nicht (mehr) zur Verfügung steht.

Hier wird die soziale Dimension des Verhältnisses von Delegieren, Erweitern und Vermindern deutlich: Menschen können durch Prozesse des Delegierens bzw. Ersetzens unterschiedlich betroffen sein. Man denke hier etwa an den Einsatz von Entscheidungsunterstützungssystemen in

der Verwaltung, der Handlungsmöglichkeiten von Fachkräften und unterschiedlichen Betroffenen auf sehr verschiedene Weise erweitern oder vermindern kann. Auch im schulischen Kontext betrifft das Delegieren von Handlungen an Maschinen Lernende anders als Lehrkräfte oder gegebenenfalls noch weitere Personen. Beide Kontexte sind von asymmetrischen Beziehungs- und Machtverhältnissen gekennzeichnet, denen bei der Bewertung der Effekte des Einsatzes von Technologien Rechnung getragen werden muss. Dies gilt umso mehr, wenn bestimmte Personengruppen aufgrund ihrer Situierung besonders vulnerabel sind, beispielsweise Kinder im Schulkontext oder gar im Falle einer Wohlgefährdung. Bei der Beurteilung technologischer Entwicklungen gilt es also, den Blick nicht nur auf, diejenigen, die Technik verwenden und direkt davon Betroffene, sondern auch auf indirekt Betroffene zu richten. Darüber hinaus gilt es, die Rolle von weiteren Beteiligten zu beleuchten, die eher im Hintergrund agieren und entweder in besonderer Weise von Einführung und Nutzung der Technologien profitieren oder durch ihre Entscheidungen zur Gestaltung oder zum Einsatz von Technik Möglichkeiten und Risiken für andere Akteure präkonfigurieren. Man denke hier sowohl an die großen Plattformen und Softwareanbieter, aber auch an Data Broker, die im Hintergrund Daten zu verschiedensten, möglicherweise auch sachfremden Zwecken verarbeiten oder verkaufen.

Die Herausforderungen stecken also wie so oft im Detail, genauer: in den Details der Technik, der Einsatzkontexte sowie der institutionellen und sozio-technischen Umgebung. Neben den allgemeinen Forderungen nach Transparenz und Nachvollziehbarkeit, dem Schutz der Privatsphäre und der Minimierung von Diskriminierung und systematischen Verzerrungen (Bias) müssen das ethische Design und der ethische Einsatz von Technologien also immer auch die konkreten institutionellen und organisationalen Rahmenbedingungen sowie individuelle Erfordernisse unterschiedlicher Akteure in den Blick nehmen. Im Folgenden werden die Erkenntnisse aus den Kapiteln 5 bis 8 zusammengefasst und in Kapitel 10 dann übergreifende Querschnittsthemen und Empfehlungen extrahiert. Die detaillierteren, sektorspezifischen Empfehlungen befinden sich jeweils am Ende der Kapitel 5 bis 8.

9.2 Einsichten aus den Anwendungsfeldern

Die Verwendung von Technologien verändert und erweitert die Handlungsmöglichkeiten von Menschen seit Jahrtausenden. Der Einsatz von Technologien hat dabei nicht nur intendierte Folgen im Nahbereich, sondern wirkt im Positiven wie im Negativen auch in großer räumlicher und zeitlicher Distanz. Ähnlich verhält es sich mit dem Einsatz von Künstlicher Intelligenz. Die Delegation von Entscheidungen, Handlungen und Handlungssegmenten an Software im Allgemeinen und KI im Besonderen schafft neue Möglichkeiten, birgt aber auch zahlreiche Risiken.

In dieser Stellungnahme hat sich der Deutsche Ethikrat exemplarisch mit einigen Anwendungen in der Medizin, der schulischen Bildung, der öffentlichen Kommunikation und der öffentlichen Verwaltung beschäftigt. Dieser Auswahl liegt die Überzeugung zugrunde, dass erst ein Blick in die jeweiligen Anwendungskontexte die besonderen Chancen und Risiken des Einsatzes von KI zutage fördert. Es wurden bewusst Sektoren ausgewählt, in denen die Durchdringung durch KI-basierte Technologien sehr unterschiedlich ausfällt. Am einen Ende des Spektrums befindet sich der Bereich der Sozialen Medien mit seinen Auswirkungen auf öffentliche Kommunikation und Meinungsbildung. Hier sind alle, die diese Angebote nutzen, bereits jetzt in ihrem Alltag auf vielfältige Weise bei ihrer Informationsauswahl von den Auswirkungen der dort verwendeten Technologien betroffen. Am anderen Ende befindet sich der Bereich der schulischen Bildung, in dem der Einsatz von KI gegenwärtig, jedenfalls in Deutschland, noch eher die Ausnahme darstellen dürfte.

Es wurden in allen Sektoren Beispiele gewählt, die unterschiedliche Ausmaße des Ersetzens vormals menschlicher Handlungen durch KI veranschaulichen. In allen vier Sektoren gibt es Einsatzszenarien, die durch teils erhebliche Beziehungs- und Machtasymmetrien gekennzeichnet sind, was einen verantwortungsvollen Einsatz von KI und die Berücksichtigung der Interessen und des Wohls insbesondere vulnerabler Personengruppen umso wichtiger macht. Diese Unterschiedlichkeit der Art und Weise des KI-Einsatzes sowie des Ausmaßes der Delegation an Maschinen in den Blick zu nehmen, erlaubt es, nuancierte ethische Betrachtungen anzustellen.

Im besonders sensiblen Bereich der *Medizin* (vgl. Kapitel 5) kommen digitale Produkte mit KI-Komponenten in rasch wachsendem Umfang zum Einsatz. Dabei sind unterschiedliche Akteursgruppen zu unterscheiden, die vom KI-Einsatz unterschiedlich betroffen sind oder verschiedene Verantwortlichkeiten besitzen: diejenigen, die KI-Instrumente entwickeln, klinisch tätige Personen, Patientinnen und Patienten. Das Spektrum der Delegation an KI-Systeme reicht hier von der punktuellen Unterstützung ärztlichen Handelns über dessen ständige Begleitung bis hin zur weitgehenden oder gar vollständigen Ersetzung ärztlicher Fachkräfte durch ein KI-System. Um einen verantwortlichen Einsatz von KI-Systemen zu gewährleisten, ist es daher erforderlich, die gesamte Praxis von der Entwicklung der KI-Systeme über ihren Einsatz in der Forschung bis zur Implementierung in der medizinischen Versorgung an ethischen Standards auszurichten.

Ein wichtiges Einsatzfeld für KI-Systeme ist die Unterstützung ärztlicher Entscheidungen in der Diagnostik, da diese große Datenmengen und eine Vielzahl relevanter Parameter auswerten

und in Gestalt präziser Mustererkennung großes diagnostisches und therapeutisches Potenzial entwickeln können. Der verbesserten Diagnostik stehen allerdings auch Risiken gegenüber, etwa der Verlust eigener Diagnosekompetenz aufseiten des ärztlichen Fachpersonals im Vertrauen auf technische Systeme und die Vernachlässigung eigener Urteilskraft. Auch der Aufklärungsbedarf aufseiten der Patientinnen und Patienten steigt durch den Einsatz von KI-Instrumenten und muss berücksichtigt werden, um das Vertrauen und die personale Zuwendung im Arzt-Patienten-Verhältnis nicht zu beschädigen. Besonders ambivalent ist etwa die Ersetzung psychotherapeutischer Diagnose oder Therapie durch Chatbots. Einerseits erhofft man sich davon einen Beitrag zur Entlastung des Gesundheitssektors und einen erleichterten Zugang zur psychotherapeutischen Erstversorgung. Andererseits stellen sich zahlreiche Probleme, nicht zuletzt die Gefahr einer Personalisierung von Maschinen, die falsche Projektionen begünstigen kann.

Auch im Bereich der *schulischen Bildung* (vgl. Kapitel 6) kommen KI-gestützte Technologien zum Einsatz, insbesondere um personalisiertes Lernen und Lehren zu unterstützen oder Informationen zum Lerngeschehen bereitzustellen. Die maschinelle Ersetzung bestimmter pädagogischer Handlungssegmente kann auch hier sowohl zur Erweiterung, aber auch zur Verminderung der Handlungsoptionen von Lernenden und Lehrkräften führen. Die Vision einer vollständigen Ersetzung des Lehrpersonals durch Maschinen ist jedoch mit dem interpersonalen Charakter des pädagogischen Geschehens und dem Fokus auf Persönlichkeitsbildung nicht in Einklang zu bringen. Auch dürfen durch den Einsatz digitaler Tools weder die ohnehin oft dominierenden Vorstellungen einer technisch verstandenen Optimierung von Lernprozessen verstärkt noch der Blick auf das (Ab-)Prüfbare verengt werden. Im Mittelpunkt der Bildungspraxis muss vielmehr weiterhin die Herausbildung von mündigen und freien Personen stehen, die urteilsfähig und verantwortlich handeln können. Der Einsatz digitaler Systeme muss diesen Zielen verpflichtet sein und darf nicht lediglich Selbstzweck sein. Er kann dabei je nach Ziel, Gestaltung und technischen Grundlagen sehr unterschiedlich ausfallen. So kann er etwa darauf gerichtet sein, Wissen zu vermitteln oder Fähigkeiten zu trainieren, Feedback zum Lernfortschritt zu geben oder Impulse für das pädagogische Geschehen zu geben. Dabei sollte verstärkt auf das Inklusionsziel allgemeiner Bildung geachtet werden: KI-Systeme können Schülerinnen und Schüler mit besonderen Bedürfnissen und deren individuellen Lerngeschichten adaptiv begleiten, sie unterstützen, ihre spezifischen pädagogischen Bedürfnisse zu artikulieren, oder neue Möglichkeiten eröffnen, im Krankheitsfall aus der Ferne immersiv am Unterrichtsgeschehen teilzunehmen. Das Unterrichtsgeschehen als Praxis der Verständigung, des Austausches von Gründen, der Stärkung der Urteilskraft und der Persönlichkeitsentwicklung sollte durch den

Einsatz digitaler Tools erweitert und nicht vermindert werden. Gleichwohl können den Lehrkräften wie den Lernenden dadurch neue Handlungsoptionen eröffnet und der personalisierte Austausch vertieft werden.

Prozesse der *öffentlichen Kommunikation und Meinungsbildung* (vgl. Kapitel 7) sind in zunehmendem Maße durch Software im Allgemeinen und KI im Speziellen geprägt. Auf der einen Seite erweitern Soziale Medien mit ihren vielfältigen und wirkmächtigen sozio-technischen Prozessen der Generierung, Kuratierung und Sortierung von Inhalten menschliche Handlungsmöglichkeiten, etwa durch den Zugang zu einer zuvor ungekannten Menge und Diversität verfügbarer Informationen und Kommunikationsmöglichkeiten. Dem gegenüber steht jedoch auch die – häufig weniger sichtbare – Gefahr der Verminderung menschlicher Handlungsmacht. Suchmaschinen, Newsfeeds und Empfehlungssoftware steuern beabsichtigt oder unbeabsichtigt menschliches Verhalten, beeinflussen unseren Zugang zu Information und Kommunikation und somit zur Realität. Solche möglichen Einbußen informationeller Selbstbestimmung sind das Resultat der zuvor geschilderten sozio-technischen und häufig opaken Prozesse. Diese Prozesse sind primär auch nicht an den Interessen der Personen, die sie zur Information und Kommunikation nutzen, ausgerichtet, sondern an denen der Plattformen selbst sowie der dort aktiven Werbetreibenden.

Von den hier behandelten Sektoren ist der Bereich von Information und Kommunikation derjenige, der am stärksten von KI durchdrungen ist und in dem sich Möglichkeiten und Risiken bereits entsprechend deutlich abzeichnen. Viele Anwendungen wären ohne den Einsatz algorithmischer Systeme, die Informationen sortieren, nicht möglich. Diese starke Durchdringung unserer Lebenswelt mit digitalen Technologien führt jedoch zu einer individuellen und kollektiven Abhängigkeit und schafft Sachzwänge, welche die Vulnerabilität der Gesellschaft erhöhen.

Der vierte Sektor, der in dieser Stellungnahme exemplarisch behandelt wurde, ist die *öffentliche Verwaltung* (vgl. Kapitel 8). Hier lässt sich in den vergangenen Jahren in vielen Ländern ein zunehmender Einsatz von Software zur Entscheidungsunterstützung in einer Vielzahl unterschiedlicher Bereiche beobachten, beispielsweise bei der Bewertung der Arbeitsmarktchancen Jobsuchender, der Prüfung und Vergabe von Sozialleistungen, bei Vorhersagen im Sozialwesen oder auch in diversen polizeilichen Einsatzkontexten. Auch wenn in diesen Fällen Entscheidungen nach wie vor von Menschen getroffen werden, so präkonfigurieren maschinelle Diagnosen und Prognosen diese doch in zunehmendem Ausmaß.

Häufig wird der Einsatz solcher Systeme einerseits mit einer Steigerung der Effizienz von Verwaltungsvorgängen begründet und andererseits mit einer möglichen Verbesserung der Qualität von – teils sehr tiefgreifenden – Entscheidungen, insbesondere bei komplexen Datenlagen. Demgegenüber stehen zahlreiche ethische Fragen und Probleme. Es ist nicht erwiesen, dass die Verwendung von Entscheidungssystemen zwangsläufig zu besseren Entscheidungen führt. Zudem konnte in zahlreichen Studien gezeigt werden, dass insbesondere datenbasierte Systeme existierende gesellschaftliche Ungleichheiten oftmals reproduzieren. Durch den Einbau in scheinbar neutrale Entscheidungssysteme wird daher Diskriminierung zum einen perpetuiert und zum anderen unsichtbar gemacht. Gerade in Bezug auf Entscheidungen, die eine hohe Tragweite haben oder besonders vulnerable Gruppen betreffen, ist daher Sorge zu tragen, dass diese Systeme nicht nur genau sind, sondern auch offengelegt wird, mittels welcher Methoden mögliche Diskriminierungseffekte geprüft mitigiert werden.

Umgekehrt kann der Einsatz von KI auch existierende Diskriminierungen in menschlicher Entscheidungspraxis identifizieren: Mit der Analyse früherer Entscheidungen wird es möglich, gesellschaftliche Diskriminierung aufzudecken, zu belegen und Änderungen einzufordern. Die Verbesserung von Entscheidungen, die Minimierung von Fehlern und das Verhindern bzw. Ausgleichen existierender Diskriminierung sind wichtige Ziele für den Einsatz von algorithmischen Systemen in der öffentlichen Verwaltung. Damit diese die Handlungsmöglichkeiten der verschiedenen Beteiligten jedoch erweitern und nicht vermindern, ist es entscheidend, dass Qualitätsstandards verbindlich festgelegt werden und deren Überprüfung durch externe Akteure möglich ist. Gerade in staatlichen Kontexten müssen hier höchste Anforderungen an Transparenz und Diskriminierungsminimierung angelegt werden.

10 Entfaltung von Querschnittsthemen und Empfehlungen

Die Darstellung der sozio-technischen Entwicklungen und deren ethische Analyse in den vier Anwendungskontexten zeigen, dass es eine Reihe von Querschnittsthemen und -herausforderungen gibt, die sich – teils in unterschiedlichen Ausprägungsgraden und Formvarianten – durch alle vier Bereiche ziehen. Sie dürften auch in anderen Anwendungskontexten von „intelligenten“ Maschinen eine Rolle spielen, die in dieser Stellungnahme nicht behandelt wurden. Um im Hinblick auf die Erweiterung menschlicher Handlungsfähigkeit und Autorschaft zukünftig einen guten gesellschaftlichen Umgang und entsprechende Gestaltung zu gewährleisten, können solche Querschnittsfragen nicht nur innerhalb der einzelnen Bereiche – sektoral – angegangen werden. Es werden darüber hinaus vernetzte, bereichsübergreifende Ansätze notwendig, um den im Folgenden skizzierten Themen und Fragen adäquat gerecht zu werden, die gleichzeitig ausreichende Feinkörnigkeit erlauben, um der ebenso erforderlichen Kontextsensitivität zu entsprechen. Solches gleichermaßen horizontale wie vertikale gestaltende Denken stellt eine Herausforderung insbesondere für die Politikgestaltung und die etwaige zukünftige Regulierung dar. Während dies bei einzelnen Themen bereits erkannt und – wenngleich nicht immer hinreichend – umgesetzt ist, beispielsweise in Bezug auf Probleme und Lösungen hinsichtlich eines verantwortlichen Umgangs mit Daten, bestehen mit Blick auf andere Themen noch erhebliche Defizite, beispielsweise in Bezug auf (Pfad-)Abhängigkeiten und Resilienz technologischer Infrastrukturen. Die folgende Darstellung von Querschnittsthemen und die Empfehlungen sollen daher als Anregung für eine breitere Debatte dienen, wie für zukünftige Politik- und Technikgestaltung gleichzeitig und im Zusammenspiel mit sektoralen Aspekten immer auch übergreifende Fragen in den Blick genommen werden können und müssen.

10.1 Querschnittsthema 1: Erweiterung & Verminderung von Handlungsmöglichkeiten

Obwohl KI-Anwendungen in allen in dieser Stellungnahme erwähnten Bereichen neue Handlungsoptionen erschließen können, fallen die Potenziale für die tatsächliche Erweiterung menschliche Handlungsmöglichkeiten bislang sehr unterschiedlich aus – nicht nur zwischen den einzelnen Feldern, sondern auch innerhalb verschiedener Segmente ein und desselben Bereichs.

Wie das Beispiel der *Medizin* zeigt, können insbesondere datenintensive Fächer wie die Onkologie von der Implementierung solcher Tools zur Verbesserung der Diagnostik und Therapie zum Wohle der Patientinnen und Patienten erheblich profitieren. Entscheidend ist dabei jedoch

deren Einbettung in eine vertrauensvolle Arzt-Patienten-Beziehung, um der individuellen Besonderheit der jeweiligen Lebensumstände gerecht zu werden.

Ganz ähnlich ist auch der Einsatz von KI-Anwendungen im weiten Bereich der *schulischen Bildung* differenziert zu beurteilen. Während sich etwa auf dem Gebiet von Lehr-Lern- sowie intelligenten Tutor-Systemen bereits gegenwärtig sinnvolle Einsatzgebiete beispielsweise zur Verbesserung individueller Lernprozesse abzeichnen, ist das Nutzen-Schaden-Potenzial bei anderen Anwendungen – wie zum Beispiel die Privatsphäre der Lernenden stark beeinträchtigenden Classroom Analytics – kritischer zu beurteilen.

Auch im Bereich der *öffentlichen Kommunikation und Meinungsbildung* ergibt sich ein breites Spektrum zwischen einer Erweiterung der Handlungsmöglichkeiten und ihrer Verengung. Dies zeigt sich exemplarisch daran, dass die Möglichkeiten, sich breit und umfassend zu informieren und mit vielen Menschen schnell und kostengünstig zu kommunizieren, so groß sind wie nie zuvor. Im Gegenzug kann es jedoch auch zur Verminderung von Handlungsfähigkeit kommen, etwa wenn die freie Meinungsbildung gestört oder gar manipuliert wird. Auch wenn Existenz, Beschaffenheit und Ursachen von Effekten wie Filterblasen, Echokammern oder (Wahl-)Manipulation genauso kontrovers diskutiert werden wie mögliche und Gegenmaßnahmen, so gilt es doch, Sorge zu tragen, dass Informations- und Kommunikationstechnologien so gestaltet werden, dass menschliche Autorschaft gestärkt und nicht vermindert wird.

In der *öffentlichen Verwaltung* werden die Auswirkungen von Technologien auf Handlungsfähigkeiten ebenso deutlich. Softwaresysteme wie beispielsweise Risk Scores haben das Ziel, Personen dabei zu unterstützen, bessere Entscheidungen zu treffen. Idealerweise sollen solche Systeme die Handlungsfähigkeit also erweitern. In der Realität ist dies gleichwohl nicht immer der Fall: Zum einen kann durch blindes Befolgen algorithmischer Empfehlungen (Automation Bias) die Handlungsfähigkeit entscheidenden Personen verringert werden. Zum anderen ist auch zu betrachten, inwiefern die Delegation von Entscheidungen an Softwaresysteme die Autorschaft und Handlungsmöglichkeit derer beeinträchtigt, die von den Entscheidungen betroffen sind.

Eine sektorenübergreifende Gemeinsamkeit hinsichtlich der angestrebten Erweiterung menschlicher Handlungspotenziale besteht darin, dass die komplette Ersetzung menschlicher Akteure durch KI-Systeme sich überall dort verbietet, wo die konkrete zwischenmenschliche Begegnung eine notwendige Voraussetzung für die Erreichung der jeweiligen Handlungsziele darstellt. Darüber hinaus besteht die Notwendigkeit, die Differenzen eines KI-Einsatzes in den

einzelnen Handlungsbereichen sorgfältig zu beachten. Diese betreffen nicht nur die unterschiedliche Sensibilität von Daten und ihre jeweilige Bedeutung für das Leben und die Privatsphäre der Betroffenen, sondern auch den Bezug zu ganz verschiedenen Gütern (z. B. Gesundheit, Bildung, Teilhabe, Sicherheit).

Empfehlung

- *Empfehlung Querschnittsthema 1:* Da die Vor- und Nachteile von KI-Anwendungen für verschiedene Personengruppen sowie die Gefahr des Verlustes bestimmter Kompetenzen bei den Personen, die solche Systeme anwenden, erheblich variieren, bedarf es sowohl einer differenzierten Planung des KI-Einsatzes in unterschiedlichen Handlungsfeldern, welche die jeweiligen Zielsetzungen und Verantwortlichkeiten präzise benennt, als auch einer zeitnahen Evaluation der tatsächlichen Folgen eines solchen Einsatzes, um die Systeme besser an die spezifischen Handlungskontexte anzupassen und sie fortlaufend zu verbessern.

10.2 Querschnittsthema 2: Wissenserzeugung durch KI und der Umgang mit KI-gestützten Voraussagen

Korrelationen und Datenmuster sind nicht mit Erklärungen und Begründungen von Ursachen von Ereignissen gleichzusetzen. Vielmehr müssen Daten nicht nur qualitativ evaluiert, sondern auch normativ beurteilt werden. Diese Beurteilung setzt die Fähigkeit voraus, Handlungen als Zweckrealisierungsversuche auszuführen, zu verstehen und zu überprüfen.

Mit dem Zuwachs qualitativ hochwertiger und relevanter Daten sowie verbesserter Methoden der Datenanalyse steigt häufig die Genauigkeit von Vorhersagen. Dennoch bestehen bei probabilistischen Methoden immer Restunsicherheiten. Dadurch ergibt sich die Frage, welcher Grad an epistemischer Sicherheit gesellschaftlich oder individuell für wünschenswert bzw. unerlässlich gehalten wird, um Handlungen und Entscheidungen zu legitimieren. Auch wie Fehler erster oder zweiter Art, das heißt falsch-positive und falsch-negative Ergebnisse und deren Folgen für unterschiedliche Betroffene zu bewerten sind, ist eine normative Frage, die den methodischen Entscheidungen in der Technikentwicklung vorausgehen muss.

Eine durch KI instrumentell unterstützte Daten- und Informationsbearbeitung sowie Wissensproduktion, die den Einfluss menschlicher Schwächen (z. B. Müdigkeit, Lustlosigkeit, Ehrgeiz, Bestechlichkeit) vermindert, würde epistemische und praktische Kompetenzen des Menschen grundsätzlich erweitern. Ein KI-gestützte Software allerdings, welche auf intransparenter Pro-

babilitätsbasis praktische Kompetenz übernimmt, würde die menschliche Handlungssouveränität und Verantwortung hingegen vermindern. Wollte man versuchen, den Menschen als Handlungssouverän zu ersetzen, käme es zu einer Diffusion oder vollständigen Eliminierung von Verantwortung.

Für die ethische Beurteilung des Einsatzes von KI ist relevant, dass durch diesen in allen vier in dieser Stellungnahme angesprochenen Anwendungsbereichen erhebliche funktionale Verbesserungen erreicht wurden und weiterhin erwartbar sind. Dies gilt bei Diagnose und Therapie in der *Medizin*, im Unterricht in der schulischen *Bildung*, bei Information und Austausch von Argumenten im Bereich der *öffentlichen Kommunikation und Meinungsbildung* und schließlich in der *Öffentlichen Verwaltung*. In allen Bereichen wird jedoch eine grundsätzlich normativ problematische Schwelle überschritten, wenn funktionale Verbesserungen (eventuell sogar unbemerkt) in eine Ersetzung moralischer Kompetenz und damit verbundener Verantwortung hinübergleiten.

Diese Schwelle ist beispielsweise im Gesundheitsbereich überschritten, wenn diagnostische und therapeutische Interventionen unter tatsächlicher Umgehung einer informierten Einwilligung der Patientin oder des Patienten (unter Umständen angetrieben durch Personalmangel) in Gänze digitalen Systemen überlassen werden. Im Bereich der *Bildung* wäre dies der Fall, wenn nicht nur routinierte Lehrprozesse an Software delegiert, sondern Lehrpersonen vollständig technologisch ersetzt würden. Im Bereich der *öffentlichen Kommunikation und Meinungsbildung* wäre eine Grenze überschritten, wenn für die am öffentlichen Diskurs Teilnehmenden die moralische und rechtliche Verantwortung für Meinungen und Tatsachenbehauptungen nicht mehr erkennbar ist oder der Austausch von Argumenten durch digitale Maßnahmen praktisch aufgehalten wird. Im Bereich der *Öffentlichen Verwaltung* gilt Entsprechendes, wenn Menschen nicht erkennen können, dass Verwaltungsentscheidungen durch digitale Agenten übernommen wurden oder eine solche Übernahme effektive Kontrollen verhindert. In allen Kontexten ist das entscheidende Kriterium für eine ethische Beurteilung die Möglichkeit der Identifizierung und der moralischen bzw. rechtlichen Inanspruchnahme von Agierenden, die in der Lage und auch verpflichtet sind, Handlungsverantwortung zu übernehmen.

<u>Empfehlung</u>

- *Empfehlung Querschnittsthema 2:* Der Einsatz KI-gestützter digitaler Techniken ist im Sinne der Entscheidungsunterstützung und nicht der Entscheidungsersetzung zu gestalten, um Diffusion von Verantwortung zu verhindern. Er darf nicht zulasten effektiver Kon-

trolloptionen gehen. Den von algorithmisch gestützten Entscheidungen Betroffenen ist insbesondere in Bereichen mit hoher Eingriffstiefe die Möglichkeit des Zugangs zu den Entscheidungsgrundlagen zu gewähren. Das setzt voraus, dass am Ende der technischen Prozeduren entscheidungsbefugte Personen sichtbar bleiben, die in der Lage und verpflichtet sind, Verantwortung zu übernehmen.

10.3 Querschnittsthema 3: Die Gefährdung des Individuums durch statistische Stratifizierung

Mit auf der Grundlage von Big Data erstellten Korrelationen werden menschliche Individuen Kohorten („Strata“) zugeordnet (beispielsweise regional definierte Kohorten wie die Anwohnende einer Straße oder Alterskohorten), die ohne jede soziale Erlebnisqualität sein können. Die Bildung solcher Kohorten und die auf ihrer Basis durch Algorithmen produzierten Voraussagen können für diejenigen, die solche Ansätze verwenden, durchaus nützlich sein, versprechen sie doch über die Gesamtmenge der Entscheidungen hinweg eine Erhöhung von Effektivität und Qualität. Probleme treten freilich für Individuen auf, die von solchen kollektiven Schlüssen betroffen sind – insbesondere dann, wenn die statistisch getroffene Diagnose oder Prognose in ihrem konkreten Fall nicht zutrifft.

Im Bereich der *Medizin* zeigt sich diese strukturelle Asymmetrie, wenn beispielsweise ein Individuum aufgrund diagnostizierter Merkmale einer Kohorte zugeordnet und aufgrund von Algorithmen eine therapeutische Strategie festgelegt wird. Man kann von einem „statistischen Kollektivismus“ sprechen, gegenüber dem das Individuum geltend machen können muss, *als Individuum* betrachtet und behandelt zu werden. Das bedeutet, dass im Rahmen der Medizin als praktischer Wissenschaft eine korrelative Zuordnung eines Individuums zwar ein gutes diagnostisches Indiz sein kann, das wiederum eine gute praktische Heuristik rechtfertigt, aber grundsätzlich nicht die einzige Evidenzgrundlage für eine Behandlung sein darf. Epidemiologische Studien grenzen entsprechend denkbare prioritäre Diagnosen ein und legen auf dieser Grundlage bestimmte Therapien nahe. Sie sollten jedoch nie allein die Grundlage für eine Behandlung sein.

Die gleichen Probleme betreffen auch die datenbasierte Diagnostik und Prognose im Bereich der *öffentlichen Verwaltung* oder der *schulischen Bildung*. Auch wenn solche Systeme das Ziel haben, nicht nur die Effektivität, sondern auch die Qualität von Entscheidungen zu verbessern, so bleibt ein unüberbrückbarer Spalt zwischen der je individuellen Lebenssituation und deren statistischer Annäherung. Auch wenn 99 von 100 Personen mit einem bestimmten Risikoprofil

erneut straffällig werden oder keinen Schulabschluss erreichen werden, so wissen wir doch nicht, ob diese eine Person, über die zu entscheiden ist, nicht genau die eine Ausnahme ist.

Das Individuum entzieht sich einer vollständigen statistischen Annäherung. Jedoch wirkt jeder Versuch einer statistischen Annäherung möglicherweise auf das Individuum zurück. Genauer: Die Probleme der datenbasierten Voraussagen werden noch einmal um eine Dimension verschärft, wenn das Referenzobjekt, auf das sich Wahrscheinlichkeitsaussagen beziehen, durch diese Aussagen selbst beeinflussbar ist, wie das für alle Kontexte menschlichen Handelns, so zum Beispiel in der Volkswirtschaft oder im Gesundheitssystem, zutrifft. Dies gilt sowohl für Individuen als auch für Kollektive. Die Rückwirkungen von Klassifikationen und statistischen Vorhersagen insbesondere in Medizin, Sozialwissenschaften und Psychologie auf die betroffenen Individuen hat Ian Hacking mit dem Terminus *looping effect* beschrieben.[423] Aber auch auf kollektiver Ebene können beispielsweise aus epidemiologischen Studien gewonnene konditionierte Voraussagen bei passender öffentlicher Verbreitung kollektive Verhaltensweisen erheblich verändern – derart, dass die Datengrundlage nicht mehr verlässlich ist.

Diese Rückwirkung von Klassifikation und Prognostik auf Individuen wie Kollektive zeigt sich nicht zuletzt auch in Bezug auf Soziale Medien im Bereich *der öffentlichen Kommunikation und Meinungsbildung*. Der Begriff des Hypernudge verweist hierbei auf die Besonderheit datenbasierter Systeme, die im Hintergrund und in Echtzeit die Informationsumgebung von Individuen, basierend auf statistischen Analysen, permanent anpassen und dadurch die weitere Informationsauswahl und die Entscheidungsmöglichkeiten präkonfigurieren.[424]

Empfehlung

- *Empfehlung Querschnittsthema 3:* Neben einer Analyse der konkreten und naheliegenden Probleme datenbasierter Software, beispielsweise in Bezug auf den Schutz der Privatsphäre oder die Verhinderung von Diskriminierung, gilt es, auch die langfristigen Auswirkungen dieser statistischen Präkonfiguration von Individuen sowie deren Rückwirkung – im Sinne einer Erweiterung oder Verminderung der Handlungsmöglichkeiten – auf Individuen wie Kollektive für alle Sektoren sorgfältig zu beleuchten.

[423] Hacking, I. (1986): Making up people. In: Heller, T. C. (Hg.): Reconstructing Individualism. Autonomy, Individuality, and the Self in Western Thought. Stanford, 222-236.
[424] Yeung, K. (2016): `Hypernudge `- Big Data as a mode of regulation by design. In: Information, Communication and Society, 20 (1), 118-136 (DOI: 10.1080/1369118X.2016.1186713).

Darüber hinaus gilt, dass Einzelfallbeurteilungen grundsätzlich wichtig bleiben. KI-basierte Beurteilungen und Vorhersagen können unter günstigen Bedingungen ein Hilfsmittel sein, aber kein geeignetes Instrument der *definitiven* Lagebeurteilung und Entscheidung. Pragmatische und heuristische Faktoren wie Prüfung der Kohärenz mit anderen Evidenzquellen, Erfolgseinschätzungen und anderes spielen eine nicht zu vernachlässigende Rolle.

10.4 Querschnittsthema 4: Auswirkungen von KI auf menschliche Kompetenzen und Fertigkeiten

Der Einsatz von Anwendungen der Künstlichen Intelligenz in ganz unterschiedlichen Lebensbereichen beeinflusst menschliche Handlungsfähigkeit und hat erheblichen Einfluss auf den Erwerb und den Erhalt menschlicher Kompetenzen und Fertigkeiten.

So zeigt sich im Bereich der *Medizin,* dass der fachgerechte Einsatz von KI-Anwendungen dazu geeignet ist, diagnostische Kompetenzen zu verbessern und zu einer deutlichen Reduktion von Fehldiagnosen führen kann. Auch in der *schulischen Bildung* dienen digitale Technologien in vielfältiger Weise der Kompetenzerweiterung. Im Bereich der *öffentlichen Kommunikation und Meinungsbildung* erweitern digitale Technologien die informationellen und kommunikativen Möglichkeiten der Beteiligten und erlauben damit eine eventuell kompetentere Teilnahme am öffentlichen Diskurs. In der *öffentlichen Verwaltung* wiederum sollen Softwaresysteme die Kompetenz der Beschäftigten erhöhen, gute und angemessene Entscheidungen zu treffen, etwa beim Einsatz von Risikobewertungen zur Prognose und Prävention von Gefahrenlagen.

Allerdings hat der Einsatz von KI-Anwendungen nicht nur das Potenzial, Kompetenzen zu erweitern, sondern auch, diese zu vermindern. Dies kann wiederum mit einem Beispiel aus dem medizinischen Bereich veranschaulicht werden: Eine Software, die Röntgenbilder präzise nach Auffälligkeiten untersucht und dabei statistisch deutlich besser abschneidet als der Mensch, erweist sich als signifikante Erweiterung bisheriger menschlicher Handlungsmöglichkeiten. Zugleich geht mit dem Einsatz des Systems das Risiko einher, dass der die Anwendung nutzende Mensch die eigenen Fähigkeiten einbüßt, Röntgenbilder unter diesem Gesichtspunkt zu analysieren. Dieser Effekt des Deskillings kann etwa deshalb eintreten, weil die jeweilige Fähigkeit zu ihrem Erhalt der steten Einübung bedarf, sie nunmehr aber nicht hinreichend häufig praktiziert wird, da die Technologie den Menschen mehr und mehr unterstützt bzw. seine Aufgabe vornehmlich übernimmt.[425] Ein weiterer Grund für Deskilling kann das große Vertrauen

[425] Bainbridge, L. (1983): Ironies of Automation. In: Automatica, 19 (6), 775-779, (DOI: 10.1016/0005-1098(83)90046-8).

in eine nahezu perfekte Technologie bei gleichzeitigem Misstrauen in die eigenen Fähigkeiten sein (Automation Bias). Der einzelne Mensch könnte geneigt sein, sich weniger auf das eigene als auf das technisch erzeugte Urteil zu verlassen – mit der Folge, dass wiederum Fähigkeiten nicht eingeübt werden und so mit der Zeit verloren gehen. Angesichts des verstärkten Einsatzes von KI-Anwendungen wird Deskilling in verschiedenen gesellschaftlichen Bereichen zunehmend befürchtet. Freilich ist Kompetenzverlust als Effekt nicht allein auf digitale oder gar KI-basierte Technologien begrenzt, sondern kann letztlich bei nahezu allen Werkzeugen beobachtet werden, derer Menschen sich zur Vereinfachung ihrer Aufgaben bedienen.

Den offenkundigen Vorteilen der Erweiterung von Fähigkeiten und Handlungsmöglichkeiten durch KI-basierte Systeme stünden also etwaige Nachteile gegenüber, die entstehen, wenn es zu einer Verringerung von Fähigkeiten und Fertigkeiten menschlicher Akteure in bestimmten gesellschaftlichen Bereichen käme. Wann dies in der Gesamtschau akzeptabel und wann problematisch wäre, lässt sich nicht pauschal oder übergreifend beantworten, sondern muss kontext- und fallspezifisch bestimmt werden. Weil die Nutzung von KI-Anwendungen dazu führen kann, dass bedeutsame menschliche Fähigkeiten nachlassen bzw. ganz verkümmern, können Abhängigkeiten von diesen Technologien entstehen. Es ist jedoch gerade angesichts der enormen Zunahme digitaler Technologien nicht ohne Weiteres ausgemacht, dass die Voraussetzungen für deren Einsatz, wie zum Beispiel eine stabile Stromversorgung, immer sicher gewährleistet werden können. Dies kann menschliche Fähigkeiten mitunter kurzfristig wieder bedeutsam werden lassen. Das spezifische Risiko von Kompetenzverlusten im Zusammenhang mit dem Einsatz von KI-Anwendungen liegt demnach in der Besonderheit der Tätigkeiten, die der Technik überlassen werden – handelt es sich dabei um gesellschaftlich besonders bedeutsame oder kritische Einsatzbereiche, ist ein Verlust von menschlichen Kompetenzen und Fertigkeiten ein ernstzunehmendes Risiko.

Allerdings ließe sich einem solchen Risiko entgegenwirken. Droht Kompetenzverlust und ist dieser gesellschaftlich unerwünscht, bedarf es dafür effektiver Vorkehrungen zur Vermeidung dieses Effekts. Eine Lösung kann beispielsweise darin liegen, dass die Technologie nur dann zum Einsatz gebracht werden darf, wenn sichergestellt ist, dass die Betroffenen die Fertigkeiten weiter regelmäßig trainieren. In der Medizin etwa könnte dies neben noch zu entwickelnden spezifischen Fortbildungsprogrammen bedeuten, Befundungen nie ganz an einen Algorithmus abzugeben.

Gleichwohl bleibt auch unter dieser Voraussetzung der Einwand bestehen, dass infolge eines immer durchgreifenderen Einsatzes von KI-Anwendungen in unterschiedlichen Lebensbereichen Menschen mehr und mehr dazu verleitet werden könnten, eigene Aufgaben an die Technik zu delegieren, weil diese als (vermeintlich) überlegen angesehen wird. Hieraus ergibt sich ein Risiko für die individuelle Selbstwahrnehmung. Mit zunehmender Delegation von Aufgaben an Technologien kann der Eindruck entstehen, immer mehr Kontrolle über wesentliche Bereiche des eigenen Lebens abzugeben. Im Extremfall könnte eine regelmäßige Delegation von Entscheidungen einen Effekt auf die Wahrnehmung des Selbst als Autor des eigenen Geschickes haben und sogar bürgerschaftliches Engagement reduzieren. Vor allem aber könnten „Kipppunkte" einer solchen passiven Selbstwahrnehmung schleichend erreicht und überschritten werden, ohne dass noch effektiv interveniert werden könnte.

Angesichts dieser Überlegungen sollten das Maß an Delegation, etwaige Verlagerungen von Aufgaben aus einem Lebensbereich in andere sowie die allgemeine gesellschaftliche Entwicklung infolge des verstärkten Einsatzes von KI-Anwendungen in dieser Hinsicht sorgsam beobachtet werden. Wenn das beschriebene Risiko jedoch mehr als eine denkbare Möglichkeit darstellt, sich also zu einem relevanten Faktor verdichtet, erscheinen Gegensteuerungsmaßnahmen erforderlich.

Empfehlung

- *Empfehlung Querschnittsthema 4:* Ob und inwiefern beim Einsatz von KI-Anwendungen Verluste menschlicher Kompetenz auftreten, die als unerwünscht eingestuft werden, muss sorgfältig beobachtet werden. Bei der Entwicklung und dem Einsatz neuer Technologien sind solch unerwünschte Kompetenzverluste durch eine sinnvolle Gestaltung des Zusammenspiels von Mensch und Technik, durch angemessene institutionelle und organisatorische Rahmenbedingungen sowie durch gezielte Gegenmaßnahmen wie etwa spezifische Trainingsprogramme zu minimieren bzw. zu kompensieren. Kompetenzverluste können sowohl individueller als auch kollektiver Natur sein. So gilt es zu verhindern, dass die Delegation von Aufgaben an Technologien dazu führt, dass Gesellschaften übermäßig anfällig werden, wenn diese Technologien (zeitweise) ausfallen. Jenseits dieser systemischen Aspekte müssen negative Auswirkungen solcher Delegation auf die individuelle Autonomie oder Selbstwahrnehmung mitigiert werden.

10.5 Querschnittsthema 5: Schutz von Privatsphäre und Autonomie versus Gefahren durch Überwachung und Chilling-Effekte

Ein weiteres wichtiges übergreifendes Thema, welches sich durch alle Anwendungsbereiche zieht, ist die Gefahr der Überwachung und die Möglichkeit auftretender Chilling-Effekte, die nicht nur die Privatsphäre Menschen gefährden, sondern auch deren Autonomie.[426] Viele der zuvor behandelten Technologien sind auf die Sammlung und Auswertung großer Mengen an personenbezogenen Daten angewiesen. Die Erfassung solcher Daten und die Möglichkeit, auf ihrer Basis sensible Prognosen zu erstellen, beeinträchtigt nicht nur die Privatsphäre der Personen, von denen die Daten stammen, sondern macht sie auch vulnerabel gegenüber möglichen Benachteiligungen oder Manipulationen, die aus der Verarbeitung der Daten resultieren können.

Aber auch wenn Daten nicht personenbezogen ausgewertet, das heißt keine Rückschlüsse auf das Individuum gezogen werden, kann sich bereits die Sorge vor der Möglichkeit solcher Rückschlüsse negativ auf die Autonomie und Entfaltungsfreiheit von Menschen auswirken. Der sogenannte *Chilling*-Effekt beschreibt in diesem Kontext Rückwirkungen auf Menschen, die Sorge haben, dass ihr Verhalten beobachtet, aufgezeichnet oder ausgewertet wird. Besonders sichtbar werden solche Chilling-Effekte im Kontext von *öffentlicher Kommunikation und Meinungsbildung*. So können Menschen davor zurückschrecken, nach relevanten, aber möglicherweise sensiblen Themen im Internet zu suchen oder als kontrovers wahrgenommene Inhalte zu lesen, in dem Wissen, dass ihr Onlineverhalten getrackt wird und aus der Sorge heraus, in welcher Art und Weise sich das möglicherweise nachteilig auf sie auswirken könnte.[427] Es geht hier also einerseits um die tatsächlichen negativen Effekte der Überwachung von Individuen und andererseits um Effekte, die als Anpassung an die Sorge vor dieser Überwachung entstehen.

Invasives Tracking führt nicht nur dazu, dass Menschen personalisierte, passgenaue Werbung zugespielt wird, sondern auch, dass Datenspuren, die sie online hinterlassen, im Aggregat und Zusammenspiel mit weiteren Datenspuren (sowohl aus anderen Quellen als auch von anderen Personen) etwa mittels Datenbroker zu vielfältigen, für die betroffene Person nicht nachvollziehbaren Zwecken zweitverwertet werden können. Das Geschäftsmodell von Datenbrokern

[426] Penney, J. W. (2017): Internet surveillance, regulation, and chilling effects online: a comparative case study. In: Internet Policy Review, 6 (2), 1-39. https://doi.org/10.14763/2017.2.692 [08.02.2023]; Solove, D. J. (2006): A Taxonomy of Privacy. In: University of Pennsylvania Law Review 154 (3), 477-560 (DOI: 10.2307/40041279).
[427] Büchi, M.; Festic, N.; Latzer, M. (2022): The Chilling Effects of Digital Dataveillance: A Theoretical Model and an Empirical Research Agenda. In: Big Data & Society, 9 (1), 1-14 (DOI: 10.1177/20539517211065368).

stellt eine Beeinträchtigung der Privatsphäre dar und beruht auf der Fiktion, dass eine Zustimmung zu Allgemeinen Geschäftsbedingungen von Online-Angeboten einer informierten Einwilligung gleichkommt.

Chilling-Effekte wiederum sind das Resultat des Wissens um oder auch nur der Ahnung von solchen Praktiken. Das eigene Verhalten verändert sich so, als würde man tatsächlich überwacht. Nicht nur der Informationskonsum wird einer Art Selbstzensur unterworfen, sondern auch die freie Meinungsäußerung kann online beeinträchtigt werden. Das Tracking insbesondere des Medienverhaltens von Individuen beeinträchtigt somit nicht nur deren intellektuelle Privatsphäre, sondern hat möglicherweise auch negative Auswirkungen auf die Kommunikationsfreiheit, auf individuelle Kreativität und kann somit der Persönlichkeitsentwicklung und der demokratischen Selbstverständigung abträglich sein.

Ähnliche Probleme können auch in gesundheitlichen Kontexten, bzw. an der Schnittstelle zwischen Onlineverhalten und *Medizin* entstehen. Wenn Onlineverhalten getrackt wird und Datenspuren von Brokern weiterverkauft werden, so kann die Sorge, aufgrund der möglichen Diagnose oder Prognose etwa einer psychischen Erkrankung keine private Krankenversicherung mehr zu bekommen, dazu führen, dass nicht nach Informationen gesucht wird, die etwa Depression oder Suchtverhalten nahelegen könnten. Auch bei der Verwendung von Technologien im Gesundheitswesen selbst, beispielsweise beim Einsatz von datenbasierten Systemen in der Diagnostik oder auch bei der Kommunikation mit therapeutischen Chatbots, können Menschen, die solche Angebote nutzen, annehmen – ob berechtigt oder nicht – dass sensible Daten oder Prognosen in die Hände unbefugter Dritter geraten können, und falsche Angaben machen, die möglicherweise zu falschen Diagnosen führen. Um etwaigen Chilling-Effekten entgegenzuwirken, gilt es also, unzulässige Datenweitergaben zu verhindern, insbesondere dann, wenn diese nicht im Interesse der Personen sind, von denen die Daten stammen, oder gar fremden, nicht gesundheitsbezogenen Zwecken dienen.

Auch im *schulischen Kontext* müssen Gefahren von Überwachung und mögliche Chilling-Effekte ernst genommen werden. Dies gilt insbesondere in Bezug auf Technologien, welche auf Video- oder Audioüberwachung des Klassenraums basieren. Emotionserkennung oder Aufmerksamkeitsmonitoring kann nicht nur die Privatsphäre der Lernenden sowie gegebenenfalls der Lehrkräfte auf unzulässige Art und Weise beeinträchtigen. Bereits die Sorge vor einer solchen Überwachung, selbst wenn die Daten nicht aufbewahrt und nicht personenbezogen analysiert und ausgewertet würden, hat unter Umständen Auswirkungen auf das Verhalten der Beteiligten.

Die parallelen Gefahren von Überwachung und Chilling-Effekten stellen sich schließlich auch im Bereich der *öffentlichen Verwaltung.* Paradigmatisch hierfür sind Maßnahmen im Kontext von Predictive Policing wie beispielsweise der Chat-Kontrolle. Eine anlasslose Überwachung der Telekommunikation ist grundrechtlich inakzeptabel; bereits die Sorge vor einer solchen Überwachung könnte dazu führen, dass Menschen< sich selbst zensieren.

Empfehlung

- *Empfehlung Querschnittsthema 5:* Die beschriebenen Phänomene sollten in ihrer Entstehung, Ausprägung und Entwicklung umfassend empirisch untersucht werden. Um sowohl dem Problem der Überwachung sowie den parallelen Gefahren durch etwaige Chilling-Effekte Rechnung zu tragen, müssen angemessene und effektive rechtliche und technische (beispielsweise *privacy by design*) Vorkehrungen getroffen werden, die dem übermäßigen Tracking von Onlineverhalten und dem Handel mit personenbeziehbaren Daten Einhalt gebieten. Die Interessen der Datensubjekte müssen hierbei im Mittelpunkt stehen. Insbesondere ist dabei auf besonders vulnerable Gruppen zu achten, da viele der Einsatzkontexte zudem von asymmetrischen Machtverhältnissen gekennzeichnet sind. Es muss Sorge getragen werden, dass die Erweiterung der Handlungsmöglichkeiten einiger nicht zulasten der Verminderung der Handlungsmöglichkeiten anderer, insbesondere benachteiligter Gruppen stattfindet.

10.6 Querschnittsthema 6: Datensouveränität und gemeinwohlorientierte Datennutzung

Wie zuvor erläutert, bergen viele der in dieser Stellungnahme vorgestellten Technologien insbesondere aufgrund der Nutzung vielfältiger Daten zwar zahlreiche Risiken, sie bringen aber gleichzeitig große Chancen mit sich, auf die wir als Gesellschaft ungern verzichten würden. Es müssen also Lösungen entwickelt werden, wie Daten sinnvoll für verschiedene wichtige Zwecke genutzt werden können, ohne zugleich den Schutz der Privatsphäre der Datenlieferanden unzulässig zu beeinträchtigen.[428]

[428] Diesem Zielkonflikt und den damit verbundenen, vielfältigen Herausforderungen hat sich bereits die Stellungnahme *Big Data und Gesundheit* des Deutschen Ethikrates aus dem Jahr 2017 gewidmet (Deutscher Ethikrat (2017): Big Data und Gesundheit – Datensouveränität als informationelle Freiheitsgestaltung. Berlin). Dort hat Deutsche Ethikrat die Notwendigkeit, 1) individuelle Interessen und die Privatsphäre effektiv zu schützen, 2) systematische Schadenspotentiale zu erkennen und zu verringern, sowie 3) gleichzeitig gemeinwohl-orientierten Aktivitäten in der Datennutzung zu ermöglichen, im normativen Konzept der Datensouveränität zusammengeführt. Unter dieser wird eine „den Chancen und Risiken einer digital vernetzten Welt angemessene, verantwortliche informationelle Freiheitsgestaltung“ verstanden. Der Rat hat in der

In diesem Zusammenhang stellt sich die Frage, ob das derzeitige Datenschutzrecht bzw. die herrschende Datenschutzpraxis diesen beiden Zielen gerecht wird. Hier kommt Kritik aus beiden Richtungen: Einerseits wird die Privatsphäre von datenliefernden Personen gerade im Kontext von Sozialen Medien und der zugrundeliegenden Datenökonomie nicht ausreichend geschützt. Andererseits wirft die aktuelle Datenschutzpraxis gerade in der öffentlichen Forschung vielfach Probleme auf – auch und gerade dort, wo es explizit um gemeinwohlorientierte Aktivitäten und eben nicht um (unbemerkte) Nachverfolgung geht, die die Privatsphäre und Interessen der Personen, die ihre Daten zur Verfügung stellen, verletzt. Während also in manchen Handlungsfeldern berechtigte Sorgen vor unbemerkten und weitreichenden Verletzungen von Privatsphäre und informationeller Selbstbestimmung herrschen, werden in anderen Kontexten durch strenge Auslegungen von Datenschutzregeln wichtige soziale Güter, etwa mit Blick auf Patientenversorgung und wissenschaftlichen Erkenntnisgewinn, aber auch die kommunale Daseinsvorsorge nicht oder nur sehr schwer erreicht.

Eine wesentliche Ursache für beide Probleme sind Charakteristika des Datenschutzrechts, die in der Vergangenheit umfassend kritisiert worden sind und die sich, verknappt gesprochen, aus dem zugrundeliegenden Individualismus und der damit verbundenen, überbetonten Rolle des Instruments der individuellen informierten Einwilligung zur Datennutzung ergeben.

Das geltende Datenschutzrecht ist auf die aktuellen Herausforderungen von KI nicht optimal vorbereitet. Durch den Fokus auf das initial datengebende Individuum werden systemische Risiken und Effekte auf andere Personen unzureichend berücksichtigt. Sogenannte *relationale Schäden*, die durch die Preisgabe der Daten anderer erfolgen können, können oft nicht hinreichend erfasst werden, etwa in Bezug auf Gruppen-Privatsphäre oder auch in Bezug auf Diskriminierung. Denn die angenommene enge Zweckbindung von Daten, die dieser Vorstellung zugrunde liegt, lässt sich allenfalls für die Datennutzung der ersten Instanz realisieren. Jede Weiterverwendung führt zu einer De- und Re-Kontextualisierung, zu der jeweils neue Zwecksetzungen gehören. Die Kaskade der Weiterverwendungen führt dabei sehr schnell aus dem Übersichtsbereich der Person, von der die Daten stammen, heraus. Zudem erschwert die stark auf das Individuum enggeführte Perspektive, dass systematische Chancen von Datennutzungen nicht bzw. unzureichend genutzt werden.

Stellungnahme eine Vielzahl von Empfehlungen vorgelegt, wie der Schutz der Privatsphäre unter den Bedingungen datengesättigter Lebenswelten verbessert werden kann; diese sind vielfach auch auf die hier verhandelten Anwendungsbereiche übertragbar.

Das weiterhin datenschutzrechtlich zentrale Instrument der informierten Einwilligung, das sich aus der individualistischen Perspektive auf Datennutzung ergibt, führt wiederum zweitens in verschiedenen Kontexten ins Leere bzw. in die praktische Dysfunktionalität: Insbesondere bei den beschriebenen Beispielen des Daten-Tracking im Bereich des medizinischen Online-Verhaltens, zum Beispiel bei Gesundheitsapps oder in Sozialen Medien allgemein, kann regelmäßig weder von Informiertheit noch von einer Einwilligung der Personen, die diese Angebote verwenden, ausgegangen werden. Dies hat sich auch im Gefolge der Einführung der Datenschutzgrundverordnung nicht verändert; die entsprechenden Details der Einwilligungserklärungen werden bekanntermaßen oft nicht gelesen bzw. nicht inhaltlich wahrgenommen. Dort wiederum, wo explizite informierte Einwilligungsformate unabdingbar und auch breit in der Praxis eingeführt sind – etwa im Bereich der medizinischen Forschung – können sie offenkundig sinnvolle Anwendungen stark erschweren oder verunmöglichen, weil sie, dem Primat der individuell orientierten engen Zweckbindung folgend, alle weiteren Datennutzungen verunmöglichen oder jedenfalls so ausgelegt werden. So wird etwa gerade in der öffentlich geförderten universitären Forschung durch eine restriktive Auslegungspraxis mit Blick auf Zweckbindung und informierte Einwilligungsformate die Sekundärdatennutzung von Patientendaten stark erschwert, selbst wenn Patientinnen und Patienten initial in eine intensive Nutzung ihrer Daten eingewilligt haben; der Spielraum, den die DSGVO hier bietet, wird oft nicht genutzt.

Dabei, so hat es der Deutsche Ethikrat bereits in seiner Stellungnahme zu Big Data und Gesundheit 2017 unterstrichen, sind aus ethischer Sicht nicht nur Präferenzen der individuellen Binnensphäre, sondern auch Gesichtspunkte der Solidarität im Umgang mit Datensammlung und -nutzung zu bedenken, die eine besondere Aktualität bei KI-Anwendungen erlangen. In der gängigen Metapher gesprochen, sind Daten grundsätzlich ein wertvoller Rohstoff, den die einzelne Person unter bestimmten Umständen dem Gemeinwesen zur Verfügung stellen sollte. Die Einordnung von Daten als potenzielles Gemeingut steht nicht in striktem Widerspruch zum individuellen Anspruch auf Schutz der Privatsphäre. Vielmehr widerstreitet sie dessen Absolutsetzung und pauschaler Vorrangstellung und verweist darauf, dass beide Vorgaben – kontextspezifisch – in einen angemessenen Ausgleich zu bringen sind. Daher sollte die Möglichkeit bestehen, als individuelle Akteurin oder Akteur selbstbestimmt Daten im Interesse der Allgemeinheit zur Verfügung zu stellen, beispielsweise für die medizinbezogene Grundlagenforschung. Dem entspricht der Fokus im unionalen Data Governance Act auf „Datenaltruismus“.

Diese Spannungen sind wohl bekannt im Bereich der *Medizin*. Dort ist das Ungleichgewicht zwischen – vereinfacht gesprochen – „Tracking-Wildwuchs“ etwa im App-Bereich auf der einen Seite und sehr restriktiver Praxis mit Blick auf klinische Sekundärdatennutzung von Daten

für sinnvolle, öffentliche und gemeinwohlorientierte Forschung auf der anderen Seite besonders ausgeprägt sowie bekannt.[429] Sie haben aber ebenso starke Relevanz in den anderen Anwendungsfeldern. Offenkundig ist dies auch bei datenintensiven Anwendungen im *schulischen Kontext* der Fall, wo der Schutz von Privatsphäre sowohl von Lernenden als auch von Lehrkräften essenziell ist und immer gewährleistet sein muss, etwa durch entsprechende technische Voreinstellungen (*privacy by design*). Gleichwohl ist auch hier offenkundig, dass sich aus einer stärker gemeinwohlorientierten Perspektive vielfältige wichtige Nutzungen für derart gesammelte Daten ableiten lassen. So etwa wäre die verantwortliche Auswertung von Daten aus Tutorsystemen geeignet, um Indikatoren für edukatorische Ungleichheiten oder Nachteile im Bildungssystem für bestimmte Gruppen von Lernenden systematischer, schneller und vorausschauender zu erfassen und zielgenauer anzugehen. Ähnliche Konstellationen ergeben sich auch in der *öffentlichen Verwaltung*. Dort kann eine präzisere und stärker vorausschauende Erfassung etwa von Risikofaktoren für eine Kindeswohlgefährdung oder in der Bewährungshilfe wichtigen individuellen und gesellschaftlichen Zielen dienen. Zugleich erfordert dies den hohen Schutz der Privatsphäre aller Betroffenen, um unzulässige Vorverurteilungen oder Stigmatisierungen zu vermeiden. Zusammenfassend scheinen sowohl die rechtlichen Rahmenbedingungen als auch die konkrete Anwendungspraxis in bestimmten Bereichen, insbesondere in den für die *öffentliche Kommunikation und Meinungsbildung* so zentralen Sozialen Medien und der zugrundeliegenden Datenökonomie, nicht nur die Privatsphäre der Nutzenden unzureichend zu schützen. Auch eine mögliche Datensouveränität wird ad absurdum geführt, da vielfach weder das Kriterium der Informiertheit noch jenes der Einwilligung gewährleistet ist. Dies gilt vielfach auch für das Datensammeln und -verarbeiten im Hintergrund kommerzieller Technologien, die im Bildungswesen, der öffentlichen Verwaltung und im Medizinbereich eingesetzt werden.

Demgegenüber werden wichtige Chancen insbesondere im Bereich der öffentlichen Forschung durch eine teils übermäßig rigide Datenschutzpraxis vergeben, die etwa Fortschritten mit Blick auf Bildungsgerechtigkeit, der Früherkennung von Erkrankungen oder der flächendeckenden Vermeidung von Medikamentenverschreibungsfehlern sowie optimierter und zugleich maßvoller und verantwortlicher Risikoabschätzung im Sozialwesen entgegensteht.

---

[429] Sachverständigenrat zur Begutachtung der Entwicklung im Gesundheitswesen (2021): Digitalisierung für Gesundheit. Ziele und Rahmenbedingungen eines dynamisch lernenden Gesundheitssystems – Gutachten 2021. Bern.

Empfehlung

- *Empfehlung Querschnittsthema 6:* Mit Blick auf KI-Anwendungen müssen neue Wege gefunden werden, um innerhalb der jeweiligen Kontexte und mit Blick auf die jeweils spezifischen Herausforderungen und Nutzenpotenziale die gemeinwohlorientierte Daten(sekundär)nutzung zu vereinfachen bzw. zu ermöglichen und damit die Handlungsoptionen auf diesem Gebiet zu erweitern. Zugleich ist es essenziell, einen Bewusstseinswandel sowohl in der Öffentlichkeit als auch bei den praktisch tätigen Personen, die Datennutzung gestalten, herbeizuführen – weg von einer vornehmlich individualistisch geprägten und damit verkürzten Perspektive, hin zu einer Haltung, die auch systematische und gemeinwohlbasierte Überlegungen mit einbezieht und in einen Ausgleich bringt. Eine solche Haltung ist auch für die zukünftige Politikgestaltung und Regulierung deutlich stärker als bisher zugrunde zu legen. Nur so kann es gelingen, neben den Risiken, die sich aus breiterer KI-Anwendung ohne Zweifel ergeben, zugleich die wichtigen Chancen einer verantwortlichen Nutzung nicht aus dem Blick zu verlieren.

10.7 Querschnittsthema 7: Kritische Infrastrukturen, Abhängigkeiten und Resilienz

Infrastrukturen sind zentrale Elemente moderner Gesellschaften, die für das Funktionieren gesellschaftlicher Teilbereiche essenziell sind und deren Ausfall mit großen gesamtwirtschaftlichen Schäden verbunden ist. Im Zuge der Digitalisierung werden Infrastrukturen wie beispielsweise Stromnetze zunehmend digital überwacht und über das Internet gesteuert. Auf der anderen Seite werden digitale Technologien selbst zu Infrastrukturen. Dies gilt insbesondere für digitale Medien im Kontext der *öffentlichen Kommunikation und Meinungsbildung*: Kommerzielle Plattformen stellen hier zunehmend Infrastrukturen für den öffentlichen Diskurs dar.

Im Kontext von *öffentlicher Verwaltung* und *Bildung*, aber auch in Teilen der *Medizin*, stellt sich die infrastrukturelle Bedeutung digitaler Technologien teilweise ähnlich, teilweise sehr unterschiedlich dar. Die Gemeinsamkeit besteht darin, dass es in Teilen dieselben Unternehmen sind, die nicht nur die großen Plattformen betreiben, sondern auch Technologien für die Bereiche öffentliche Verwaltung, Bildung und Medizin anbieten. Der Unterscheid besteht darin, dass sich die Vulnerabilität der öffentlichen Verwaltung und Schulen, und teilweise auch der medizinischen Versorgung, beispielsweise in der Corona-Pandemie gerade in einer unzureichenden Digitalisierung gezeigt hat.

Diese doppelte infrastrukturelle Bedeutung digitaler Technologien gilt es also zu berücksichtigen und unter ethischen Aspekten zu beleuchten. Hier ist auf folgende mögliche Problembereiche hinzuweisen:

Während Infrastrukturen von Menschen aufgebaut werden, um bestimmten Zwecken zu dienen, Menschen also die Gestalter der Systeme sind, hat der infrastrukturelle Charakter zur Folge, dass sich dieses Verhältnis allmählich verschiebt. Denn am Vorhandensein und Funktionieren der betreffenden Infrastruktur richten Menschen ihr Handeln aus, Im Zuge dieser sozialen Aneignung einer Infrastruktur entstehen Abhängigkeiten, die die menschliche Autonomie gefährden können. Dies zeigt sich beispielsweise im Kontext der Übernahme von Twitter durch Elon Musk. Dabei wurde deutlich, dass es keine vergleichbaren Alternativen gibt, auf welche im Falle grundlegender Änderungen der Geschäftsmodelle oder Moderationspraktiken ausgewichen werden könnte, ohne dass zentrale Aspekte der Plattformnutzung wegfallen würden.[430]

Die vorstehend geschilderte Problematik gilt für Infrastrukturen generell und ist nicht spezifisch auf KI bezogen. Wenn jedoch KI-gestützte ADM zusehends in die Steuerung der Infrastrukturen integriert werden, kommt eine neue Form von Abhängigkeit hinzu. KI-Systeme sind nicht vollständig transparent und nachvollziehbar. Die infrastrukturelle Abhängigkeit würde also auch die Abhängigkeit von Systemen beinhalten, die zumindest zum Teil Black Boxes sind (vgl. Abschnitt 10.10). Im Zusammenhang mit den Phänomenen menschlicher Gewöhnung an entsprechende Systeme kann es zur Verfestigung von Verhaltensmustern in sozio-technischen Konstellationen kommen, die eine kritische Reflexion und ein Hinterfragen der teils intransparenten ADM-Systeme erschweren können. Durch solche Pfadabhängigkeiten (vgl. Abschnitt 10.8) wird eine Eigendynamik eingeleitet, die den Wechsel auf andere Formen erschwert oder unmöglich macht und damit Optionen verbaut. Auch hier kann Twitter zur Illustration dienen: Sein Mehrwert besteht gerade in der Interaktion mit anderen Nutzenden; durch einen Wechsel auf ein anderes, gegebenenfalls dezentrales Medium wie Mastodon geht ein Teil des Nutzens jedoch verloren. Solche sogenannten *Netzwerkeffekte* erschweren den Wechsel zwischen Plattformen und erhöhen die Abhängigkeiten.

Die Abhängigkeit vom Funktionieren der digitalen Systeme betrifft längst nicht mehr nur die Informations- und Kommunikationsinfrastruktur. Dadurch, dass mittlerweile viele Infrastrukturen digitalisiert sind und über das Internet gesteuert werden, sind sie mit dem Internet als

[430] Zwar migrieren derzeit viele Twitter-Nutzerinnen und -Nutzer zu anderen Services wie beispielsweise Mastodon. Deren Funktionalität und Reichweite ist jedoch nicht mit Twitter deckungsgleich.

Quasi-Nervensystem zu einer Mega-Infrastruktur verbunden. Hackerangriffe, technisches Systemversagen oder sozio-technische Systemrisiken können auf diese Weise erheblich größere Reichweite haben als in separierten Infrastrukturen. Durch die fortwährende Komplexitätssteigerung der Infrastruktursysteme und ihre Steuerung steigt die gesellschaftliche und institutionelle Vulnerabilität weiter an. Demgegenüber ist die allgemeine Erfahrung, dass die Infrastrukturen zuverlässig funktionieren und Störungen vorübergehend sind. Dieses Funktionieren, so etwa die hohe Zuverlässigkeit der Stromversorgung und des Internets in westlichen Ländern, kann blind für die zunehmende Abhängigkeit von digitaler Technik machen. Schwere Wirtschaftskrisen, technische Systemeffekte, ein Kollaps der staatlichen Ordnung oder Hacker-Angriffe sind jedoch nicht unmöglich. Ethisches Vorsorgedenken gebietet, solche Abhängigkeiten zumindest bewusst zu machen und Strategien für digitale Blackouts zu entwickeln. Dies gilt umso mehr für intransparente, KI-getriebene Systeme, in denen sowohl Fehler als auch Angriffe noch schwerer zu entdecken und nachzuweisen sind.

Empfehlung

- *Empfehlung Querschnittsthema 7:* Um die Autorschaft menschlicher Akteure und deren Handlungsmöglichkeiten zu erweitern, muss die Resilienz sozio-technischer Infrastrukturen gestärkt und die Abhängigkeit von individuellen Akteuren und Systemen minimiert werden. Dies umfasst zunächst die Notwendigkeit, die infrastrukturelle Bedeutung digitaler Technologien anzuerkennen und infolgedessen dem Schutz und der Resilienz kritischer digitaler Infrastrukturen mehr Aufmerksamkeit zuteilwerden zu lassen, auch im politischen Handeln. In allen Sektoren gilt es, einseitige Abhängigkeiten zu vermeiden, welche im Krisenfalle verletzlich und angreifbar machen.

  Für Nutzerinnen und Nutzer erfordert eine Verringerung der Abhängigkeit die Möglichkeit, zwischen Alternativen zu wählen, ohne große Teile der Funktionalität einzubüßen. Dies umfasst zum einen die Notwendigkeit von Interoperabilität, um einfach zwischen Systemen wechseln zu können. Hierfür ist auch der Auf- und Ausbau alternativer Infrastrukturen von besonderer Bedeutung. Im Kontext der öffentlichen Meinungsbildung erscheint die Etablierung unabhängiger, öffentlicher digital-kommunikativer Plattformen dringend geboten. Aber auch in anderen Sektoren wie der Verwaltung, der Bildung oder der Medizin vermindert eine zu große Abhängigkeit von wenigen Systemen oder Akteuren potenziell die individuelle wie kollektive Handlungsfähigkeit.

10.8 Querschnittsthema 8: Pfadabhängigkeiten, Zweitverwertung und Missbrauchgefahren

Eine wichtige und oft vernachlässigte Rolle bei der Entwicklung und Nutzung von Technologien sind Pfadabhängigkeiten. Entscheidungen, die zu Beginn einer bestimmten Entwicklung getroffen wurden, können noch lange nachwirken und sind teils schwer wieder aufzuheben, auch wenn sich der Kontext der Nutzung möglicherweise geändert hat. Dies gilt insbesondere dann, wenn es sich hierbei um grundlegende Technologien handelt, die Infrastrukturen prägen, welche in komplexe sozio-technische Zusammenhänge eingebettet sind oder aber es durch die Nutzung bereits zu Gewöhnungseffekten gekommen ist. Von besonderer Relevanz für Infrastrukturen sind hier die Festlegungen von Normen und Standards, die zukünftige Entwicklungsmöglichkeiten begrenzen. Man denke hier etwa an standardisierte Stecker und Steckdosen, deren aktuelle Form nicht notwendigerweise besser oder schlechter als Alternativen sind, deren Änderung aber mit hohen Kosten verbunden wäre. Ein klassisches Beispiel für Pfadabhängigkeiten und Beharrungstendenzen durch Gewöhnungseffekte ist auch die Anordnung der Buchstaben auf Tastaturen. Die heute noch weit verbreitete QWERTY/Z-Tastatur wurde für mechanische Schreibmaschinen entwickelt, um die am häufigsten vorkommenden Buchstabenfolgen räumlich aufzuteilen, damit sich die mechanischen Typenhebel dieser Buchstaben weniger verhaken. Dass diese vergleichsweise wenig ergonomische Anordnung weiter Bestand hat, liegt an der Gewöhnung durch die Nutzenden.

Eine Konsequenz aus dieser Beobachtung lautet, dass grundlegenden Entscheidungen bei der Gestaltung von Infrastrukturen und dem Festsetzen von Standards und Normen eine hohe Aufmerksamkeit zukommen sollte. Für digitale Technologien, die zunehmend den Status kritischer Infrastrukturen annehmen, gilt dies umso mehr. Für Technologien, die entweder über einen weiten Verbreitungsgrad verfügen und/oder eine infrastrukturelle Dimension haben, könnte es zunehmend schwer und kostspielig werden, Änderungen einzufordern. Das Vorhandensein von Technologien und Infrastrukturen kann zudem selbst auch Erwartungen hinsichtlich der Verfügbarkeit und möglichst umfänglicher Nutzung wecken. Gerade bei kostspieligen Technologien dürfte eine Tendenz auszumachen zu sein, deren Möglichkeiten voll auszuschöpfen – auch über das ursprüngliche Anwendungsfeld hinaus.

Eine „Verführung“, technologische Möglichkeiten auch zu anderen Zwecken als zunächst intendiert zu nutzen, ist nicht prinzipiell problematisch. Mit dem Computer als Universalmaschine zeichnen sich digitale Technologien in der Regel durch vielfältige Einsatzszenarien aus. Dennoch öffnet dies auch die Tür für fremdnützigen Gebrauch oder Missbrauch. So wurden im

Kontext der Corona-Pandemie weltweit sehr unterschiedliche Contact-Tracing-Apps entwickelt. Während in Deutschland auf ein dezentrales Modell gesetzt wurde, um möglichen Missbrauch bereits technisch zu verhindern, gab es aus anderen Ländern Hinweise auf Zweckentfremdung.[431]

Sobald eine Technologie etabliert ist, kann es schwer sein, weitere, auch missbräuchliche Nutzungsszenarien auszuschließen. Diese Tendenz wird oft auch mit dem Potenzial des sogenannten *Dual Use* beschrieben. Der Begriff *Dual Use* wurde ursprünglich dafür verwendet anzuzeigen, dass Technologien sowohl für zivile als auch militärische Zwecke genutzt werden können. Mittlerweile wird der Begriff aber breiter verwendet, um darauf zu verweisen, dass Technologien oder Forschungsergebnisse sowohl für friedliche und nützliche Zwecke als auch zur absichtlichen Schädigung von Gesellschaft oder Umwelt, beispielsweise in krimineller oder terroristischer Absicht, eingesetzt werden können.[432] In den letzten Jahren sind digitale Technologien im Allgemeinen und KI im Besonderen zunehmend unter dem Blickwinkel von Dual Use beleuchtet worden. Dabei zeigt sich, dass die Konzeption von Dual Use hier um weitere Komplexitätsstufen ergänzt werden sollte, da digitale Technologien und insbesondere Grundlagentechnologien wie das maschinelle Lernen, oft sehr mannigfaltige Nutzungsmöglichkeiten eröffnen, in denen die Frage der Abgrenzung von Ge- und Missbrauch zunehmend schwieriger wird.

Empfehlung

- *Empfehlung Querschnittsthema 8:* Bei Technologien mit großen Auswirkungen oder hohem Verbreitungsgrad und vor allem dort, wo sich eine Nutzung von Technologien kaum oder gar nicht vermeiden lässt, müssen bereits zu Beginn der Entwicklungsplanung mögliche Langzeitfolgen wie Pfadabhängigkeiten im Allgemeinen sowie Dual-Use-Potenziale im Speziellen regelhaft und explizit mitgedacht und antizipiert werden. Dies gilt in besonderem Maße in der Anwendungsplanung. Dabei sind neben direkten, sektorspezifischen Schadenspotenzialen auch etwaige – natürlich deutlich schwieriger fass- und antizipierbare – sektorübergreifende Effekte zu bedenken. Hohe Standards für die Sicherheit und den Schutz der

[431] So wurden Contact Tracing Apps, welche zur Nachverfolgung von Covid entwickelt wurden zum Beispiel auch zur Unterstützung polizeilicher Ermittlungen in Singapur herangezogen. (Taylor, J. (2021): Singapore says police will be given access to Covid-19 contact tracing data. In: The Guardian. https://www.theguardian.com/world/2021/jan/05/singapore-says-police-will-be-given-access-to-covid-19-contact-tracing-data [01.02.2023]).
[432] Deutscher Ethikrat (2014): Biosicherheit – Freiheit und Verantwortung in der Wissenschaft. Berlin; Gemeinsamer Ausschuss zum Umgang mit sicherheitsrelevanter Forschung (www.sicherheitsrelevante-forschung.org).

Privatsphäre (*security by design, privacy by design*) können ebenfalls dazu beitragen, spätere missbräuchliche Anwendungen einzuhegen bzw. möglichst zu verhindern.

Bei besonders invasiven Technologien beispielsweise in der öffentlichen Verwaltung, die Bürgerinnen und Bürger gegebenenfalls verpflichtend nutzen müssen, sind besonders hohe Standards einzuhalten. Um dies sicherzustellen und überprüfen zu können, sind gegebenenfalls Open-Source-Ansätze angezeigt (vgl. Abschnitt 10. 10).

10.9 Querschnittsthema 9: Bias und Diskriminierung

Datenbasierte KI-Systeme lernen auf Basis vorhandener Daten. Resultierende Prognosen und Empfehlungen schreiben somit die Vergangenheit in die Zukunft fort, wodurch Stereotype, aber auch bestehende gesellschaftliche Ungerechtigkeiten durch den Einbau in scheinbar neutrale Technologien reproduziert und sogar verstärkt werden können. In den letzten Jahren wurden die teils diskriminierenden Effekte insbesondere datenbasierter Technologien zur Entscheidungsunterstützung in zahlreichen Sektoren nachgewiesen. Für den Bereich der *öffentlichen Verwaltung* sei hier exemplarisch auf die Debatten rund um die Software COMPAS hingewiesen (vgl. Abschnitt 8.3.1). Auch im Kontext der für *öffentliche Kommunikation und Meinungsbildung* zentralen Sozialen Medien und Suchmaschinen konnte gezeigt werden, dass algorithmische Systeme gesellschaftliche Stereotype und Ungerechtigkeiten reproduzieren können und dies in systematischen diskriminierenden Verzerrungen resultieren kann. In der *Medizin* wiederum gibt es zahlreiche Beispiele, dass verzerrte Trainingsdaten zu diskriminierenden Ergebnissen bei der Beurteilung von Patientinnen und Patienten durch KI-basierte Systeme führen können, so etwa bei der Einschätzung, wieviel Nachsorgebehandlungen Menschen nach einem Krankenhausaufenthalt benötigen – so wurde in diesem Fall die vorhandene Verzerrung in den Trainingsdaten in eine direkte Benachteiligung bestimmter Personengruppen übersetzt.[433] Auch im *Bildungsbereich* können systematische Verzerrungen nachgewiesen werden, insbesondere im Kontext von Audio- und Videoanalysen zum Zweck der Emotions- und Affekterkennung.

Die Ursachen für Diskriminierung durch KI-Systeme sind vielfältig. Oft liegt bei deren Entwicklung keine unmittelbare Diskriminierungsabsicht vor. Stattdessen sind diskriminierende Effekte das Resultat gesellschaftlicher Realitäten oder Stereotype in Kombination mit technisch-methodischen Entscheidungen, wie beispielsweise der Wahl der Zielvariablen und Labels,

[433] Obermeyer, Z. et al. (2019): Dissecting racial bias in an algorithm used to manage the health of populations. In: Science 366 (6464), 447-453. (DOI: 10.1126/science.aax2342).

der Auswahl der Trainingsdaten oder der verwendeten statistischen Analysemethoden. Dies erschwert einerseits die Anwendung rechtlicher Regulierungen, die auf Vorsätzlichkeit von Diskriminierung bauen. Andererseits bedarf es detaillierter Analysen der methodisch-technischen Ursachen für diskriminierende Effekte, die häufig schwierig zu erkennen und nachzuweisen sind.

Von besonderer Bedeutung sind hier Trainingsdaten. Mangelnde Qualität, aber auch existierende gesellschaftliche Ungleichheiten, die sich in Daten widerspiegeln, können zu diskriminierenden Modellen führen. Über- und Unterrepräsentativität sind weitere Probleme, die diskriminierende Effekte haben können. Wenn Software im Bereich der *Medizin* zur Diagnostik von Hautkrebs vor allem auf Bildern mit heller Haut trainiert wurde, kann dies zu unterschiedlicher Genauigkeit bei der Befundung von verschiedenen Hautfarben führen. Die Folge wären beispielsweise überproportional häufigere Fehldiagnosen für Patientinnen und Patienten mit dunklerer Hautfarbe.

Das Gegenteil dieser mangelnden Berücksichtigung von Personengruppen in Datensätzen ist die Überrepräsentativität. Ein Beispiel ist der Einsatz von Software zur Vorhersage von Straftaten im Kontext von prädiktiver Polizeiarbeit. Falls eine Software eine bestimmte Gegend aufgrund bisheriger Straftaten als Hoch-Risikozone kategorisiert, werden dort gegebenenfalls in der Folge Polizeikontrollen verstärkt. Gerade aufgrund der erhöhten Kontrolle können dort dann noch mehr Straftaten verzeichnet werden. Im Gegenzug bleiben gegebenenfalls Straftaten in Niedrig-Risiko-Zonen aufgrund ausbleibender Kontrollen unerkannt. Durch beide Effekte kann sich die Datenlage weiter zuungunsten der Bewohnerschaft einer Hoch-Risikozone verschieben – nämlich für jene, die durch verstärkte Kontrollen unter Umständen ungerechtfertigt stigmatisiert werden.

Darüber hinaus besteht das Problem der sogenannten redundanten Enkodierung: Sensible Attribute wie beispielsweise Geschlecht, religiöse Zugehörigkeit oder sexuelle Orientierung lassen sich teilweise aus anderen Datenpunkten, etwa Bewegungsprofilen, Details des Medienkonsums sowie Angaben über Wohnort oder Hobbys ableiten. Dadurch können diese Daten als sogenannte Proxy-, bzw. Stellvertretervariablen für die geschützten Variablen fungieren. Das Resultat ist, dass Personen auch dann aufgrund ihres Geschlechts oder ihrer sexuellen oder religiösen Orientierung von der Software diskriminiert werden können, wenn diese Angaben gar nicht erhoben wurden, eben weil diese Kategorien aus anderen erhobenen Daten ableitbar sind.

In den zuvor genannten Beispielen ist Diskriminierung der Effekt von technisch-methodischen Entscheidungen, aber nicht notwendigerweise intendiert. Es ist allerdings zumindest denkbar,

+++ Vorabfassung! Wird durch die lektorierte Fassung ersetzt. +++

dass auch explizite Diskriminierungsabsichten in komplexen Systemen versteckt werden könnten. Dies gilt umso mehr in proprietärer, das heißt rechtlich geschützter Software, in welche die Personen, die sie verwenden, nicht nur aufgrund von technischer Komplexität, sondern auch aus rechtlichen Gründen keine Einsicht haben.

Empfehlung

- *Empfehlung Querschnittsthema 9:* Zum Schutz vor Diskriminierung in Anbetracht der zuvor dargelegten Herausforderungen bedarf es *angemessener Aufsicht und Kontrolle* von KI-Systemen. Besonders in sensiblen Bereichen erfordert dies den Auf- oder Ausbau gut ausgestatteter Institutionen. Hier gilt: je größer die Eingriffstiefe und je unumgänglicher die Systeme, desto höher die Anforderungen an Diskriminierungsminimierung.

  Auch bereits bei der Entwicklung von Technologien gilt es, Diskriminierung zu minimieren bzw. Fairness, Transparenz und Nachvollziehbarkeit herzustellen. Dies sollte sowohl durch Anreize – etwa Forschungsförderung – als auch durch entsprechende gesetzliche Anforderungen befördert werden, etwa hinsichtlich der Offenlegung, welche Maßnahmen zur Diskriminierungsminimierung bei der Softwareentwicklung ergriffen wurden.[434]

  Allerdings haben technische wie regulatorische Maßnahmen zur Minimierung von Diskriminierung ihre Grenzen, unter anderem weil unterschiedliche Fairnessziele technisch nicht gleichzeitig erfüllt werden können. Es müssen also zugleich ethische und politische Entscheidungen getroffen werden, welche Kriterien für Gerechtigkeit in welchem Kontext zum Tragen kommen sollen. Diese Entscheidungen dürfen nicht den Personen, die Software entwickeln, und anderen direkt Beteiligten überlassen werden. Stattdessen bedarf es der Entwicklung geeigneter Verfahren und Institutionen, um diese Kriterien kontextspezifisch und demokratisch, gegebenenfalls immer wieder neu auszuhandeln. Je nach Anwendungskontext und Sensibilität des einzusetzenden Systems kann die Beteiligung der Öffentlichkeit erforderlich sein. Dabei sollte der Schutz der jeweils bedürftigsten bzw. von Entscheidungen besonders betroffenen Gruppen besonders berücksichtigt werden.

[434] Der derzeit diskutierte Entwurf eines EU Artificial Intelligence Act (AI Act) verfolgt bereits diesen Ansatz, auf der einen Seite die Forschung zu KI-Technologien zu fördern und auf der anderen, einen rechtlichen Rahmen für ihre Entwicklung und Anwendung zu schaffen.

10.10 Querschnittsthema 10: Transparenz und Nachvollziehbarkeit – Kontrolle und Verantwortung

KI-Systeme sind mitunter wenig transparent und nachvollziehbar. Diese Opazität hat verschiedene Ursachen, die vom Schutz geistigen Eigentums über die Komplexität und Nicht-Nachvollziehbarkeit der Verfahren bis hin zur mangelnden Durchsichtigkeit von Entscheidungsstrukturen, in die der Einsatz algorithmischer Systeme eingebettet ist, reichen. Als Reaktion auf diese vielfältigen Herausforderungen gibt es Bemühungen, die Transparenz und Nachvollziehbarkeit durch technische, organisatorische und rechtliche Mittel zu erhöhen.[435]

Fragen von Transparenz, Erklärbarkeit und Nachvollziehbarkeit sind mit Fragen von Kontrolle und Verantwortung verbunden. Zwar besteht zwischen ihnen ein gewisser Zusammenhang, doch ist die Transparenz und Nachvollziehbarkeit algorithmischer Systeme für die Kontrolle und die Verantwortung für ihren Einsatz weder zwingend notwendig noch hinreichend.

Einerseits kann es auch bei prinzipiell transparenten und nachvollziehbaren Methoden wie zum Beispiel Entscheidungsbäumen dazu kommen, dass verantwortliches Handeln und angemessene Kontrolle ausbleiben. Zudem ist in Bezug auf Offenlegungspraktiken auf das Problem strategischer Transparenz hinzuweisen. So könnten insbesondere im Bereich der öffentlichen Kommunikation und Meinungsbildung Plattformbetreibende entweder irrelevante und unzureichende Informationen transparent machen (beispielsweise in Bezug auf Prozesse und Effekte von Content-Moderation) oder aber relevante Informationen unter einer Fülle irrelevanter Information verbergen. In diesen Fällen würde Transparenz also nicht zwingend zu verantwortlichem Handeln und Kontrolle führen.

Andererseits sind Kontrolle und Verantwortung auch ohne vollständige Transparenz möglich. So können bei der Nutzung von Softwaresystemen, die auf Deep Learning-Ansätzen beruhen, Herstellern oder Anwendern die volle Verantwortung für den Einsatz dieser Systeme zugewiesen werden, selbst wenn ihnen die Details der Verarbeitung unbekannt sind. Sie trügen dann die Verantwortung, derartige Systeme zum Einsatz gebracht zu haben, und müssten begründen, warum diese Intransparenz akzeptabel ist – etwa weil der mögliche Schaden gering oder der zusätzliche Nutzen dieser Systeme (beispielsweise in Bezug auf eine höhere Genauigkeit der Prognosen) die Nachteile der Intransparenz überwiegt. So kann es einerseits sein, dass im medizinischen Kontext aus guten Gründen Software in der Krebsdiagnostik eingesetzt wird, deren

[435] Schlagworte dieser Debatte sind neben Transparenz (transparency) insbesondere auch Erklärbarkeit (explainability/explainable AI/explicability), Beobachtbarkeit (observability) und Nachvollziehbarkeit sowie Verantwortung und Haftung (responsibility, accountability und liability).

Prognosen durch diejenigen, die sie einsetzen, zwar nicht mehr nachvollziehbar erklärt werden können, deren höhere Genauigkeit jedoch diesen Nachteil überwiegt. Umgekehrt könnte in anderen Kontexten bzw. bei sensiblen Entscheidungen in der öffentlichen Verwaltung die Notwendigkeit der vollständig nachvollziehbaren Begründung von Entscheidungen dazu führen, dass intransparente Systeme nicht eingesetzt werden dürfen.

Anforderungen an Transparenz, Erklärbarkeit und Nachvollziehbarkeit sind dementsprechend in Abhängigkeit von den jeweiligen Zielen, die mit Transparenz und Nachvollziehbarkeit verfolgt werden, nach den Personen, die Informationen erhalten, und nach dem jeweiligen Anwendungskontext zu konkretisieren. Eine Nutzerin, die Auskunft darüber verlangt, warum sie keinen Kredit bekommen hat, bedarf einer anderen Art der Erklärung als eine Aufsichtsbehörde, die prüfen muss, ob eine Software in ihren Prognosen der Kreditwürdigkeit systematisch Frauen diskriminiert. Besonders hohe Anforderungen gelten für Systeme, die in hoch sensiblen Bereichen eingesetzt werden, beispielsweise bei Entscheidungen mit hoher Tragweite für das Leben von Menschen. Auch dort, wo Systeme eine Quasi-Monopolstellung erlangen, sind hohe Anforderungen an Transparenz, Erklärbarkeit und Nachvollziehbarkeit zu stellen, die möglicherweise den Einsatz von besonders intransparenten Systemen (beispielsweise basierend auf Deep Learning) ausschließen und nach nachvollziehbareren Verfahren verlangen (beispielsweise Entscheidungsbäumen). Bei der Entwicklung der Software wiederum müssen technische und organisatorische Voraussetzungen geschaffen und eingefordert werden, beispielsweise durch verpflichtende Dokumentations- und Offenlegungspflichten, damit diese Erklärungen später überhaupt erst geliefert werden können.

Die Spezifizierung und ausgewogene Ausgestaltung von Offenlegungspflichten stellt eine besondere Herausforderung dar; muss doch sichergestellt werden, dass einerseits relevante Informationen geteilt werden, aber andererseits weder die Geschäftsinteressen von Anbietern über Gebühr unterminiert werden, noch Flanken für Angriffe und Sicherheitslücken eröffnet werden.

<u>Empfehlung</u>

- *Empfehlung Querschnittsthema 10:* Es bedarf der Entwicklung ausgewogener aufgaben-, adressaten- und kontextspezifischer Standards für Transparenz, Erklärbarkeit und Nachvollziehbarkeit und ihrer Bedeutung für Kontrolle und Verantwortung sowie für deren Umsetzung durch verbindliche technische und organisatorische Vorgaben. Dabei muss den Anforderungen an Sicherheit und Schutz vor Missbrauch, Datenschutz sowie dem Schutz von intellektuellem Eigentum und Geschäftsgeheimnissen in angemessener Weise Rechnung

getragen werden. Je nach Kontext sind hier unterschiedliche Zeitpunkte (ex-ante, ex-post, real-time) sowie unterschiedliche Verfahren und Grade der Offenlegung zu spezifizieren.

10.11 Fazit

Im Rahmen dieser Stellungnahme wurden die Auswirkungen einer zunehmenden Delegation menschlicher Tätigkeiten an digitale Technologien, insbesondere KI-basierte Softwaresysteme, analysiert. In zahlreichen Beispielen aus den Bereichen der Medizin, der schulischen Bildung, der öffentlichen Kommunikation und Meinungsbildung sowie der öffentlichen Verwaltung zeigte sich, dass dieses Delegieren sowohl mit Erweiterungen als auch mit Verminderungen menschlicher Handlungsmöglichkeiten einhergeht und sich dadurch sowohl förderlich als auch hinderlich auf die Realisierung menschlicher Autorschaft auswirken kann. Eine Berücksichtigung dieser Auswirkungen sollte daher jedweder Entscheidung über eine Delegation menschlicher Tätigkeiten an Softwaresysteme – bis möglicherweise hin zu einer vollständigen Ersetzung des Menschen durch algorithmische Systeme – vorausgehen.

Ziel und Richtschnur ethischer Bewertung muss hierbei immer die Erhöhung menschlicher Autorschaft sein. Dabei ist zu berücksichtigen, dass die Erweiterung von Handlungsmöglichkeiten für eine Personengruppe mit deren Verminderung für andere einhergehen kann. Diesen Unterschieden ist Rechnung zu tragen, insbesondere in Hinblick auf den Schutz und die Verbesserung der Lebensbedingungen vulnerabler oder benachteiligter Gruppen. Die hier vorgelegte Analyse hat zahlreiche übergreifende Themen offengelegt, die in allen vier untersuchten Sektoren aufscheinen – wenn auch nicht immer in gleicher Art und Weise. Letztlich zeigt sich, dass die normativen Anforderungen an die Gestaltung und den Einsatz solcher Technologien, beispielsweise in Bezug auf Anforderungen hinsichtlich Transparenz und Nachvollziehbarkeit, den Schutz der Privatsphäre sowie die Verhinderung von Diskriminierung, zwar in allen Bereichen und für alle Betroffenen von hoher Bedeutung sind, sie jedoch sektor-, kontext-, und adressatenspezifisch konkretisiert werden müssen, um angemessen zu sein und wirksam werden zu können.

Mitglieder des Deutschen Ethikrates

Prof. Dr. med. Alena Buyx (Vorsitzende)

Prof. Dr. iur. Dr. h. c. Volker Lipp (Stellvertretender Vorsitzender)

Prof. Dr. phil. Dr. h. c. Julian Nida-Rümelin (Stellvertretender Vorsitzender)

Prof. Dr. rer. nat. Susanne Schreiber (Stellvertretende Vorsitzende)

Prof. Dr. iur. Steffen Augsberg

Regionalbischöfin Dr. phil. Petra Bahr

Prof. Dr. theol. Franz-Josef Bormann

Prof. Dr. rer. nat. Hans-Ulrich Demuth

Prof. Dr. iur. Helmut Frister

Prof. Dr. theol. Elisabeth Gräb-Schmidt

Prof. Dr. rer. nat. Dr. phil. Sigrid Graumann

Prof. Dr. rer. nat. Armin Grunwald

Prof. Dr. med. Wolfram Henn

Prof. Dr. rer. nat. Ursula Klingmüller

Stephan Kruip

Prof. Dr. theol. Andreas Lob-Hüdepohl

Prof. Dr. phil. habil. Annette Riedel

Prof. Dr. iur. Stephan Rixen

Prof. Dr. iur. Dr. phil. Frauke Rostalski

Prof. Dr. theol. Kerstin Schlögl-Flierl

Dr. med. Josef Schuster

Prof. Dr. phil. Mark Schweda

Prof. Dr. phil. Judith Simon

Prof. Dr. phil. Muna Tatari

Unterausschuss
Globale Gesundheit

Deutscher Bundestag

Deutscher Bundestag
Ausschuss f. Gesundheit

Ausschussdrucksache
20(14)126

04.07.2023

Überblick Sitzungen und Fachgespräche 20. Wahlperiode

Stand: 4. Juli 2023

Sitzungen		
Termin	**Thema**	**Gesprächspartner**
1.Sitzung 6. April 2022	**Konstituierung**	
2. Sitzung 25.04.2022	**Bericht der Bundesregierung zu den G7-Aktivitäten zur Globalen Gesundheit**	**PStSn Sabine Dittmar** (BMG) **PStSn Dr. Bärbel Kofler** (BMZ) **Andreas Bergfelder** (AA)
	Bericht der WHO zur Gesundheitsversorgung in der Ukraine	**Dr. Gerald Rockenschaub** (WHO) **Dr. Jarno Habicht** (WHO)
3. Sitzung 16.05.2022	**Bericht der Bundesregierung zur Umsetzung der Strategie zur globalen Gesundheit**	**PStS Prof. Dr. Edgar Franke** (BMG) **Paul Zubeil** (BMG) **Dr. Tania Rödiger-Vorwerk** (BMZ)
	Prioritäten nichtstaatlicher Akteure in der globalen Gesundheit (Akteursgruppen des Global Health Hub Germany, GHHG)	**Roland Göhde** (GHHG) **Paul Schnase** (GHHG) **Fiona Uellendahl** (GHHG) **Maike Voss** (GHHG) **Kristina Knispel** (GHHG) **Sarah Splettstößer** (GHHG)

	Wahl der Mitglieder des Lenkungskreises des Global Health Hub Germany (GHHG)	
4.Sitzung 20.06.2022	**Bericht der Bundesregierung zur Weltgesundheitsversammlung** **Bericht zur Weiterentwicklung des** WHO Hub for Pandemic and Epidemic Intelligence **Vorstellung des WHO-Netzwerks für Parlamentarier**	**Björn Kümmel** (BMG) **Dr. Bernhard Braune** (BMZ) **Dr. Chikwe Ihekweazu** (WHO Hub for Pandemic and Epidemic Intelligence) **Dr. Gaudenz Silberschmidt** (WHO)
5. Sitzung 19.09.2022	**Sekundäre Effekte der Covid-19-Pandemie auf die Gesundheit im Globalen Süden** **Weiterentwicklung der globalen Gesundheitsfinanzierung**	**Aidan O'Leary** (Global Polio Eradiction Initiative (GPEI)) **Bruno Rivalen** (Globale Financing Facility(GFF)) **Dr. Johannes Hunger** (Globale Fonds zur Bekämpfung von AIDS, Tuberkulose und Malaria (GFATM)) **Florence Bull** (Christian Health association, Sierra Leone) **Tilmann Rüppel** (Missionsärztliches Institut Würzburg) **Dr. Christoph Benn** (Joep Lange Institute)
6. Sitzung 17.10.2022	**Gespräch mit Expertinnen zu den Folgen des Klimawandels und ökologischer Krise für die Gesundheit**	**Sophie Gepp** (Centre for Planetary Health Policy) **Prof. Dr. Claudia Traidl-Hoffmann** (Universität Augsburg,

		Helmholtz Zentrum München) **Dr. Catharina Boehme** (WHO) **Dr. Anna Babette Stier** (BMG) **Dr. Bernhard Braune** (BMZ)
7. Sitzung 28.11.2022	**Bericht der Bundesregierung zu aktuellen Entwicklungen der Globalen Gesundheit** **Gespräch zu der globalen Arzneimittelversorgung und Lieferketten im Globalen Süden** **Bericht zur internationalen Forschungsvernetzung**	**Ingo Behnel** (BMG) **Dr. Tania Rödiger-Vorwerk** (BMZ) **Esteban Burrone** (Medicines Patent Pool) Martin H. Friede (WHO) **Dr. Richard Lessells** (Universität KwaZulu-Natal)
8. Sitzung 12.12.2022	**Bericht der Bundesregierung zu aktuellen Entwicklungen in der Globalen Gesundheit** **Bericht der Bundesregierung zur G7/G20 Gesundheitspolitik – Die Globale Gesundheitsarchitektur vor dem Hintergrund einer Vorbereitung auf die Pandemien der Zukunft** **Bericht zur Weiterentwicklung der globalen Gesundheitsarchitektur**	**PStS Niels Annen** (BMZ) **Helga Barth** (AA) **PStS Niels Annen** (BMZ) **StS Dr. Thomas Steffen** (BMG) **Prof. Dr. Ilona Kickbusch** (Graduate Institute of International and Development Studies Geneva) **Dr. Ayoade Alakija** (WHO)
9. Sitzung 23.01.2023	**Bericht der Bundesregierung zu aktuellen Entwicklungen in der Globalen Gesundheit**	**Dr. Tania Rödiger-Vorwerk** (BMZ)

	Forschung und Innovation für vernachlässigte und armutsassoziierte Krankheiten	**Prof. Dr. Jürgen May** (Bernhard-Nocht-Institut für Tropenmedizin (BNITM)) **Joelle Tanguay** (Drugs for Neglected Diseases initiative (DNDi))
	Forschung und Arzneimittel-Produktion. Welche Lehren sind aus der Covid 19-Pandemie für zukünftige Pandemien zu ziehen?	**Dr. Richard Hatchett** (Coalition for Epidemic Preparedness and Innovation (CEPI)) **Lisa Hedman** (WHO)
10. Sitzung 27.02.2023	**Bericht zur Lage in der Türkei und in Syrien nach dem Erdbeben**	**Helga Barth** (AA)
	Bericht der Bundesregierung zu aktuellen Entwicklungen in der Globalen Gesundheit	**Dr. Tania Rödiger-Vorwerk** (BMZ) **Dagmar Reitenbach** (BMG)
	Genderbasierte Gewalt und sexuelle und reproduktive Gesundheit und Rechte (SRGR) in konfliktbetroffenen und fragilen Kontexten	**Dr. Pascale Allotey** (WHO) **Dr. Claudia Garcia-Moreno** (WHO) **Mathilde Boddaert** (WHO)
	Geschlechtergerechte Ansätze zur Erreichung von SDG 3	**Dr. Julitta Onabanjo** (United Nations Population Fund, UNFPA)
11. Sitzung 27.03.2023	**Bericht von Prof. Dr. Axel Radlach Pries** (Präsident World Health Summit) **zum World Health Summit Regional Meeting und World Health Summit 2023**	**Prof. Dr. Axel Radlach Pries** (World Health Summit)
	Globale und nationale Herausforderungen und Lösungsansätze bei der Bekämpfung multiresistenter Tuberkulose/ Die globale Perspektive	**Dr. Lucica Ditiu** (Stop TB Partnership) **Meike Schwarz** (Ärzte ohne Grenzen) **Dr. Mel Spigelman** (TB Alliance)

	Globale und nationale Herausforderungen und Lösungsansätze bei der Bekämpfung multiresistenter Tuberkulose/ Die nationale Perspektive	**Prof. Dr. Torsten Bauer** (Deutsches Zentralkomitee zur Bekämpfung der Tuberkulose (DZK) **Prof. Dr. Michael Hoelscher** (LMU München) **Prof. Dr. Dr. Timo Ulrichs** (Akkon Hochschule)
12. Sitzung 24.04.2023	**Beteiligung Taiwans an der WHO und der Weltgesundheitsversammlung**	**Andreas Bergfelder** (AA)
	Bericht der Bundesregierung zu aktuellen Entwicklungen in der Globalen Gesundheit	**PStS Prof. Dr. Edgar Franke** (BMG)
	75 Jahre Weltgesundheitsorganisation	**Dr. Catharina Böhme** (WHO) **Prof. Dr. Anna Holzscheiter** (Technische Universität Dresden)
	Pandemievertrag und Reform der Internationalen Gesundheitsvorschriften	**Steven Solomon** (WHO) **Dr. Clare Wenham** (The London School of Economics and Social Science) **Dr. Viviana Munoz-Tellez** (The South Centre)
13. Sitzung 22.05.2023	**Bericht der Bundesregierung zu aktuellen Entwicklungen in der Globalen Gesundheit**	**Ingo Behnel** (BMG) **Dr. Tania Rödiger-Vorwerk** (BMZ)
	Gesundheitsfachkräfte	**Dr. Tine Hanrieder** (London School of Economics and Political Science, LSE) **Mousa Othman** (Faire Integration) **Norbert Grote** (Bundesverband privater

		Anbieter sozialer Dienste e. V., bpa)
	Geplant: Europa als globaler Gesundheitsakteur	**Francisco Pérez-Cañado** (Europäische Kommission, Generaldirektion Gesundheit und Lebensmittelsicherheit) **abgesagt**
14. Sitzung 19.06.2023	**Bericht der Bundesregierung zu aktuellen Entwicklungen in der Globalen Gesundheit, Schwerpunkt: High Level Meetings der Generalversammlung der Vereinten Nationen zu Universal Health Coverage“ (UHC), Tuberkulose (TB) und Pandemieprävention, -vorsorge und -bewältigung (PPR) und SDG-Summit**	**Helga Barth** (AA)
	Luftverschmutzung als globale Gesundheitsgefahr	**Prof. Dr. med. Babette Simon** (Universität Paris) **Prof. Dr. med. Barbara Hoffmann** ((Universität Duisburg-Essen) **Dr. Dorota Jarosińska** (WHO)
	Hitze als globale Gesundheitsgefahr	**Dr. Hedi Katre Kriit** (Umeå University Institute of Global Health (HIGH) and Interdisciplinary Center for Scientific Computing (IWR), Universität Heidelberg) **Dr. med. Susanne Johna** (Bundesärztekammer)

Fachgespräche		
Termin	**Thema**	**Gesprächspartner**
20.06.2022	**Austausch mit VertreterInnen des Access to COVID-19 Tools Accelerator (ACT-A) zur Eindämmung der COVID-19-Pandemie und den aktuellen Herausforderungen**	**Bruce Aylward (**WHO) **Seth Berkley** (Global Alliance for Vaccines and Immunisation (GAVI)) **Ted Chaiban** (Global Lead Coordinator for COVID Vaccine Country Readiness and Delivery) **Robert Matiru** (Unitaid) **Thabani Maphosa** (GAVI) **Bill Rodriguez** (Foundation for Innovative New Diagnostics (FIND)) **Diane Stewart** (Globaler Fonds für HIV, Tuberkulose und Malaria (GFATM) **Dr. Fifa Rahman** (ACT-A)
17.10.2022	**Gemeinsames Fachgespräch mit der WHO und internationalen ParlamentarierInnen zur nachhaltigen Finanzierung der WHO**	**Dr. Tedros Adhanom Ghebreyesus** (WHO-Generaldirektor) **Dr. Michael Ryan** (WHO) **Dr. Gaudenz Silberschmidt** (WHO)
18.10.2022	**Gemeinsames Fachgespräch mit der WHO und internationalen ParlamentarierInnen zu Global Health for Peace und Schutz vulnerabler Gruppen in fragilen Kontexten**	**Pascale Allotey** (WHO) **Mathilde Boddaert** (WHO)

Deutscher Bundestag

20. Wahlperiode

Rechtsausschuss

Deutscher Bundestag
Rechtsausschuss

Ausschussdrucksache
20(6)60

3. Juli 2023

Drucksache 20/[…]

[Datum]

Änderungsantrag

der Abgeordneten Dr. Lars Castellucci, Ansgar Heveling, Dr. Kirsten Kappert-Gonther, Dr. Konstantin von Notz, Petra Pau, Stephan Pilsinger, Benjamin Strasser, Kathrin Vogler und weiterer Abgeordneter

zu dem Gesetzentwurf der Abgeordneten Dr. Lars Castellucci, Ansgar Heveling, Dr. Kirsten Kappert-Gonther, Dr. Konstantin von Notz, Petra Pau, Stephan Pilsinger, Benjamin Strasser, Kathrin Vogler und weiterer Abgeordneter
– Drucksache 20/904 –

Entwurf eines Gesetzes zur Strafbarkeit der geschäftsmäßigen Hilfe zur Selbsttötung und zur Sicherstellung der Freiverantwortlichkeit der Entscheidung zur Selbsttötung

Der Bundestag wolle beschließen:

den Gesetzentwurf mit folgenden Maßgaben, im Übrigen unverändert anzunehmen:

1. Artikel 1 wird wie folgt gefasst:

Artikel 1

Änderung des Strafgesetzbuches

§ 217 des Strafgesetzbuches in der Fassung der Bekanntmachung vom 13. November 1998 (BGBl. I S. 3322), das zuletzt durch Artikel 4 des Gesetzes vom 4. Dezember 2022 (BGBl. I S. 2146) geändert worden ist, wird wie folgt gefasst:

„§ 217

Geschäftsmäßige Förderung der Selbsttötung

(1) Wer in der Absicht, die Selbsttötung einer anderen Person zu fördern, dieser hierzu geschäftsmäßig die Gelegenheit gewährt, verschafft oder vermittelt, wird mit Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft.

(2) Die Förderungshandlung im Sinne des Absatzes 1 ist nicht rechtswidrig, wenn

1. die zur Selbsttötung entschlossene Person volljährig und einsichtsfähig ist,
2. nach der fachlichen Überzeugung einer dritten Person mit fachärztlicher Qualifikation der Fachrichtungen Psychiatrie oder Psychotherapie oder mit psychotherapeutischer Qualifikation, die nicht an der Selbsttötung beteiligt ist, keine die autonome Entscheidungsfindung beeinträchtigende psychische Erkrankung vorliegt und nach fachlicher Überzeugung das Sterbeverlangen freiwilliger, ernsthafter und dauerhafter Natur ist; dies setzt in der Regel mindestens zwei Termine mit einem Mindestabstand von drei Monaten voraus,
3. mindestens ein individuell angepasstes, umfassendes und ergebnisoffenes Beratungsgespräch nach Maßgabe der dritten Person gemäß Nummer 2 mit einem multiprofessionellen und interdisziplinären Ansatz bei einem weiteren Arzt oder einer weiteren Ärztin, einem Psychotherapeuten oder einer Psychotherapeutin, einer psychosozialen Beratungsstelle, einer Suchtberatung oder einer Schuldenberatung stattgefunden hat, welches mindestens die folgenden Punkte umfasst:
   a) Aufklärung über den mentalen und physischen Zustand,
   b) Möglichkeiten der medizinischen Behandlung und Alternativen zur Selbsttötung,
   c) Hinweis auf weitere Beratungsmöglichkeiten und
   d) mögliche psychologische und physische Auswirkungen eines fehlgeschlagenen Selbsttötungsversuchs sowie soziale Folgen einer durchgeführten Selbsttötung,
4. zwischen der letzten Untersuchung gemäß Nummer 2, die auf das letzte Beratungsgespräch gemäß Nummer 3 folgt, und der Selbsttötung eine Wartefrist von mindestens zwei Wochen liegt und die Selbsttötung höchstens zwei Monate nach der letzten psychiatrischen oder psychotherapeutischen Untersuchung nach Nummer 2 erfolgt.

Wenn nach der fachlichen Überzeugung der dritten Person gemäß Satz 1 Nummer 2 die Durchführung von zwei Untersuchungsterminen für die zur Selbsttötung entschlossene Person nicht zumutbar ist, insbesondere bei Vorliegen einer nicht heilbaren, fortschreitenden und weit fortgeschrittenen Erkrankung bei einer zugleich begrenzten Lebenserwartung, und von einer weiteren Untersuchung offensichtlich keine weitere Erkenntnis zur Freiwilligkeit, Ernsthaftigkeit und Dauerhaftigkeit des Sterbeverlangens zu erwarten ist, kann die Feststellung nach Satz 1 Nummer 2 bereits nach einem Untersuchungstermin erfolgen. Alle benannten Punkte sind zu dokumentieren.

(3) Als Teilnehmer bleibt straffrei, wer selbst nicht geschäftsmäßig handelt und entweder Angehöriger oder Angehörige der in Absatz 1 genannten Person ist oder dieser nahesteht.“ ‘

2. Nach Artikel 2 werden die folgenden Artikel 3 bis 6 eingefügt:

Artikel 3

Änderung des Heilmittelwerbegesetzes

Nach § 1 Absatz 1 Nummer 2 Buchstabe c des Heilmittelwerbegesetzes in der Fassung der Bekanntmachung vom 19. Oktober 1994 (BGBl. I S. 3068), das zuletzt durch Artikel 2 des Gesetzes vom 11. Juli 2022 (BGBl. I S. 1082) geändert worden ist, wird folgender Buchstabe d eingefügt:

„d) auf Hilfe zur Selbsttötung,“.

Artikel 4

Änderung des Elften Buchs Sozialgesetzbuch

Dem § 11 des Elften Buches Sozialgesetzbuch – Soziale Pflegeversicherung – (Artikel 1 des Gesetzes vom 26. Mai 1994, BGBl. I S. 1014, 1015), das zuletzt durch Artikel 2a des Gesetzes vom 19. Juni 2023 (BGBl. 2023 I Nr. 155) geändert worden ist, wird folgender Absatz 4 angefügt:

„(4) Förderungshandlungen zur Selbsttötung nach § 217 Absatz 1 des Strafgesetzbuches stellen keine Leistungen im Sinne des Dritten Kapitels des Fünften Buches dar. Träger von Diensten und Einrichtungen des Gesundheits- und Sozialwesens sind vorbehaltlich der Bestimmungen des Wohn- und Betreuungsvertragsgesetzes nicht verpflichtet, an einer Selbsttötung mitzuwirken oder die Durchführung von Förderungshandlungen zur Selbsttötung in ihren Räumlichkeiten zu dulden.“

Artikel 5

Änderung des Wohn- und Betreuungsvertragsgesetzes

Dem § 3 des Wohn- und Betreuungsvertragsgesetz vom 29. Juli 2009 (BGBl. I S. 2319), das zuletzt durch Artikel 12 des Gesetzes vom 30. November 2019 (BGBl. I S. 1948) geändert worden ist, wird folgender Absatz 6 angefügt:

„(6) Der Unternehmer ist berechtigt, in der Darstellung seines Leistungsangebots darauf hinzuweisen, dass er Förderungshandlungen zur Selbsttötung nach § 217 Absatz 1 des Strafgesetzbuches weder selbst vornimmt noch deren Vornahme oder Werbung hierfür in seinen Räumlichkeiten, die nicht dem Verbraucher als Wohnraum überlassen sind, duldet. Die Rechte des Verbrauchers in dem ihm vom Unternehmer überlassenen Wohnraum bleiben hiervon unberührt.“

Artikel 6

Evaluierung

Die Bundesregierung berichtet dem Deutschen Bundestag innerhalb von fünf Jahren nach Verkündung dieses Gesetzes über dessen Wirkungen.

3. Der bisherige Artikel 3 wird Artikel 7.

Berlin, den

Dr. Lars Castellucci, Ansgar Heveling, Dr. Kirsten Kappert-Gonther, Dr. Konstantin von Notz, Petra Pau, Stephan Pilsinger, Benjamin Strasser, Kathrin Vogler (…)

Begründung

Zu Artikel 1:

Neben redaktionellen Änderungen in § 217 StGB wird in dessen Abs. 1 Satz 1 Nr. 2 nun klargestellt, dass für die Frage, ob eine die autonome Entscheidungsfindung beeinträchtigende psychische Erkrankung vorliegt, die fachliche Überzeugung eines nicht an der Selbsttötung beteiligten Facharztes oder einer nicht an der Selbsttötung beteiligten Fachärztin für Psychiatrie und Psychotherapie oder eines nicht an der Selbsttötung beteiligten Psychotherapeuten oder einer nicht an der Selbsttötung beteiligten Psychotherapeutin maßgeblich ist. Gegenüber dem ursprünglichen Gesetzentwurf wird der Kreis derjenigen Personen, deren fachliche Überzeugung insoweit maßgeblich ist, um Psychotherapeuten und Psychotherapeutinnen erweitert. Liegt nach Überzeugung derjenigen Personen, deren fachliche Überzeugung insoweit maßgeblich ist, keine die autonome Entscheidungsfindung beeinträchtigende psychische Erkrankung vor, ist davon auszugehen, dass die zur Selbsttötung entschlossene Person einsichtsfähig im Sinne von § 217 Abs. 2 Satz 1 Nr. 1 StGB ist. Dementsprechend wird auch in § 217 Abs. 2 Satz 2 StGB klargestellt, dass für die Frage, ob die Durchführung von zwei Untersuchungsterminen für die zur Selbsttötung entschlossene Person zumutbar ist, die fachliche Überzeugung des untersuchenden Facharztes oder der untersuchenden Fachärztin für Psychiatrie und Psychotherapie oder der untersuchenden Psychotherapeutin oder des untersuchenden Psychotherapeuten gemäß Satz 1 Nr. 2 maßgeblich ist. Dadurch ist ein klares abgekürztes Verfahren für die Situation vorgesehen, dass die zur Selbsttötung entschlossene Person sich in der Situation einer schweren und belastenden Erkrankung mit voraussichtlich begrenzter Lebenserwartung befindet.

Zu Artikel 3 – neu –

Die Änderung des Heilmittelwerbegesetzes erfolgt im Gegenzug zur Streichung von § 217a StGB (Werbung für die Hilfe zur Selbsttötung im ursprünglichen Gesetzentwurf).

Zu Artikel 4 – neu – und Artikel 5 – neu –

Durch die Änderung des Elften Buchs Sozialgesetzbuch sowie des Wohn- und Betreuungsvertragsgesetzes wird klargestellt, dass Einrichtungen des Gesundheits- und Sozialwesens grundsätzlich nicht verpflichtet sind, an einer Selbsttötung mitzuwirken oder die Durchführung von Förderungshandlungen zur Selbsttötung in ihren Räumlichkeiten zu dulden. Die Einrichtungen haben als Unternehmer im Sinne des Wohn- und Betreuungsvertragsgesetzes die Möglichkeit, dies gegenüber dem Verbraucher bzw. der Verbraucherin im Einzelnen vertraglich auszugestalten. Eine vertragliche Einschränkung der Möglichkeit, in Räumlichkeiten der Einrichtungen des Gesundheits- und Sozialwesens Hilfe zur Selbsttötung in Anspruch zu nehmen, ist durch den hohen Rang, den das Grundgesetz dem Leben und der Autonomie beimisst, grundsätzlich gerechtfertigt, zumal ihnen im Bereich der Suizidhilfe besondere Gefahren drohen.

Zu Artikel 6 – neu –

Durch den neuen Artikel 6 wird eine Evaluierungspflicht eingeführt. Dadurch wird der Bundestag in die Lage versetzt, die Durchführung und die Folgen des vorliegenden Gesetzes sowie daraus eventuell folgenden Anpassungsbedarf nach Ablauf von fünf Jahren auf Basis des Evaluierungsberichts zu erörtern.

Deutscher Bundestag
Rechtsausschuss

Ausschussdrucksache
20(6)63

4. Juli 2023

Änderungsantrag der Initianten der Gesetzesentwürfe auf Drucksache 20/2332 und 20/2293 zum Zwecke der Zusammenführung in der nachstehenden Fassung:

Entwurf eines Gesetzes zum Schutz des Rechts auf selbstbestimmtes Sterben und zur Regelung der Hilfe zur Selbsttötung sowie zur Änderung weiterer Gesetze

Vom …

Der Bundestag hat mit Zustimmung des Bundesrates das folgende Gesetz beschlossen:

Artikel 1

Gesetz zum Schutz des Rechts auf selbstbestimmtes Sterben und zur Regelung der Hilfe zur Selbsttötung

(Suizidhilfegesetz – ShG)

§ 1

Recht auf Hilfe zur Selbsttötung

(1) Jeder, der aus autonom gebildetem, freiem Willen sein Leben eigenhändig beenden möchte, hat das Recht, hierbei Hilfe in Anspruch zu nehmen.

(2) Dieses Gesetz soll eine unwürdige, unzumutbare und nicht von freiem Willen getragene Umsetzung des Sterbewunsches verhindern sowie eine autonome und vollinformierte Entscheidungsfindung suizidwilliger Personen sicherstellen. Es eröffnet suizidwilligen Personen einen sicheren Zugang zu Arznei- und Betäubungsmitteln zum Zwecke der Selbsttötung.

§ 2

Recht zur Hilfeleistung

(1) Jeder darf einem anderen, der aus autonom gebildetem, freiem Willen sein Leben eigenhändig beenden möchte, auf dessen Wunsch Hilfe zur Selbsttötung leisten und ihn bis zum Eintritt des Todes begleiten.

(2) Niemand kann verpflichtet werden, Hilfe zur Selbsttötung zu leisten.

(3) Keiner Person darf aufgrund ihrer Berufszugehörigkeit untersagt werden, Hilfe zur Selbsttötung zu leisten oder die Leistung von Hilfe zur Selbsttötung zu verweigern. Beschäftigte dürfen nicht benachteiligt werden, weil sie Hilfe zur Selbsttötung leisten, ihre Bereitschaft zur Leistung von Hilfe zur Selbsttötung bekunden oder die Leistung von Hilfe zur Selbsttötung verweigern.

§ 3

Autonom gebildeter, freier Wille

Ein autonom gebildeter, freier Wille liegt vor, wenn

1. die suizidwillige Person die Fähigkeit besitzt, ihren Willen frei und unbeeinflusst von einer akuten psychischen Störung bilden und nach dieser Einsicht handeln zu können,
2. der suizidwilligen Person alle entscheidungserheblichen Gesichtspunkte bekannt sind; hierzu ist erforderlich, dass sie über sämtliche Informationen verfügt, die sie befähigen, auf einer hinreichenden Beurteilungsgrundlage realitätsgerecht das Für und Wider einer Suizidentscheidung abzuwägen, und insbesondere, dass sie Handlungsalternativen zum Suizid kennt, ihre jeweiligen Folgen bewerten kann und ihre Entscheidung in Kenntnis aller erheblichen Umstände und Optionen trifft,

3. der Entschluss zur Selbsttötung ohne unzulässige Einflussnahme oder Druck gebildet worden ist und
4. der Entschluss zur Selbsttötung von einer gewissen Dauerhaftigkeit und inneren Festigkeit getragen ist.

§ 4

Beratung

(1) Jeder hat das Recht, sich zu Fragen der Hilfe zur Selbsttötung beraten zu lassen. Die Beratung ist ergebnisoffen zu führen, darf nicht bevormunden und muss vom Grundwert jedes Menschenlebens ausgehen. Eine Person mit akuten Suizidgedanken ist unverzüglich zu beraten.

(2) Die Beratung soll die für eine Entscheidung für oder gegen eine Selbsttötung erheblichen Gesichtspunkte vermitteln. Sie umfasst insbesondere Informationen über

1. die Bedeutung und die Tragweite der Selbsttötung,
2. Handlungsalternativen zur Selbsttötung, sofern die zu beratende Person in der Beratung ihrerseits eine Erkrankung angibt, auch über in Betracht kommende alternative therapeutische Maßnahmen sowie pflegerische oder palliativmedizinische Möglichkeiten,
3. die Möglichkeiten zur und Voraussetzungen für eine Inanspruchnahme von Hilfe zur Selbsttötung,
4. die Folgen einer Selbsttötung und eines fehlgeschlagenen Selbsttötungsversuches auch für das nähere persönliche und familiäre Umfeld,
5. Möglichkeiten der Inanspruchnahme von Unterstützungs- und Betreuungsangeboten sowie
6. alle weiteren nach Sachlage erforderlichen medizinischen, sozialen und juristischen Informationen.

(3) Die zu beratende Person kann auf ihren Wunsch gegenüber der sie beratenden Person anonym bleiben.

(4) Die Beratung darf nicht von einer Person vorgenommen werden, die an einer späteren Hilfe zur Selbsttötung beteiligt ist.

(5) Soweit erforderlich sind zur Beratung im Einvernehmen mit der zu beratenden Person andere Personen hinzuzuziehen, insbesondere ärztlich, fachärztlich, psychologisch, sozialpädagogisch, sozialarbeiterisch oder juristisch ausgebildete Fachkräfte und nahe Angehörige.

§ 5

Beratungsstellen

(1) Für die Beratung nach § 4 haben die Länder ein ausreichendes Angebot an Beratungsstellen sicherzustellen. Sie haben dafür Sorge zu tragen, dass für suizidwillige Personen mit körperlichen Einschränkungen ein aufsuchendes Beratungsangebot besteht. Die Beratungsstellen bedürfen besonderer staatlicher Anerkennung nach Absatz 2. Als Beratungsstellen können auch Einrichtungen freier Träger sowie Ärzte und Ärztinnen anerkannt werden.

(2) Eine Beratungsstelle darf nur anerkannt werden, wenn sie die Gewähr für eine fachgerechte Beratung nach § 4 bietet und zur Durchführung der Beratung in der Lage ist, insbesondere

1. über hinreichend persönlich und fachlich qualifiziertes und der Zahl nach ausreichendes Personal verfügt,
2. sicherstellt, dass zur Durchführung der Beratung erforderlichenfalls kurzfristig eine ärztlich, fachärztlich, psychologisch, sozialpädagogisch, sozialarbeiterisch oder juristisch ausgebildete Fachkraft hinzugezogen werden kann,
3. mit allen Stellen zusammenarbeitet, die öffentliche und private Hilfen für suizidwillige Personen gewähren, und
4. mit keiner Einrichtung, in der Hilfe zur Selbsttötung geleistet wird, derart organisatorisch oder durch wirtschaftliche Interessen verbunden ist, dass hiernach ein materielles Interesse der Beratungseinrichtung an der Durchführung von Hilfe zur Selbsttötung nicht auszuschließen ist.

(3) Die zur Sicherstellung eines ausreichenden Angebotes erforderlichen Beratungsstellen haben Anspruch auf eine angemessene öffentliche Förderung der Personal- und Sachkosten.

(4) Die Beratungsstelle hat nach Abschluss der Beratung der beratenen Person, sofern sie nicht nach § 4 Absatz 3 anonym bleiben möchte, eine mit Namen und Datum versehene Bescheinigung darüber auszustellen, dass eine Beratung stattgefunden hat. Hat die beratende Person begründete Zweifel daran, dass

die beratene Person ihre Suizidentscheidung aus autonom gebildetem, freiem Willen im Sinne des § 3 treffen wird, hat sie dies auf der Bescheinigung zu vermerken.

(5) Die Beratung ist für die beratene Person und die nach § 4 Absatz 5 hinzugezogenen Personen unentgeltlich.

(6) Die Beratungsstellen sind verpflichtet, die ihrer Beratungstätigkeit zugrunde liegenden Maßstäbe und die dabei gesammelten Erfahrungen jährlich in einem schriftlichen Bericht niederzulegen. Als Grundlage für den schriftlichen Bericht nach Satz 1 hat die beratende Person über jedes Beratungsgespräch eine Aufzeichnung zu fertigen. Diese darf keine Rückschlüsse auf die Identität der beratenen Person und der zum Beratungsgespräch hinzugezogenen weiteren Personen ermöglichen. Sie hält den wesentlichen Inhalt der Bera-tung und angebotene Hilfsmaßnahmen fest.

(7) Die zuständige Behörde hat mindestens im Abstand von drei Jahren zu überprüfen, ob die Voraussetzungen für die Anerkennung nach Absatz 2 noch vorliegen. Sie kann sich zu diesem Zweck die Berichte nach Absatz 6 Satz 1 vorlegen lassen und Einsicht in die nach Absatz 6 Satz 2 anzufertigenden Aufzeichnungen nehmen. Liegt eine der Voraussetzungen des Absatzes 2 nicht mehr vor, ist die Anerkennung zu widerrufen.

(8) Näheres regelt das Landesrecht.

§ 6

Verschreibung eines Arznei- oder Betäubungsmittels zum Zwecke der Selbsttötung

(1) Ein Arzt oder eine Ärztin darf einer volljährigen Person, die die deutsche Staatsangehörigkeit besitzt oder ihren Wohnsitz oder gewöhnlichen Aufenthalt in Deutschland hat und aus autonom gebildetem, freiem Willen im Sinne des § 3 ihr Leben eigenhändig beenden möchte, ein Arznei- oder Betäubungsmittel zum Zweck der Selbsttötung verschreiben.

(2) Der Arzt oder die Ärztin ist verpflichtet, die suizidwillige Person mündlich und in verständlicher Form über sämtliche für die Selbsttötung wesentlichen medizinischen Umstände aufzuklären. Dazu gehören insbesondere der voraussichtliche Ablauf der Selbsttötung und die Risiken der bei der Selbsttötung angewendeten medizinischen Methode. Im Rahmen der Aufklärung ist, sofern die suizidwillige Person unter einer Erkrankung leidet, auch auf Behandlungsmöglichkeiten und Möglichkeiten der Palliativmedizin hinzuweisen.

(3) Der Arzt oder die Ärztin hat sich durch Vorlage der Bescheinigung nach § 5 Absatz 4 nachweisen zu lassen, dass sich die suizidwillige Person höchstens zwölf Wochen vor der Verschreibung in einer Beratungsstelle nach § 5 hat beraten lassen.

(4) Der Arzt oder die Ärztin darf die Verschreibung erst vornehmen, wenn seit der Beratung mindestens drei Wochen vergangen sind.

(5) Der Arzt oder die Ärztin ist verpflichtet, sämtliche für die Beurteilung des Vorliegens eines autonom gebildeten, freien Willens wesentlichen Gesichtspunkte und die Aufklärung der suizidwilligen Person nach Absatz 2 zu dokumentieren. Die Beratungsbescheinigung nach § 5 Absatz 4 ist zur Dokumentation zu nehmen.

§ 7

Verschreibung eines Arznei- oder Betäubungsmittels zum Zwecke der Selbsttötung in Härtefällen

(1) Abweichend von § 6 Absatz 3 und 4 darf der Arzt oder die Ärztin in besonderen Härtefällen ein Arznei- oder Betäubungsmittel zum Zweck der Selbsttötung ohne Vorlage einer Bescheinigung nach § 5 Absatz 4 verschreiben, wenn ein zweiter Arzt oder eine zweite Ärztin, der oder die nicht der Weisung des anderen Arztes oder der anderen Ärztin unterliegt, ebenfalls zu der Einschätzung gelangt, dass ein besonderer Härtefall vorliegt. Die Einschätzung muss schriftlich dokumentiert werden und ist zur Dokumentation nach § 6 Absatz 5 Satz 1 zu nehmen.

(2) Ein besonderer Härtefall ist gegeben, wenn sich die suizidwillige Person gegenwärtig in einem existentiellen Leidenszustand mit anhaltenden Symptomen, die die Person in ihrer gesamten Lebensführung dauerhaft beeinträchtigen, befindet oder in absehbarer Zeit befinden wird, insbesondere bei Vorliegen einer nicht heilbaren, fortschreitenden und weit fortgeschrittenen Erkrankung und zugleich begrenzter Lebenserwartung.

§ 8

Erteilung einer Erlaubnis zum Erwerb eines Arznei- oder Betäubungsmittels zum Zweck der Selbsttötung

Eine nach Landesrecht zuständige Stelle hat einer suizidwilligen Person, die aus autonom gebildetem, freiem Willen im Sinne des § 3 ihr Leben eigenhändig beenden möchte, eine einer ärztlichen Verschreibung gleichstehende Erlaubnis zum Erwerb eines Arznei- oder Betäubungsmittels zum Zweck der Selbsttötung zu erteilen, sofern ein Arzt oder eine Ärztin oder eine andere bei der zuständigen Stelle beschäftigte gleichermaßen qualifizierte Person das Vorliegen der Voraussetzungen des § 6 Absatz 1 und 3 bestätigt, die suizidwillige Person entsprechend § 6 Absatz 2 aufgeklärt hat, die Verschreibung unter Einhaltung der Frist nach § 6 Absatz 4 erfolgt und die suizidwillige Person glaubhaft macht, dass eine Verschreibung nach § 6 Absatz 1 für sie nicht in zumutbarer Weise zu erlangen ist. § 6 Absatz 5 gilt für die zuständige Stelle entsprechend.

§ 9

Verordnungsermächtigung

Die Bundesregierung wird ermächtigt, durch Rechtsverordnung mit Zustimmung des Bundesrates das Nähere zur Evaluation, zu Meldepflichten und Dokumentation, zur sicheren Aufbewahrung der Arznei- oder Betäubungsmittel sowie der Vergütung der Hilfe zur Selbsttötung und der Prävention gegen die Etablierung rein auf Gewinnstreben ausgerichteter Angebote zu regeln. Die Zulassung organisierter Angebote von Hilfe zur Selbsttötung kann in der Rechtsverordnung nach Satz 1 von einer Zuverlässigkeitsprüfung abhängig gemacht werden. Die Bundesregierung kann die Ermächtigung nach Satz 1 durch Rechtsverordnung auf das Bundesministerium für Gesundheit und das Bundesministerium der Justiz übertragen.

§ 10

Berichtswesen, Evaluation

(1) Die Wirkungen dieses Gesetzes sind durch die Bundesregierung unter Heranziehung interdisziplinären externen Sachverstands zu evaluieren.

(2) Die Bundesregierung erstattet dem Bundestag spätestens drei Jahre nach Inkrafttreten dieses Gesetzes und anschließend regelmäßig im Abstand von drei Jahren einen schriftlichen Evaluationsbericht. Der Bericht soll die Entwicklungen in der Hilfe zur Selbsttötung, auch hinsichtlich potenzieller rein auf Gewinnstreben ausgerichteter Angebote, erfassen.

§ 11

Übergangsvorschrift

Bis ausreichend anerkannte Beratungsstellen im Sinne des § 5 eingerichtet sind, längstens in einem Zeitraum bis zum … [einsetzen: Datum zwei Jahren nach Inkrafttreten gemäß Artikel 6], darf jeder Arzt und jede Ärztin abweichend von § 5 Absatz 1 Satz 3, eine Beratung, die den Anforderungen des § 4 entspricht, vornehmen und eine Bescheinigung entsprechend § 5 Absatz 4 ausstellen, ohne dass es einer besonderen staatlichen Anerkennung im Sinne von § 5 Absatz 2 hierfür bedarf.

Artikel 2

Änderung des Betäubungsmittelgesetzes

Das Betäubungsmittelgesetz in der Fassung der Bekanntmachung vom 1. März 1994 (BGBl. I S. 358), das zuletzt durch Artikel 1 der Verordnung vom 1. Juni 2023 (BGBl. 2023 I Nr. 143) geändert worden ist, wird wie folgt geändert:

1. § 4 wird wie folgt geändert:
   a) Nach Absatz 1 Nummer 3 Buchstabe a wird folgender Buchstabe b eingefügt:
      „b) auf Grund einer nach § 8 des Suizidhilfegesetzes erteilten Erlaubnis,".
   b) Die bisherigen Buchstaben b und c werden die Buchstaben c und d.

2. § 12 wird wie folgt geändert:
   a) Nach Absatz 3 Nummer 1 Buchstabe a wird folgender Buchstabe b eingefügt:
      „b) auf Grund einer nach § 8 des Suizidhilfegesetzes erteilten Erlaubnis im Rahmen des Betriebes einer Apotheke,".
   b) Die bisherigen Buchstaben b und c werden die Buchstaben c und d.
3. § 13 wird wie folgt geändert:
   a) Nach Absatz 1 Satz 2 wird folgender Satz eingefügt:
      „Die §§ 6 bis 8 des Suizidhilfegesetzes bleiben unberührt."
   b) Absatz 2 wird wie folgt geändert:
      aa) In Satz 1 werden nach der Angabe „Absatz 1" die Wörter „oder nach den §§ 6 und 7 des Suizidhilfegesetzes" eingefügt.
      bb) Nach Satz 1 wird folgender Satz eingefügt:
      „Das gleiche gilt für Betäubungsmittel, zu deren Erwerb eine Erlaubnis nach § 8 des Suizidhilfegesetzes erteilt worden ist, mit der Maßgabe, dass statt der Verschreibung die nach § 8 des Suizidhilfegesetzes erteilte Erlaubnis vorzulegen ist."
4. In § 29 Absatz 1 Satz 1 Nummer 6 wird die Angabe „§ 13 Abs. 1" durch die Wörter „§ 13 Absatz 1 Satz 1 oder Satz 4" ersetzt.

Artikel 3

Änderung des Strafgesetzbuches

In § 203 Absatz 1 Nummer 5 des Strafgesetzbuches in der Fassung der Bekanntmachung vom 13. November 1998 (BGBl. I S. 3322), das zuletzt durch Artikel 4 des Gesetzes vom 4. Dezember 2022 (BGBl. I S. 2146) geändert worden ist, werden vor dem Komma am Ende die Wörter „oder § 5 des Suizidhilfegesetzes" eingefügt.

Artikel 4

Änderung des Arzneimittelgesetzes

Das Arzneimittelgesetz in der Fassung der Bekanntmachung vom 12. Dezember 2005 (BGBl. I S. 3394), das zuletzt durch Artikel 8c des Gesetzes vom 20. Dezember 2022 (BGBl. I S. 2793) geändert worden ist, wird wie folgt geändert:

1. In § 48 Absatz 1 Satz 1 werden im Satzteil vor der Aufzählung nach den Wörtern „tierärztlichen Verschreibung" die Wörter „oder einer nach § 8 des Suizidhilfegesetzes erteilten Erlaubnis" eingefügt.
2. In § 81 wird nach den Wörtern „des Anti-Doping-Gesetzes" ein Komma und werden die Wörter „des Suizidhilfegesetzes" eingefügt.

Artikel 5

Änderung des Heilmittelwerbegesetzes

Nach § 1 Absatz 1 Nummer 2 Buchstabe c des Heilmittelwerbegesetzes in der Fassung der Bekanntmachung vom 19. Oktober 1994 (BGBl. I S. 3068), das zuletzt durch Artikel 2 des Gesetzes vom 11. Juli 2022 (BGBl. I S. 1082) geändert worden ist, wird folgender Buchstabe d eingefügt:

„d) auf Hilfe zur Selbsttötung,".

Artikel 6

Inkrafttreten

Dieses Gesetz tritt am Tag nach der Verkündung in Kraft.

Berlin, den […]

Begründung

Allgemeiner Teil

Problem

Das allgemeine Persönlichkeitsrecht umfasst als Ausdruck persönlicher Autonomie ein Recht auf selbstbestimmtes Sterben. Zum Recht auf selbstbestimmtes Sterben gehört auch die Freiheit, sich das Leben zu nehmen und hierbei Hilfe in Anspruch zu nehmen, sofern sie angeboten wird. Es obliegt dem Gesetzgeber, selbstbestimmte Entscheidungen zu achten. Einen gegen die persönliche Autonomie gerichteten Lebensschutz kann und darf es nicht geben. So hat es das Bundesverfassungsgericht im Jahr 2020 in einem wegweisenden Urteil entschieden (Urteil vom 26.2.2020 – 2 BvR 2347/15 - BVerfGE 153, 182 ff.).

Die Entscheidung des Einzelnen, so das Bundesverfassungsgericht, dem eigenen Leben entsprechend dem individuellen Verständnis von Lebensqualität und Sinnhaftigkeit der eigenen Existenz ein Ende zu setzen, ist im Ausgangspunkt als Akt autonomer Selbstbestimmung von Staat und Gesellschaft zu respektieren. Maßgeblich sei der Wille des Grundrechtsträgers, der sich einer Bewertung anhand allgemeiner Wertvorstellungen, religiöser Gebote, gesellschaftlicher Leitbilder für den Umgang mit Leben und Tod oder Überlegungen objektiver Vernünftigkeit entziehe (a. a. O. Rn. 210).

Das vormalige strafbewehrte Verbot der geschäftsmäßigen Förderung der Selbsttötung in § 217 Absatz 1 des Strafgesetzbuches (StGB) war nichtig, weil es selbstbestimmt handelnden Menschen mit Sterbewunsch in Ermangelung real existenter anderweitiger Zugangswege zur Suizidhilfe faktisch unmöglich machte, ihren verfassungsrechtlich geschützten Wunsch in zumutbarer Weise zu verwirklichen.

Der derzeit unregulierte Zustand der Suizidhilfe erfordert gesetzgeberisches Handeln, um die unwürdige, unzumutbare und nicht von freiem Willen getragene Umsetzung von Sterbewünschen möglichst zu verhindern, eine autonome und vollinformierte Entscheidungsfindung suizidwilliger Personen sicherzustellen, ihnen einen sicheren Zugang zu Arznei-/Betäubungsmitteln zum Zwecke der Selbsttötung zu ermöglichen und so Rechtssicherheit und klare Regelungen für Suizidwillige und auf deren Wunsch zur Hilfe bereite Personen zu schaffen.

Lösung

Der vorliegende Gesetzentwurf sichert in seinem Artikel 1 (Suizidhilfegesetz) das Recht auf einen selbstbestimmten Tod und stellt klar, dass Hilfe zur eigenhändigen Selbsttötung straffrei und erlaubt ist. Das strafrechtliche Verbot der Tötung auf Verlangen (§ 216 StGB) bleibt unberührt.

Das Regelungskonzept schafft innerhalb des vom Bundesverfassungsgericht beschriebenen Normierungsspielraums einen klaren Rechtsrahmen für Menschen, die ernstlich sterben möchten und diesen Wunsch frei und eigenverantwortlich im Vollbesitz ihrer geistigen Kräfte gebildet haben, sowie für Personen, die bereit sind, auf Wunsch von Sterbewilligen dabei zu helfen. Dazu formuliert der Entwurf Voraussetzungen, unter denen sich Menschen zukünftig auf eine Begleitung bis zum Lebensende verlassen können und schafft sichere Zugangsmöglichkeiten zu Arznei-/Betäubungsmitteln zur Selbsttötung. Dabei wird, wie es das Bundesverfassungsgericht explizit für zulässig gehalten hat (a.a.O. Rn.340), auch nach Lebenssituationen Betroffener differenziert. Es macht also in den Voraussetzungen einen Unterschied, ob der Tod z.B. wegen einer weit fortgeschrittenen unheilbaren Erkrankung oder aus anderen Gründen angestrebt wird.

Zugleich wird der Entwurf dem staatlichen Schutzauftrag gerecht. Durch Fest-legung von Entscheidungskriterien für Ärztinnen und Ärzte sowie klaren Verfahrensregeln einschließlich einer Beratungspflicht schafft der Entwurf einen ausgewogenen Ausgleich zwischen der Gewährleistung des Rechts auf selbstbestimmtes Sterben einerseits und dem Schutz der Autonomie Suizidwilliger andererseits. Die Verschreibungsmöglichkeit wird unter anderem an das Kriterium der Volljährigkeit geknüpft, wodurch der Gesetzgeber seinem besonderen Schutzauftrag gegenüber Minderjährigen nachkommt.

In den folgenden Artikeln schließen sich notwendige Änderungen des Betäubungsmittelgesetzes (Öffnung für Verschreibungsmöglichkeit zum Zwecke der Selbsttötung) und des Arzneimittelgesetzes an sowie des Strafgesetzbuches (Schutz von Privatgeheimnissen der in Beratungsstellen zum Thema Suizid Ratsuchenden) und des Heilmittelwerbegesetzes (Schutz vor unzulässiger Werbung).

Der dringliche bedarfsdeckende Ausbau palliativmedizinischer Versorgung bleibt weiteren gesetzgeberischen Schritten vorbehalten. Dazu wird auf den parallel eingebrachten Entschließungsantrag verwiesen.

Zielsetzung und Notwendigkeit der Regelungen

Das allgemeine Persönlichkeitsrecht umfasst nach der Rechtsprechung des Bundesverfassungsgerichts (BVerfGE 153, 182 ff.) als Ausdruck persönlicher Autonomie ein Recht auf selbstbestimmtes Sterben (a. a. O. Leitsatz 1a.). Die Achtung und der Schutz der Menschenwürde und der Freiheit sind grundlegende Prinzipien der grundgesetzlichen Verfassungsordnung, die den Menschen als eine zu Selbstbestimmung und Eigenverantwortung fähige Persönlichkeit begreift. Im Mittelpunkt dieses Gesetzentwurfs stehen daher Menschen, die sterben möchten, ihre Angehörigen und Freunde sowie Menschen, die helfen möchten.

Dem ernstlichen und dauerhaften Wunsch von Menschen, ihrem Leben selbstbestimmt ein Ende zu setzen, einerseits mit Respekt zu begegnen und andererseits die Achtung der Selbstbestimmung mit dem Schutz dieser Menschen vor einer Beeinflussung durch Dritte in Einklang zu bringen, ist eine gesetzgeberische Aufgabe.

Menschen mit Suizidwunsch werden derzeit weitgehend allein gelassen, sei es am Anfang eines Willensbildungsprozesses, wenn bereits ein explizierter Sterbewunsch besteht oder auch wenn die Entscheidung, das eigene Leben beenden zu wollen, abschließend gefasst ist. Ungeachtet der Lebenssituation gilt jedoch: Menschen, die Suizidgedanken haben oder sich dem Thema Sterben und Suizidhilfe widmen, verdienen gesellschaftliche Beachtung und Unterstützung. Sie sollen sich des Schutzes und der Fürsorge der Gesellschaft sicher sein, die eine solche Entscheidung in Ausübung des individuellen Selbstbestimmungsrechts respektiert. Denn Fakt ist: Suizide sind Teil unserer gesellschaftlichen Realität.

Aber obwohl das Bundesverfassungsgericht und auch drei Jahre zuvor das Bundesverwaltungsgericht (BVerwG, Urteil vom 2.3.2017 – BVerwG 3 C 19.15 - BVerwGE 158, 142) sich zweifelsfrei hinter das Selbstbestimmungsrecht des Einzelnen gestellt haben, bestehen praktische Rechtsunsicherheiten und faktische Hürden, die mit dem Gesetzentwurf beseitigt werden sollen. Deshalb soll legislativ klargestellt werden, dass die Hilfe zur von einem autonom gebildeten, freien Willen ausgehenden Selbsttötung regelmäßig straffrei ist.

In diesem Kontext ist zu betonen, dass die Regelung zur Strafbarkeit wegen Tötung auf Verlangen (§ 216 StGB) durch diesen Entwurf nicht berührt wird. Wer einen anderen Menschen tötet, wird weiterhin auch dann wegen eines Tötungsdelikts bestraft, wenn er auf Verlangen des Opfers handelt. Deshalb wird in Artikel 1 (Suizidhilfegesetz) §§ 1 Abs. 1, 2 Abs. 1, 6 Abs. 1 und 8 Satz 1 jeweils vorgegeben, dass die Umsetzung des Sterbewunsches im Sinne dieses Gesetzes eigenhändig erfolgen muss: Der oder die Sterbewillige muss die zuletzt handelnde Person bleiben. Verschreibungen nach dem Suizidhilfegesetz berechtigen den Arzt oder die Ärztin oder andere Beteiligte nicht zu einer hiervon abweichenden Verabreichung des zum Tode führenden Arznei-/Betäubungsmittels. Bei Suizidwilligen, die zu Eigenhändigkeit im Wortsinne nicht in der Lage sind, kann die Möglichkeit einer technischen Unterstützung in Betracht kommen.

Kein Handlungsbedarf besteht auch mit Blick auf die Rechtslage zur straffreien Sterbehilfe in den Fällen einer Inkaufnahme eines früheren unbeabsichtigten Todeseintritts bei einem sterbenden oder todkranken Menschen infolge einer medizinisch indizierten schmerz- oder in sonstiger Weise leidensmindernden Therapie und einer aktiven oder passiven Begrenzung oder Beendigung einer lebenserhaltenden oder lebensverlängernden medizinischen Maßnahme im Einklang mit dem tatsächlichen oder mutmaßlichen Patientenwillen.

Es ist allerdings nicht hinnehmbar, dass es für freiverantwortlich handelnde Betroffene trotz des grundgesetzlich verankerten Rechts auf einen selbstbestimmten Tod keine Möglichkeit gibt, in einem geordneten Verfahren Zugang zu Arznei-/Betäubungsmitteln zur Selbsttötung zu erlangen. Trotz der

Feststellung des Bundesverwaltungsgerichts im Jahr 2017, dass unheilbar kranken Patienten und Patientinnen im Extremfall der Anspruch auf Medikamente zur schmerzlosen Selbsttötung nicht verwehrt werden darf (BVerwG Urteil vom 02.03.2017 - 3 C 19.15 - BVerwGE 158, 142), ist von den im Zeitraum zwischen März 2017 und August 2021 beim Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM) gestellten 223 Anträgen auf Erwerb eines tödlich wirkenden Betäubungsmittels kein Antrag bewilligt worden (BT-Drs. 19/32360, 6).

Häufig sind die gewählten Suizid-Methoden sowohl bei Frauen als auch bei Männern risikoreiche, unsichere oder schmerzhafte Methoden (wie z.B. Erhängen oder auch der Verzicht auf Essen und Trinken). Menschen, die sterben möchten, sollten aber nicht auf solche Möglichkeiten der Selbsttötung verwiesen werden, wenn es Mittel gibt, die nach dem Stand der Wissenschaft regelmäßig einen sanfteren und würdevolleren Tod ermöglichen und auch im Ausland genutzt werden dürfen. Daher ist es erforderlich, eine ärztliche Verschreibung bzw. staatliche Erlaubnis zum Erwerb entsprechender Arznei-/Betäubungsmittel zu ermöglichen.

Verständlicherweise wünschen sich Menschen auf ihrem letzten Weg Beistand und Unterstützung. Hilfe zur Selbsttötung sollte nicht mit Repression, dem Strafrecht oder Sühne geahndet werden, sondern bedarf tiefen Respekts gegenüber der Entscheidung der suizidwilligen Person und der Bereitschaft der hilfeleistenden Person.

Gleichwohl ist niemand zur Suizidhilfe verpflichtet. Die freie Entscheidung eines Arztes oder einer Ärztin, aus Gewissensgründen keine Suizidhilfe zu leisten, ist zu respektieren. Gleicher Respekt ist aber auch den Ärzten und Ärztinnen entgegenzubringen, die bereit sind, Menschen in ihrem selbstbestimmten Sterbewunsch zu begleiten.

Mit dem Gesetzentwurf verfolgt der Gesetzgeber seinen verfassungsrechtlichen Schutzauftrag. Durch die gesetzliche Ausgestaltung eines Verfahrens wird suizidwilligen Personen ein individuell als würdevoll erachtetes Sterben ermöglicht und abgesichert, dass der Sterbewunsch autonom gefasst wird und keine Drucksituation von außen besteht. Die Freiverantwortlichkeit muss im Mittelpunkt stehen. Die Menschen, die sterben möchten und oft allein mit ihren Gedanken und Überlegungen sind, müssen die Möglichkeit erhalten, über ihre Gedanken und Gefühle zu sprechen und Auswege zu diskutieren. Sie müssen echte Unterstützung auf ihrem selbstbestimmten Weg erfahren.

Mit dem Gesetzentwurf soll deshalb eine organisierte Beratungsinfrastruktur aufgebaut werden. Dadurch sollen Betroffene die Möglichkeit erhalten, sich qualifiziert mit dem Thema Suizid und Suizidhilfe auseinanderzusetzen, ohne in eine bestimmte Entscheidungsrichtung gedrängt zu werden. Weiterhin kann so sichergestellt werden, dass die Betroffenen über umfassende Informationen verfügen, die sie als mündige Individuen befähigen, eine informierte Entscheidung zu treffen. Ebenso soll der Gesetzentwurf sicherstellen, dass Menschen, die sterben möchten, nicht auf ein Angebot rein auf Gewinnstreben orientierter Einrichtungen angewiesen sind. Die Bereitstellung geregelter Verfahren, um einen frei gefassten Suizidwunsch verwirklichen zu können, kann schließlich auch ein Beitrag zum Lebensschutz sein, weil diese Verfahren zu einer Reduzierung von Suiziden führen können. Dies gilt schon deshalb, weil sich die Betroffenen in solchen Verfahren mit anderen über ihren Suizidwunsch auseinandersetzen müssen.

Das Recht auf selbstbestimmtes Sterben wurzelt in der Würde des Menschen. Es gilt daher, die nötigen gesetzlichen Regeln für ein würdevolles, selbstbestimmtes Sterben zu schaffen und suizidwillige Personen dabei gleichermaßen zu schützen. Suizidhilfe braucht Menschlichkeit und einen sicheren rechtlichen Rahmen.

Wesentlicher Inhalt des Entwurfs

Der vorliegende Gesetzentwurf beantwortet die Frage, unter welchen Voraussetzungen Hilfe zur Selbsttötung in Anspruch genommen beziehungsweise von Dritten geleistet werden darf. Er regelt den sicheren Zugang zu Arzneimitteln und bzw. oder Betäubungsmitteln (also auch eine Kombination) zum Zweck der Selbsttötung, um eine unwürdige, unzumutbare und nicht von einem freien Willen getragene Umsetzung des Sterbewunsches zu verhindern sowie eine autonome und vollinformierte Entscheidungsfindung suizidwilliger Personen sicherzustellen.

Der Entwurf adressiert Menschen, die sterben möchten und Menschen, die – unabhängig von der Berufszugehörigkeit – zur Hilfeleistung bereit sind, gleichermaßen.

Im Mittelpunkt stehen dabei die Autonomie und die Freiverantwortlichkeit der sterbewilligen Person und der Respekt vor jedweder autonomen Entscheidung, ohne dass externe Moralvorstellungen angelegt werden dürfen. Es ist daher verfassungsrechtlich unerheblich, ob eine suizidwillige Person ihr Leben aufgrund von

Krankheit oder anderen Motiven beenden möchte, solange eine autonome und freiverantwortliche Entscheidung erfolgt.

Der Gesetzentwurf soll solche selbstbestimmten Entscheidungen gewährleisten, aber auch sichern. Deshalb sieht er auch einen Schutz vor übereilten und nicht autonom gebildeten Suizidentscheidungen vor und orientiert sich dabei an den vom Bundesverfassungsgericht aufgestellten Kriterien (BVerfGE 153, 182 ff. – Rn. 240 - 244). Zudem soll er auch die Entscheidung, ob jemand Hilfe zur Selbsttötung leisten will oder nicht, schützen und Menschen, die Hilfe anbieten, absichern.

Die vorgesehene Beratung soll die suizidwillige Person bei ihrer mündigen informierten Meinungsbildung unterstützen. Dabei muss die Beratung ergebnisoffen sein und vom Grundwert jedes Menschenlebens ausgehen. Sie zeigt auch mögliche Alternativen zur Selbsttötung auf, ohne die suizidwillige Person zu bevormunden. Ein wichtiges Schutzelement ist die Trennung von Beratung und anschließender Suizidhilfe. Die Beratungsstellen müssen zudem staatlich anerkannt werden. Die vorgesehenen Wartefristen sollen vor übereilten Entscheidungen schützen und gleichzeitig einen zeitlichen Zusammenhang von Beratung und Selbsttötung sicherstellen.

Der Gesetzentwurf eröffnet die Möglichkeit der ärztlichen Verschreibung von Medikamenten zur Selbsttötung sowie der Erteilung einer verschreibungsähnlichen Erlaubnis zum Erwerb eines Arznei-/ Betäubungsmittels zum Zweck der Selbsttötung, sofern eine Verschreibung durch einen Arzt oder eine Ärztin nicht in zumutbarer Weise zu erlangen ist und die Voraussetzungen für eine solche Verschreibung aber vorliegen. Diese Dualität berücksichtigt zum einen die besondere fachliche Expertise von Ärzten und Ärztinnen sowie das zu den Patienten und Patientinnen bestehende Vertrauensverhältnis, lässt aber auch Menschen, die keinen Zugang zu einer ärztlichen Verschreibung erhalten, nicht allein.

Mit der Härtefallregelung wird berücksichtigt, dass suizidwillige Personen sich in besonders schwerwiegenden und leidträchtigen Situationen befinden können, die eine Verkürzung des Verfahrens notwendig machen. Die Möglichkeit zu einer solchen Differenzierung wurde seitens des Bundesverfassungsgerichts explizit aufgezeigt (a.a.O – Rn. 340).

Alternativen

Da der Kern der Problematik nicht im Strafrecht liegt, kommt eine Regelung im Strafgesetzbuch nicht in Betracht und dabei insbesondere nicht die Wiedereinführung des vom Bundesverfassungsgericht für nichtig erklärten § 217 StGB. Es wäre für alle Beteiligten unzumutbar, wenn zulässige Hilfe beim Suizid generell im Strafrecht verortet würde. Zudem ist das Strafgesetzbuch nicht der sachgerechte Regelungsort für die zum Schutz selbstbestimmten Sterbens und zur Regelung der Suizidhilfe nötigen Bestimmungen. Im Übrigen werden die zum Schutz des Lebens und der körperlichen Unversehrtheit vorhandenen Straftatbestände im StGB sowie im Betäubungsmittelgesetz (BtMG) dem strafrechtlichen Schutzaspekt bereits vollumfänglich gerecht.

Eine Regelung im ärztlichen Berufsrecht scheidet schon deshalb aus, weil angesichts der Länderzuständigkeit die erforderliche Rechtseinheit im Bund nicht gewährleistet werden könnte.

Gesetzgebungskompetenz

Die Gesetzgebungskompetenz des Bundes ergibt sich aus Artikel 74 Absatz 1 Nummern 1 und 19 des Grundgesetzes.

Vereinbarkeit mit dem Recht der Europäischen Union und völkerrechtlichen Verträgen

Es sind keine Kollisionen mit dem Recht der Europäischen Union oder völkerrechtlichen Verträgen ersichtlich.

Gesetzesfolgen

Rechts- und Verwaltungsvereinfachung

Keine.

Nachhaltigkeitsaspekte

Keine.

Haushaltsausgaben ohne Erfüllungsaufwand

Durch den Gesetzentwurf entstehen Kosten bei Bund und Ländern, die nicht näher abgeschätzt werden können.

Erfüllungsaufwand

Erfüllungsaufwand für Bürgerinnen und Bürger

Das Gesetz hat keine Auswirkungen auf den Erfüllungsaufwand für Bürgerinnen und Bürger.

Erfüllungsaufwand für die Wirtschaft

Das Gesetz hat keine Auswirkungen auf den Erfüllungsaufwand für die Wirtschaft.

Erfüllungsaufwand für die Verwaltung

Das Gesetz hat keine Auswirkungen auf den Erfüllungsaufwand für die Verwaltung.

Weitere Kosten

Für die Sicherstellung des ausreichenden Angebots an Beratungsstellen nach Artikel 1 § 5 Suizidhilfegesetz und hierbei insbesondere für angemessene Förderung der Personal- und Sachkosten gemäß § 5 Absatz 3 entstehen Kosten, die zum jetzigen Zeitpunkt jedoch nicht beziffert werden können. Die Länder haben im Rahmen ihrer Vollzugsverpflichtung die sachgerechte Erledigung des sich aus dem Gesetz ergebenden Aufgabenbestandes sicherzustellen.

Weitere Gesetzesfolgen

Keine.

Befristung; Evaluierung

Eine Befristung des Gesetzes ist nicht vorgesehen. Hingegen soll eine regelmäßige Evaluierung vorgenommen werden.

Besonderer Teil

Zu Artikel 1 (Gesetz zum Schutz des Rechts auf selbstbestimmtes Sterben und zur Regelung der Suizidhilfe)

Zu § 1 (Recht auf Hilfe zur Selbsttötung)

Das Recht auf selbstbestimmtes Sterben ist ein Grundrecht und steht unbestritten jedem zu (BVerfG, a.a.O. - Leitsatz 1a). Es stellt sicher, dass der Einzelne über sich entsprechend dem eigenen Selbstbild autonom bestimmen und damit seine Persönlichkeit wahren kann (a.a.O. - Rn. 209). Das Bundesverfassungsgericht hat in der vorgenannten Entscheidung eindeutig klargestellt, dass dieses Recht auch die Inanspruchnahme von Hilfeleistung von Dritten inkludiert, sofern diese angeboten wird (a.a.O. – Leitsatz 1c). Die aus einer ernsthaften Auseinandersetzung mit dem eigenen Tod erwachsende und autonom getroffene Bestimmung des Ob und Wie des eigenen Sterbens weist einen engen Bezug zum Kern der Persönlichkeit und deren Würde auf (a.a.O. – Rn. 128). Daher muss die Entscheidung, das eigene Leben mit Hilfe anderer zu beenden, zwingend auf einem autonom gebildeten, freien Willen beruhen, da anderenfalls ein Schutz ebendieser Autonomie verfehlt würde.

Zur Inanspruchnahme von zulässiger Hilfe zur Selbsttötung in Wohn- und Betreuungseinrichtungen bzw. Einrichtungen des Gesundheits- und Sozialwesens siehe unten zu § 2.

Zu Absatz 1

Absatz 1 soll deutlich machen, dass das grundgesetzlich verankerte Recht auf selbstbestimmtes Sterben der Ausgangspunkt für den Gesetzentwurf ist und den autonom gebildeten, freien Willen in den Mittelpunkt stellen. Menschen, die sterben möchten, sollen Sicherheit bekommen, dass sie mit ihrer Entscheidung nicht alleingelassen werden, sondern angebotene Hilfe in Anspruch nehmen dürfen. Gleichzeitig soll die Betonung der Eigenhändigkeit klarstellen, dass die tatsächliche Umsetzung des Sterbewunsches allein in den Händen der suizidwilligen Person liegt. Nur sie selbst und nicht Andere dürfen das Mittel zur Selbsttötung in ihren Körper einbringen. Die Grenze zur strafbaren Tötung auf Verlangen (§ 216 StGB) bleibt bestehen. Dabei

schließt die „Eigenhändigkeit“ die Nutzung von unterstützenden Mechanismen, etwa im Falle einer Querschnittslähmung, nicht aus.

Zu Absatz 2

Absatz 2 konkretisiert die Intention des Gesetzes. Es soll einerseits sichergestellt werden, dass suizidwillige Personen sich nicht unwürdiger und quälender Mittel bedienen müssen, um ihr Ziel zu erreichen. Deshalb wird die Möglichkeit eines sicheren Zugangs zu Arznei-/Betäubungsmitteln eröffnet, so dass sichergestellt ist, dass suizidwillige Personen sich bei der Verwirklichung ihres Todeswunsches nicht in Situationen begeben müssen, die einem würdevollen Sterben entgegenstehen.

Auf der anderen Seite soll durch das Gesetz aber auch der staatliche Schutzauftrag erfüllt und ein umfassendes Schutzkonzept für Betroffene implementiert werden. Hierzu gehört die Schaffung eines Rahmens für eine autonome und vollinformierte Entscheidungsfindung.

Zudem soll auch einer möglichen Stigmatisierung entgegengewirkt werden. Der Gesetzgeber achtet das Recht auf selbstbestimmtes Sterben und gestaltet es so aus, dass Sterbewillige rechtssicher und selbstbestimmt ihr Leben beenden können.

Zu § 2 (Recht zur Hilfeleistung)

§ 2 beinhaltet Regelungen zur Hilfe durch Dritte zur Selbsttötung, also jene Personen, die bereit sind, suizidwilligen Personen auf deren Wunsch hin zu helfen und ihnen beizustehen.

Die Regelungen gelten auch in Einrichtungen bzw. bei juristischen Personen, weil diese die Grundrechtsverwirklichung von in Suizidfragen Rat- und Hilfesuchenden und Suizidwilligen nicht grundsätzlich ausschließen dürfen. Ein Schutz der Einrichtungsnutzer und -nutzerinnen vor ungewollter Konfrontation mit Suizidhilfe(angeboten) kann auf andere geeignete Weise gewährleistet werden. Eventuelle Konflikte zwischen den Grundrechten der Einrichtungsnutzer und -nutzerinnen und der Einrichtungsbetreiber müssen in möglichst schonenden Ausgleich gebracht werden.

Zu Absatz 1

Absatz 1 formuliert den Umkehrschluss zu § 1 Absatz 1 und bildet den Leitsatz 1c des Bundesverfassungsgerichturteils gesetzlich ab (a.a.O. – Leitsatz 1c). Jeder darf Hilfe zur Selbsttötung leisten, sofern die suizidwillige Person ihr Leben selbst freiverantwortlich und eigenhändig beenden möchte und Hilfe wünscht.

Zugleich soll Absatz 1 den Menschen, die bereit sind, suizidwilligen Personen zu helfen und ihnen beizustehen, Rechtssicherheit bieten und hervorheben, dass sie nichts Verbotenes tun. Das soll explizit – klarstellend mit Blick auf die Entscheidung des Großen Strafsenats des Bundesgerichtshofs vom 10.4.1954 (BGH, Beschluss vom 10. März 1954 – GSSt 4/53 - BGHSt 6, 147,153) zur Rettungspflicht nach Eintritt der Bewusstlosigkeit der sich suizidierenden Person – auch nach der Hilfeleistung bis zum Zeitpunkt des Todeseintritts fortgelten. Denn niemand soll aus Sorge vor Strafverfolgung den Raum verlassen müssen, statt einem Menschen am Lebensende beizustehen.

Absatz 1 verdeutlicht somit, dass die suizidwillige Person grundsätzlich Hilfe zum Vollzug der von ihr beherrschten finalen Handlung annehmen kann, solange sie selbst das Geschehen aktiv in den Händen hält. So darf sie etwa das Angebot des Arztes oder der Ärztin nutzen, den Sterbewunsch in den bereitgestellten Räumen und in der Anwesenheit der hilfeleistenden Person zu verwirklichen. Das gilt grundsätzlich für entsprechende Angebote anderer Dritter gleichermaßen. „Hilfe zur Selbsttötung leisten“ im Sinne dieses Gesetzes ist also nicht als aktive Handlung zu verstehen, die final den Todeseintritt herbeiführt. Wie bereits in § 1 Absatz 1 klargestellt, obliegt die Selbsttötung allein der suizidwilligen Person. Die Strafbarkeit für eine Tötung auf Verlangen (§ 216 StGB) bleibt unberührt. „Hilfe leisten“ ist damit als Unterstützung der Person zu verstehen, die selbstbestimmt entscheidet, ihr eigenes Leben zu beenden, etwa durch vorherige (psychologische) Begleitung, physische Anwesenheit während der Suizidierung, Verschreibung des Mittels oder anderer organisatorischer Maßnahmen.

Zu Absatz 2

Spiegelbildlich zu Absatz 1 gilt, dass niemand zu einer Hilfeleistung zur Selbsttötung verpflichtet werden darf. Das grundgesetzlich verankerte Recht auf selbstbestimmtes Sterben sichert zwar jedem zu, angebotene Hilfe in Anspruch zu nehmen, es bestimmt aber nicht, dass auch jeder zur Hilfe bereit sein muss. Denn nicht

nur die Entscheidung, selbstbestimmt zu sterben, sondern ebenso die bewusste Entscheidung einer Person, an einer Hilfeleistung nicht teilhaben zu wollen, verlangt Respekt und Akzeptanz.

Absatz 2 gilt auch für alle nach diesem Gesetz zulässigen Handlungen, die Suizidwilligen Zugang zu Arzneimitteln zum Zweck der Selbsttötung ermöglichen.

Zu Absatz 3

§ 2 Absatz 3 Satz 1 soll die vorangehenden Absätze insoweit konkretisieren, als dass auch niemand aufgrund seiner Berufszugehörigkeit zur Suizidhilfe verpflichtet werden oder sie ihm aufgrund dieser verboten werden darf. Die Norm zielt insbesondere auf Ärzte und Ärztinnen ab, die aufgrund ihres Berufes vermehrt mit Sterbewünschen konfrontiert werden und auch durch diesen Gesetzentwurf mit einer Sonderstellung in der Suizidhilfe betraut werden.

Der Deutsche Ärztetag hat nach dem Urteil des Bundesverfassungsgerichts das berufsrechtliche Verbot der Suizidassistenz aus der Musterberufsordnung für Ärzte und Ärztinnen gestrichen (§ 16 Satz 3 der (Muster-)Berufsordnung wurde aufgehoben). Diese Streichung ist inzwischen auch von allen Landesärztekammern, die das Verbot der Suizidassistenz aus der rechtlich unverbindlichen Musterberufsordnung in ihre Berufsordnung übernommen hatten, umgesetzt worden. Ärzten und Ärztinnen ist es damit im gesamten Bundesgebiet berufsrechtlich nicht mehr untersagt, Hilfe zu einem freiverantwortlichen Suizid zu leisten. Mit einer gesetzlichen Klarstellung wird verdeutlicht, dass der Bundesgesetzgeber die Suizidhilfe ausdrücklich als legal anerkennt und Ärzte und Ärztinnen in ihrer freien Entscheidung absichert.

Die Norm stellt klar, dass es ebenso zu respektieren ist, wenn Ärzte und Ärztinnen bereit sind, Menschen in ihrem selbstbestimmten Sterbewunsch zu begleiten, wie wenn sie aus Gewissensgründen keine Suizidhilfe leisten möchten.

Satz 2 normiert ein Diskriminierungsverbot aller Beschäftigten, um deren Entscheidungsfreiheit für oder gegen eine Hilfeleistung zu Selbsttötung abzusichern. Der hier verwendete Beschäftigtenbegriff orientiert sich an dem des § 6 Allgemeines Gleichbehandlungsgesetz (AGG). Damit fallen etwa auch Belegärztinnen und -ärzte in den Anwendungsbereich. Auch in der Vergangenheit liegende Hilfeleistungen dürfen bei zukünftigen Bewerbungen nicht zu Benachteiligungen führen.

Zu § 3 (Autonom gebildeter, freier Wille)

Grundlage einer selbstbestimmten Entscheidung ist der autonom gebildete, freie Wille eines jeden Einzelnen. Suizidhilfe nach diesem Gesetz kann deshalb nur dann in Anspruch genommen werden, wenn die Entscheidung auch auf einem solchen Willen beruht. Das Bundesverfassungsgericht stellt eindeutig fest, dass der Staat dafür Sorge zu tragen habe, dass der Entschluss, begleiteten Suizid zu begehen, auf einem freien Willen beruht (BVerfGE 153, 182 ff. – Rn. 232). § 3 definiert die Voraussetzungen für die Annahme eines autonom gebildeten, freien Willens und orientiert sich im Wesentlichen an den vom Bundesverfassungsgericht skizzierten Anforderungen. Auch der deutsche Ethikrat betont, dass aufgrund der Irreversibilität einer Suizidentscheidung an deren Freiverantwortlichkeit besonders hohe Anforderungen zu stellen seien. Gleichzeitig merkt er an, dass die Anforderungen nicht das Recht auf ein selbstbestimmtes Sterben entleeren dürften, indem sie der betroffenen Person faktisch durch zu hohe Anforderungen den Verfügungsspielraum über das höchstpersönliche Gut ihres Lebens nehmen (Deutscher Ethikrat, Stellungnahme: Suizid – Verantwortung, Prävention und Freiverantwortlichkeit, 22.09.2022, S. 15).

Betreuerinnen und Betreuer können für die von ihnen betreute Person keine für die Verschreibung eines Arzneimittels/Betäubungsmittels zum Zwecke der Selbsttötung konstitutiven Erklärungen abgeben. Ist für den Aufgabenkreis der Gesundheitssorge Betreuung mit Einwilligungsvorbehalt angeordnet, ist in der Regel davon auszugehen, dass die Voraussetzungen des § 3 nicht vorliegen. Zur Diskussion vgl. z.B. Dodegge, Suizid und rechtliche Betreuung, FamRZ 2021, 5 ff, Henking, Ist Suizidhilfe bald auch eine Aufgabe des Betreuers? – Der Betreuer zwischen Suizidverhinderung und Suizidunterstützung, in BtPrax 2021, 92 ff.

Zu Nummer 1

Eine freie Suizidentscheidung setzt nach dem Urteil des Bundesverfassungsgerichts die Möglichkeit voraus, seinen Willen frei und unbeeinflusst von einer akuten psychischen Störung bilden und auch entsprechend handeln zu können (a.a.O. – Rn. 241). Unter Bezugnahme auf die ständige Rechtsprechung des Bundesgerichtshofs zur Einwilligung in medizinische Behandlung ist von der Einwilligungsfähigkeit, der Fähigkeit zur freien Willensbildung nur auszugehen, wenn die betreffende Person in der Lage ist, Art, Bedeutung und Tragweite (Risiken) der Entscheidung zu erfassen (BGH, Urteil. vom. 16.11.1971 – VI ZR

76/70 - NJW 1972, 335). Diese ist mithin nicht an das Vorliegen formaler Geschäftsfähigkeit (vgl. § 104 Nr. 2 BGB) gebunden.

Ob Erkrankungen, etwa psychische Erkrankungen oder eine Demenzerkrankung, der Fähigkeit zur freien Willensbildung entgegenstehen, ist im konkreten Einzelfall zu beurteilen. Das Merkmal der akuten psychischen Störung soll aber verdeutlichen, dass nicht pauschal jede psychische Erkrankung einer autonomen Willensbildung entgegensteht. Viele Menschen mit einer psychischen Erkrankung zeigen kein suizidales Verhalten und nicht alle Menschen, die sich ihr Leben nehmen, haben eine psychische Erkrankung (Brieger P, Menzel S, Hamann J., Wird die Rolle von psychischen Erkrankungen beim Suizid überbewertet? Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz, 2022 Jan;65(1):25-29). Auch der Deutsche Ethikrat betont, dass eine psychische Erkrankung als solche die Freiverantwortlichkeit der Suizidentscheidung nicht ausschließe. Ohne eine einzelfallbezogene substantiierte Begründung für eine normativ erhebliche Beeinträchtigung der Selbstbestimmungsfähigkeit sei auch die Entscheidungsfreiheit psychisch erkrankter Personen zu respektieren (Deutscher Ethikrat, Stellungnahme: Suizid – Verantwortung, Prävention und Freiverantwortlichkeit, 22.09.2022, S. 18).

Zu Nummer 2

Weitere Voraussetzung für die Bildung eines autonomen und freien Willens ist, dass die suizidwillige Person alle entscheidungserheblichen Tatsachen kennt und das Für und Wider des Suizidentschlusses ordnungsgemäß gegeneinander abwägen kann. Die freie Willensbildung setzt insbesondere voraus, dass der Entscheidungsträger bzw. die Entscheidungsträgerin Handlungsalternativen zum Suizid erkennt, ihre jeweiligen Folgen bewertet und die Entscheidung in Kenntnis aller erheblichen Umstände und Optionen trifft. Insoweit sollen dieselben Grundsätze wie bei einer Einwilligung in eine Heilbehandlung (vgl. § 630d und § 630e BGB) gelten (a.a.O. – Rn. 242).

Welche Informationen im Einzelnen vorausgesetzt werden, um die suizidwillige Person zur Entscheidung auf Basis einer autonomen und freien Willensbildung zu befähigen, hängt jeweils von den Umständen des Einzelfalls ab. Medizinisch wird im Falle einer Erkrankung regelmäßig das Wissen über etwaige alternative Behandlungsmöglichkeiten und Optionen der Palliativmedizin erforderlich sein. In jedem Fall muss die suizidwillige Person über Kenntnisse zu dem Vorgang der Selbsttötung und damit einhergehende Risiken verfügen. Die erforderlichen Informationen über Handlungsalternativen können auch Angaben über soziale Hilfs- und Unterstützungsangebote betreffen.

Zu Nummer 3

Ein autonom gebildeter und freier Wille kann nur gebildet werden, wenn weder unzulässige Einflussnahme noch Druck von Seiten Dritter ausgeübt wird (vgl. a.a.O.- Rn. 243). Die Entscheidung über den eigenen Tod ist die persönlichste Entscheidung einer Person. Der Suizidwillige darf sich zwar mit Dritten beraten und ihre Meinung ebenso wie die Folgen seiner Entscheidung auf Dritte berücksichtigen, jede Fremdbestimmung in der Entscheidung selbst gilt es aber zu verhindern. Denn obwohl intensive und möglichweise auch kontroverse Gespräche über die eigenen suizidalen Wünsche oft erst eine eigenständige und damit freiverantwortliche Entscheidung ermöglichen, kann der Einfluss anderer Personen die freiverantwortliche Entscheidung der suizidwilligen Person auch gefährden oder gar ausschließen (Deutscher Ethikrat, Stellungnahme: Suizid – Verantwortung, Prävention und Freiverantwortlichkeit, 22.09.2022, S. 19).

Von einer unzulässigen Ausübung von Druck ist jedenfalls dann auszugehen, wenn dem Betroffenen in Aussicht gestellt wird, dass er für den Fall des Weiterlebens Nachteile erleiden würde (vgl. BVerfG, Beschluss vom 23.3.2011 – 2 BvR 882/09 - Rn. 41- BVerfGE 128, 282 - 322).

Zu Nummer 4

Zudem soll von einem autonomen und freien Willen nur dann ausgegangen werden, wenn der Suizidentschluss auf einer gewissen Dauerhaftigkeit und inneren Festigkeit beruht. Diese Voraussetzung ergibt sich ebenfalls aus dem Urteil des Bundesverfassungsgerichts, folgt aber auch aus früheren Erwägungen des Bundesgerichtshofes (BVerfGE 153, 182 ff. - Rn. 244; BGH, Urteil vom 03.7.2019 – 5 StR 132/18 - Rn. 21 –BGHSt 64, 121-135). Der Suizidentschluss muss sich mithin in einem gewissen Maße im Willen manifestiert haben, sodass ein beiläufiger und gegebenenfalls nur einmalig geäußerter Suizidwille nicht ausreichen darf. Das Kriterium der Dauerhaftigkeit ist dabei geeignet, die Ernsthaftigkeit eines Suizidwunsches nachvollziehbar zu machen und sicherzustellen, dass er nicht etwa auf einer vorübergehenden Lebenskrise beruht (a.a.O. - Rn. 244). Empirische Daten belegen, dass ein kurzfristig gefasster Suizidentschluss, wenn die Selbsttötung misslingt, im Nachhinein von den Betroffenen selbst in etwa 80 bis

90 % der Fälle als Fehlentscheidung gewertet und revidiert wird. Durch Nummer 4 werden Entscheidungen, die als Kurzschlusshandlungen auf rein affektiven Impulsen beruhen oder im Rahmen vorübergehender Lebenskrisen gebildet werden, adressiert und ausgeschlossen. Die Dauerhaftigkeit und innere Festigkeit sind somit Ausprägung des dem Entwurf zugrundeliegenden Schutzgedankens. Insofern steht § 3 Nummer 4 auch in Verbindung zu § 6 Absatz 4, der eine Mindestdauer von drei Wochen zwischen Beratung und Verschreibung vorgibt. Dieser Übereilungsschutz bringt den staatlichen Schutzauftrag für das Schutzgut Leben mit dem Selbstbestimmungsrecht der suizidwilligen Personen in Einklang. Ob eine gewisse Dauerhaftigkeit und innere Festigkeit vorliegen, bleibt dabei Gegenstand einer einzelfallabhängigen Bewertung und kann nicht pauschal festgestellt werden.

Zu § 4 (Beratung)

Aus Artikel 1 Absatz 1 Satz 2 GG in Verbindung mit Artikel 2 Absatz 2 Satz 1 GG folgt die Pflicht des Staates, die individuelle Selbstbestimmung bei einer Suizidentscheidung und hierdurch mittelbar das Leben selbst zu schützen (a.a.O. - Rn. 232). Es gilt daher, Sterbewillige nicht allein zu lassen, sondern ihnen ein konkretes Informations-, und auch praktisches Unterstützungsangebot zu machen. Ein umfassendes Beratungsangebot ermöglicht es Personen, die es in Anspruch nehmen, Informationen zu erhalten, die sie zur autonomen Bildung eines freien Willens befähigen. Da die individuelle Autonomie als grundrechtlicher Mittler des Lebensschutzes fungiert, darf es folgerichtig einen gegen die Autonomie gerichteten Lebensschutz nicht geben. Dies muss die angebotene Beratung widerspiegeln.

Zu Absatz 1

Das Recht, sich zu Fragen der Suizidhilfe beraten zu lassen, soll jeder Person eröffnet werden, da eine ausreichende Information inhärenter Teil des Grundrechts auf selbstbestimmtes Sterben ist. Gleichwohl knüpft § 6 die Verschreibung des Arznei-/Betäubungsmittels einerseits als Ausdruck des staatlichen Schutzkonzepts an die Volljährigkeit, und andererseits an die deutsche Staatsangehörigkeit bzw. den Wohnsitz / gewöhnlichen Aufenthalt in Deutschland an, was einem Suizidtourismus entgegenwirken soll.

Diese Diskrepanz des persönlichen Anwendungsbereichs für die Beratung einerseits und die Verschreibung des Arznei-/Betäubungsmittels andererseits ist dabei bewusst gewählt. Jeder soll die Möglichkeit haben, sich bei Gedanken, die sich mit Suizid beschäftigen, an eine entsprechende Beratungsstelle wenden zu können. Minderjährige Personen können sodann an entsprechende Beratungsangebote / Präventionsangebote speziell für Minderjährige verwiesen werden.

Auch hinsichtlich Personen, die zwar keine Verschreibung erlangen können, da sie beispielsweise ihren Wohnsitz in einem anderen Land haben und keine deutsche Staatsangehörigkeit besitzen, ist es sinnvoll, das Beratungsangebot auch für sie zu eröffnen, da diesen Personen so ermöglicht wird, sich auszutauschen und auch hier ggf. eine Weitervermittlung an ein Suizidpräventionsangebot erfolgen kann.

Schließlich sollen sich auch Menschen, die selbst gar keine Suizidgedanken haben, beraten lassen können. Denkbar ist, dass Angehörige oder andere Betroffenen nahestehende Personen entsprechenden Beratungsbedarf haben oder auch, dass es Menschen gibt, die darüber nachdenken, ob sie einer suizidwilligen Person Hilfe leisten möchten und sich hierzu besprechen möchten.

Die vorgeschriebene Ergebnisoffenheit und das Verbot der Bevormundung sollen sicherstellen, dass die Beratung nicht darauf abzielt, die Person von ihrer Haltung abzubringen, da dies mit dem Selbstbestimmungsrecht nicht vereinbar wäre. Die Erwägungen der ratsuchenden Person für oder gegen eine Selbsttötung sind wertneutral hinzunehmen. Gleichzeitig umfasst eine ergebnisoffene Beratung selbstverständlich auch suizidpräventive Elemente. Die Formulierung betont zudem, dass der Gesetzgeber vom Grundwert jedes Menschenlebens ausgeht und intendiert eine klare Abgrenzung zu der insbesondere in der Zeit des Nationalsozialismus praktizierten Ideologie „lebensunwerter Leben“. Gleichzeitig soll die Formulierung der Gefahr entgegentreten, dass die Beratung eher zur Durchführung des assistierten Suizides ermutigt, als ergebnisoffen zu beraten.

Weiterhin wird sichergestellt, dass Personen mit akuten Suizidgedanken priorisiert beraten werden, denn sie brauchen akute Unterstützung. Empirisch belegt ist, dass die meisten suizidwilligen Personen (ca. 80 bis 90 %), deren Suizid misslingt, diesen im Nachhinein als Fehlentscheidung werten. Daraus lässt sich schließen, dass ein nicht unerheblicher Anteil der Suizidwünsche nicht anhaltend sind (a.a.O. -, Rn. 244). Suizidwillige Personen, die akute Suizidgedanken haben, brauchen deshalb umgehend Informationen zum Schutz ihrer Entscheidungsautonomie – auch und vor allem über Handlungsalternativen. Entsprechend kurzfristig muss eine Terminvergabe bzw. ad-hoc Beratung in den Beratungsstellen ermöglicht werden.

Zu Absatz 2

Absatz 2 folgt vor allem den Erwägungen des § 3 Nummer 2 und soll die suizidwillige Person befähigen, jede Tatsache und Information zu generieren, die für eine autonome Bildung eines freien Willens über eine Suizidentscheidung von Belang ist.

Absatz 2 normiert daher verschiedene Aspekte, über die die Beratung insbesondere aufklären soll. Der Katalog der Beratungsgegenstände ist nicht abschließend. Vielmehr bedarf es stets der individuellen Betrachtung und einer individuellen Anpassung der Beratungsinhalte an den konkreten Einzelfall. Die Vermittlung einiger wesentlicher Inhaltselemente der Beratung setzt voraus, dass sich die suizidwillige Person der sie beratenden Person öffnet, insbesondere ihre Beweggründe, derentwegen sie einen Suizid in Betracht zieht, mitteilt. Der Beratungscharakter schließt es aber selbstverständlich aus, eine solche Bereitschaft der suizidwilligen Person einzufordern oder gar zu erzwingen.

Zu Absatz 3

Niemand soll aus Sorge davor, durch Inanspruchnahme der Beratung unter eigenem Namen Nachteile zu erleiden, von der Nutzung des Beratungsangebotes abgehalten werden. Die beratene Person kann deshalb, wenn sie dies wünscht, gegenüber der beratenden Person anonym bleiben. Dies ist unproblematisch, weil eine Identifikation der Person zur Durchführung der Beratung nicht erforderlich ist. Gleichzeitig ist in diesem Kontext darauf hinzuweisen, dass eine für die Verschreibung des Medikaments notwendige Bescheinigung der Beratung (vgl. § 5 Absatz 4) auf den Namen der zu beratenden Person ausgestellt werden muss. Sofern sich die Person also entschließt, die Beratung anonym durchzuführen, steht dies der späteren Verschreibung des Mittels insoweit entgegen.

Zu Absatz 4

Die Überschneidung von einerseits neutraler Beratung und andererseits der Hilfeleistung wäre nicht im Sinne einer auf einem autonomen, freien Willen basierenden Suizidhilfe. Die Trennung von Beratung und Suizidhilfe unterstreicht zudem die präventive Komponente der Beratung. Die Hilfe zur Selbsttötung umfasst die Verschreibung des Arznei-/Betäubungsmittels nach § 6 und § 7, die Einschätzung, dass ein besonderer Härtefall im Sinne des § 7 vorliegt, die Erteilung einer Erlaubnis zum Erwerb eines Arznei-/Betäubungsmittels nach § 8 wie auch anderweitige Hilfeleistungen.

Zu Absatz 5

Diese Regelung trägt, wie auch Absatz 2, vor allem dem § 3 Nummer 2 Rechnung. Durch die Hinzuziehung ärztlicher, fachärztlicher, psychologischer, sozialpädagogischer, sozialarbeiterischer oder juristischer Expertise soll die Vielgestaltigkeit einer individuellen Beratungsnotwendigkeit berücksichtigt werden können. Die suizidwillige Person soll ein möglichst diverses Informationsportfolio erhalten, das sodann zu einer autonomen und freien Suizidentscheidung beiträgt. Weil die Entscheidung über Leben und Tod einerseits von der Einbettung in das soziale Umfeld abhängen kann und andererseits ein Suizid zumeist erhebliche Wirkung auf Angehörige und andere nahestehende Personen hat, sollen solche Personen, wenn beratene und beratende Person dies einvernehmlich für geboten halten, in die Beratung eingebunden werden können.

Zu § 5 (Beratungsstellen)

§ 5 widmet sich vor allem der Errichtung einer Infrastruktur, die für eine bundesweite Suizidberatung notwendig ist, und den notwendigen Anforderungen an diese Beratungsstellen. Während § 4 die Modalitäten der Beratung regelt, trifft § 5 Regelungen zu den die Beratung anbietenden Stellen.

Zu Absatz 1

Beim Ausbau der Infrastruktur ist darauf zu achten, dass Beratungsstellen barrierefrei erreichbar sind. Gerade für pflegebedürftige und schwerstkranke Betroffene ist ein aufsuchendes Angebot zwingend notwendig. Um eine ausreichende Qualität der Beratung sicherzustellen und auch der Gefahr einer bevormundenden Beratung vorzubeugen, ist eine staatliche Anerkennung erforderlich. Dies schließt keineswegs aus, dass Beratungsstellen in freier Trägerschaft unterhalten werden dürfen. Ein entsprechendes Angebot wäre vielmehr im Sinne eines flächendeckenden Beratungsangebots. Die Beratungsstellen dürfen keine mit dem

Beratungsauftrag konfligierenden Aufgaben wahrnehmen. Auch Ärztinnen und Ärzte können sich als Beratungsstelle anerkennen lassen. Wenn sie in dieser Funktion agieren, ist § 4 Absatz 4 zu beachten.

Zu Absatz 2

Die staatliche Anerkennung darf nur erteilt werden, wenn die Beratungsstelle ausreichend persönlich und fachlich qualifiziertes Personal vorweisen kann. Ebenso muss sie nachweisen, dass sie kurzfristig auf zusätzliche Expertise gemäß § 4 Absatz 5 zurückgreifen kann und mit weiteren Stellen, die anderweitige Hilfe für suizidwillige Personen anbieten, zusammenarbeitet. Zusätzlich ist eine staatliche Anerkennung dann ausgeschlossen, wenn eine Beratungsstelle mit einer Einrichtung derart zusammenarbeitet oder anderweitig organisatorisch – etwa personell – verbunden ist, dass ein materielles Interesse an der Durchführung von Suizidhilfe besteht.

Zu Absatz 3

Die Beratungsstellen haben einen Anspruch auf eine angemessene Förderung der Personal- und Sachkosten.

Zu Absatz 4

Die Beratung einer suizidwilligen Person muss auf ihren Wunsch hin von der Beratungsstelle bescheinigt werden. Dies ist nur möglich, wenn die Person nicht nach § 4 Absatz 3 anonym bleiben möchte. Die Bescheinigung ist notwendige Voraussetzung für die Verschreibung des Arznei-/Betäubungsmittels zum Zwecke der Selbsttötung (vgl. § 6 Absatz 3). Durch dieses Sicherungselement wird der staatlichen Schutzpflicht zur vollinformierten Entscheidungsfindung nachgekommen.

Wichtig ist zu betonen, dass es nicht Aufgabe der Beratungsstelle sein soll, Betroffene zu begutachten. Sie soll auf Augenhöhe beraten und unterstützen. Sofern sich für die beratende Person dennoch der Eindruck aufdrängt, dass der suizidwilligen Person die Fähigkeit zur freien Willensbildung fehlt oder aber unzulässig auf sie Einfluss genommen oder Druck ausgeübt wird, hat sie dies auf der Beratungsbescheinigung in geeigneter Form zu vermerken.

Möchte der Betroffene später die Verschreibung eines tödlichen Medikaments nach § 6 erwirken, ist es der Arzt oder die Ärztin, der beziehungsweise die sich von der Entscheidung zum Suizid aus autonom gebildetem, freiem Willen zu überzeugen hat. Es erschiene jedoch unbillig, wenn die Beratungsstelle offensichtliche Hinweise nicht weitergeben dürfte. Dem Arzt oder der Ärztin obliegt sodann eine besonders sorgfältige Prüfung.

Zu Absatz 5

Die Möglichkeit, eine Beratung in Anspruch zu nehmen, darf nicht von der finanziellen Situation der suizidwilligen Person abhängig sein. Die Beratung muss daher unentgeltlich erfolgen.

Zu Absatz 6

Die Beratungsstellen sind verpflichtet, einen jährlichen Bericht über ihre Beratungstätigkeit zu verfassen. Dieser Bericht ist auf Grundlage der Aufzeichnungen, die über die Beratungsgespräche anzufertigen sind, zu verfassen, wobei der Bericht keinerlei Rückschlüsse auf die Identität der beratenen Personen ermöglichen darf.

Zu Absatz 7

Die Anerkennung einer Beratungsstelle muss alle drei Jahre überprüft werden. Bewertungsmaßstab sind die Voraussetzungen des § 5 Absatz 2 sowie die Berichte des § 5 Absatz 6 dieses Gesetzes.

Zu Absatz 8

Die Länder haben im Rahmen ihrer Vollzugsverpflichtung die sachgerechte Erledigung des sich aus dem Gesetz ergebenden Aufgabenbestandes sicherzustellen.

Zu § 6 (Verschreibung eines Arznei- oder Betäubungsmittels zum Zwecke der Selbsttötung)

Mit dieser Norm wird die Möglichkeit, ein Arznei-/Betäubungsmittel zum Zwecke der Selbsttötung zu erhalten, eröffnet.

Ärzte und Ärztinnen verfügen über die notwendigen Kenntnisse und Kompetenzen zur Verschreibung von Arznei-/Betäubungsmitteln zur Selbsttötung. Das Gesetz sieht daher vor, dass eine Verschreibung primär über sie zu erfolgen hat. Zudem besteht zwischen Ärzten und Ärztinnen und ihren Patienten und Patientinnen häufig ein besonderes, teilweise lange gewachsenes Vertrauensverhältnis, das eine Einschätzung erleichtern kann.

Gleichzeitig ist zu betonen, dass die Entscheidung für ein selbstbestimmtes Sterben nach diesem Gesetz explizit keine Krankheit voraussetzt. Die Verschreibung durch einen Arzt oder eine Ärztin soll daher nicht implizieren, dass stets von einer vorangegangenen Krankheit ausgegangen wird. Vielmehr sind Ärzte und Ärztinnen grundlegend in der Lage zu erkennen, ob die suizidwillige Person aus autonom gebildetem, freiem Willen heraus handelt, und können eine der Tragweite der Entscheidung angemessene medizinische Aufklärung leisten.

Zu betonen ist in diesem Kontext, dass aus dem in § 2 Absatz 2 statuierten Grundsatz, dass niemand zur Suizidhilfe verpflichtet werden darf, unmittelbar folgt, dass kein Arzt und keine Ärztin zur Verschreibung eines tödlich wirkenden Arznei-/Betäubungsmittels verpflichtet werden darf. Sofern er oder sie aber zu einer solchen Verschreibung bereit ist, eröffnet § 6 diese Möglichkeit.

Zu Absatz 1

Absatz 1 bildet die Rechtsgrundlage für die ärztliche Verschreibung eines Arznei-/Betäubungsmittels zum Zwecke der Selbsttötung, sofern der Entschluss zur Selbsttötung auf einem autonom gebildeten, freien Willen fußt, wie er in § 3 dieses Gesetzes definiert ist. Die Beurteilung, ob ein solcher autonomer, freier Wille besteht, obliegt dem Arzt bzw. der Ärztin. Bei auch nur geringen Zweifeln an einer freien Willensbildung ist eine Verschreibung abzulehnen. Zweifel können gegebenenfalls durch ärztliche Konsile oder ggf. zusätzliche Gutachten ausgeräumt werden. Im Übrigen besteht nach bestehenden Normen im Strafgesetzbuch das Risiko strafrechtlicher Verurteilung für den Arzt oder die Ärztin.

Da - wie der Deutsche Ethikrat betont - die Freiverantwortlichkeit einer Entscheidung strikt von der moralischen Richtigkeit des Entscheidungsergebnisses zu unterscheiden ist, sind Wertvorstellungen der suizidwilligen Person bei der Beurteilung der Freiverantwortlichkeit uneingeschränkt zu respektieren – selbst dann, wenn sie von den allgemein in der Gesellschaft vorherrschenden Vorstellungen erheblich abweichen und die auf ihnen beruhende Suizidentscheidung deshalb im Ergebnis nicht nachvollziehbar erscheinen mag (Deutscher Ethikrat, Stellungnahme: Suizid – Verantwortung, Prävention und Freiverantwortlichkeit, 22.09.2022, S. 15-16).

Ferner wird die Verschreibungsmöglichkeit unter anderem an das Kriterium der Volljährigkeit geknüpft. Der Gesetzgeber kommt hier seinem besonderen Schutzauftrag gegenüber Minderjährigen nach und gestattet die Abgabe eines Mittels zur Selbsttötung nur an Personen, die das 18. Lebensjahr vollendet haben. Der Schutzaspekt ist bei Minderjährigen, bei denen die Gefahr besteht, dass sie aufgrund der noch nicht abgeschlossenen Persönlichkeitsentwicklung Bedeutung und Tragweite einer solch endgültigen Entscheidung noch nicht überblicken können, im Rahmen der Abwägung der Verfassungsgüter besonders – und stärker als bei Erwachsenen – zu gewichten, sodass von einer Möglichkeit zur Verschreibung von Arznei-/Betäubungsmitteln zur Selbsttötung vor Erreichen der Volljährigkeit abgesehen wurde.

Weitere Voraussetzung für die Verschreibung ist, dass die Person die deutsche Staatsangehörigkeit besitzt oder ihren Wohnsitz oder ihren gewöhnlichen Aufenthalt in Deutschland hat. Erklärtes Ziel des Gesetzentwurfs ist es, das Recht auf selbstbestimmtes Sterben zu gewährleisten und das individuelle Selbstbestimmungsrecht zu sichern. Es ist hingegen nicht gewollt, dass sich Deutschland zum Land des internationalen Sterbehilfetourismus entwickelt. Daher wird ein gewisser dauerhafterer Bezug zu Deutschland gefordert. Einen Wohnsitz hat jemand dort, wo er eine Wohnung unter Umständen innehat, die darauf schließen lassen, dass er die Wohnung beibehalten und benutzen wird. Den gewöhnlichen Aufenthalt hat jemand dort, wo er sich unter Umständen aufhält, die erkennen lassen, dass er an diesem Ort oder in diesem Gebiet nicht nur vorübergehend verweilt.

Die Anknüpfung an die Staatsangehörigkeit soll sicherstellen, dass auch Deutsche, die ihren Wohnsitz im Ausland haben, die Möglichkeit bekommen, zum Lebensende nach Deutschland zurückzukehren.

Zu Absatz 2

Dem verschreibenden Arzt oder der verschreibenden Ärztin obliegt es, die suizidwillige Person medizinisch aufzuklären (vgl. § 630e BGB). Die medizinische Aufklärung ergänzt die durch die Beratungsstellen vorgenommene umfassende Beratung nach § 4. Inhalt der ärztlichen Aufklärung müssen insbesondere die Darstellung der Reaktionen des Körpers, die auf die Einnahme des Arzneimittels zum Zwecke der Selbsttötung folgen, sowie der voraussichtliche Ablauf einer Selbsttötung durch Einnahme des Arznei-/Betäubungsmittels sein. Ferner ist auf mögliche Risiken, insbesondere das Fehlschlagen der Selbsttötung und das Eintreten schwerster physischer Schäden, einzugehen. Falls die suizidwillige Person unter einer Erkrankung leidet, hat der Arzt oder die Ärztin auf alternative Behandlungsmöglichkeiten und palliativmedizinische Behandlungsmöglichkeiten hinzuweisen und diese mit der suizidwilligen Person zu erörtern. Durch ärztliche Aufklärung soll sichergestellt werden, dass die suizidwillige Person sich der Bedeutung und Tragweite der Einnahme des Arznei-/Betäubungsmittels und der möglichen Risiken bewusst ist.

Zu Absatz 3

Die Verschreibung eines Arznei-/Betäubungsmittels zum Zwecke der Selbsttötung ist nur zulässig, sofern sich der verschreibende Arzt oder die Ärztin die Beratungsbescheinigung gemäß § 5 Absatz 4 hat vorlegen lassen. Dabei darf die Beratung nicht länger als zwölf Wochen zurückliegen. Diese Frist gewährleistet einerseits, dass die Beratungssituation regelmäßig noch mit der aktuellen Lebenssituation der sterbewilligen Person, vor deren Hintergrund beraten worden ist, übereinstimmt und andererseits, dass der beratenden Person die Erwägungen aus dem Gespräch noch präsent sind.

Ist keine Beratung erfolgt oder liegt diese schon länger als zwölf Wochen zurück, darf der Arzt oder die Ärztin das Arznei-/Betäubungsmittel zum Zwecke der Selbsttötung nicht verschreiben. Dabei ist keinesfalls ausgeschlossen, dass eine suizidwillige Person sich mehrfach über einen längeren Zeitraum beraten lässt. Nur die letzte bescheinigte Beratung darf nicht länger als zwölf Wochen zurückliegen. Mithin ist für die Verschreibung eines tödlichen Arznei-/Betäubungsmittels nach § 6 mindestens eine Beratung mit entsprechender Bescheinigung obligatorisch.

Wenngleich es aus Gründen der Humanität geboten ist, Menschen, die selbstbestimmt sterben möchten, Zugang zu möglichst schonenden und risikoarmen Arznei-/Betäubungsmitteln zu ermöglichen, hat der Staat dann, wenn er Suizidhilfe wie mit der Verschreibung eines Arznei-/Betäubungsmittels in ein staatlich geregeltes Verfahren einbettet, besondere prozedurale Sicherungsmaßnahmen zu treffen. Gleichzeitig ist es wichtig zu betonen, dass die Beratungsbescheinigung keine Erlaubnis zum Suizid darstellt, sondern allein bestätigt, dass eine Beratung stattgefunden hat, was die Ärztin oder der Arzt sodann anhand der Bescheinigung überprüft.

Zu Absatz 4

Die Verschreibung eines Arznei-/Betäubungsmittels zum Zwecke der Selbsttötung darf nur erfolgen, wenn seit der Beratung mindestens drei Wochen vergangen sind. Dieser Zeitraum gibt der suizidwilligen Person die Möglichkeit, die Informationen aus der Beratung abzuwägen und in den eigenen Entscheidungsprozess einfließen zu lassen. Die Regelung stellt einen weiteren Sicherungsmechanismus dar, um Menschen vor einer übereilten Suizidentscheidung zu schützen. Dieser Übereilungsschutz bringt den staatlichen Schutzauftrag für das Schutzgut Leben mit dem Selbstbestimmungsrecht der suizidwilligen Personen in Einklang. Denn eine Suizidentscheidung kann nur dann auf einem autonom gebildeten freien Willen beruhen, wenn sie auch von einer gewissen Dauerhaftigkeit und inneren Festigkeit getragen ist (vgl. § 3 Nr.4). Dabei stellt die Dreiwochenfrist die absolute Minimalfrist zwischen Beratung und (regulärer) Verschreibung dar. Gleichwohl ist davon auszugehen, dass der Zeitraum zwischen Beratung und Verschreibung in aller Regel länger als drei Wochen sein wird und eine suizidwillige Person sich über einen längeren Zeitraum mehrfach beraten lassen wird. Gerade bei einem seit langem bestehenden Arzt- beziehungsweise Ärztin-Patientenverhältnis kann der Arzt oder die Ärztin die Bewertung der Festigkeit des Sterbewunsches auch auf eine lange Behandlungshistorie stützen. Dem Arzt oder der Ärztin ist es unbenommen, mehr als drei Wochen verstreichen zu lassen, um seiner-/ ihrerseits wiederum mehrere Gespräche mit der suizidwilligen Person führen zu können.

Zu Absatz 5

Wie die Beratungsstelle muss auch der Arzt oder die Ärztin die Feststellungen über den autonomen und freien Willen sowie den Inhalt der vorgenommenen Aufklärung dokumentieren. Die Beratungsbescheinigung, die

die suizidwillige Person dem Arzt oder der Ärztin aushängen muss, ist im Original in die Dokumentation aufzunehmen.

Zu § 7 (Verschreibung eines Arznei- oder Betäubungsmittels zum Zwecke der Selbsttötung in Härtefällen)

Die Vorschrift schafft eine Ausnahmeregelung für Menschen, die aufgrund besonderer Umstände die vorgesehene externe Beratung durch eine Beratungsstelle und die zwischen der Beratung und der Verschreibung liegende Zeit von mindestens drei Wochen nicht einhalten können, da ihr Leidensdruck besonders hoch ist (Härtefallregelung). Aus diesem Grund werden die Anforderungen an die Verschreibung in diesen Fällen, insbesondere mit Blick auf den Übereilungsschutz, herabgesetzt. Die Regelung stellt eine besondere Ausnahme dar und soll der Befürchtung betroffener Personen, in einer solchen Situation keine geeigneten Mittel zu einer zumutbaren Selbsttötung erhalten zu können, Rechnung tragen. Das Bundesverfassungsgericht hat insoweit explizit klargestellt, dass es dem Gesetzgeber freisteht, je nach Lebenssituation unterschiedliche Anforderungen an den Nachweis der Ernsthaftigkeit und Dauerhaftigkeit eines Selbsttötungswillens zu stellen (a.a.O. Rn. 340).

Wer im Rahmen von § 7 die Verschreibung eines Arznei-/Betäubungsmittels zum Zwecke der Selbsttötung nicht erlangt, weil die speziellen Voraussetzungen nicht vorliegen, der Arzt oder die Ärztin dazu nicht bereit ist oder etwa die Zweitbegutachtung negativ ist, ist nicht gehindert, nach § 6 oder § 8 eine Verschreibung zu erlangen.

Zu Absatz 1

Jede Verschreibung nach § 7 setzt einen besonderen Härtefall voraus. Die Entscheidung, ob ein solcher Härtefall vorliegt, obliegt dem Arzt oder der Ärztin. Zwar verbietet es sich, die Zulässigkeit einer Hilfe zur Selbsttötung grundsätzlich vom Vorliegen einer unheilbaren oder tödlich verlaufenden Krankheit abhängig zu machen. Dies hat das Bundesverfassungsgericht explizit betont: „Aufgrund der verfassungsrechtlichen Anerkennung des Rechts auf Selbsttötung, welche die einem individuellen Suizidentschluss zugrundeliegenden Motive einschließt und diese damit einer Beurteilung nach Maßstäben objektiver Vernünftigkeit entzieht, verbietet es sich aber, die Zulässigkeit einer Hilfe zur Selbsttötung materiellen Kriterien zu unterwerfen.“ Je nach Lebenssituation können aber unterschiedliche Verfahrensanforderungen gestellt werden (vgl. a.a.O. - Rn. 340).

Der Arzt oder die Ärztin hat selbstverständlich auch im Rahmen des Verfahrens nach § 7 zu überprüfen, ob die Person aus autonom gebildetem, freiem Willen im Sinne des § 3 ihr Leben beenden möchte (vgl. § 6 Absatz 1) und muss sie entsprechend § 6 Absatz 2 aufklären. Eine gesonderte Beratung durch eine Beratungsstelle im Sinne des § 5 muss hingegen nicht erfolgen, weshalb auch die in § 6 Absatz 3 und 4 normierten Fristen von mindestens drei Wochen und maximal zwölf Wochen nach der Beratung keine Anwendung finden. Um ein zusätzliches Schutzelement - auch für Ärzte und Ärztinnen - einzubauen, ist vorgesehen, dass die Einschätzung, ob ein Härtefall vorliegt, zusätzlich durch einen zweiten unabhängigen Arzt oder eine zweite unabhängige Ärztin getroffen werden muss (Erfordernis des Vier-Augen-Prinzips).

Zu Absatz 2

In Absatz 2 wird definiert, wann ein besonderer Härtefall vorliegt. Dabei wird einerseits subjektiv an den Leidenszustand der suizidwilligen Person und die jeweiligen Auswirkungen auf ihr Leben angeknüpft. Erfasst werden soll auch ein Leidenszustand, bei dem Funktions-, Partizipations- und Kontrollverlust im Spätstadium einer Krankheit als unvereinbar mit den eigenen Vorstellungen vom Lebensende und ggf. als quälende Erfahrung oder Angst empfunden werden (etwa im Sinne von: keine Lebenszeit ohne Lebensqualität). Konkretisiert („insbesondere“) wird dies andererseits durch Bezugnahme auf die (objektive) Palliativdefinition (vgl. § 37b Absatz 1 Satz 1 SGB V). Damit sollen die Voraussetzungen, wann ein Arzt oder eine Ärztin nach § 7 verschreiben darf, möglichst konkret definiert und so eine rechtssichere Entscheidungsgrundlage geschaffen werden.

Zu § 8 (Erteilung einer Erlaubnis zum Erwerb eines Arznei- oder Betäubungsmittels zum Zweck der Selbsttötung)

§ 8 trägt dem Umstand Rechnung, dass es für einige suizidwillige Personen aus individuellen Gründen nicht möglich sein kann, eine Verschreibung durch einen Arzt oder eine Ärztin nach § 6 zu erlangen. Diese Gründe können vielfältiger Natur sein. Eine Verschreibung ist etwa dann nicht zumutbar zu erlangen, wenn die

Person bereits mehrere Ärzte oder Ärztinnen aufgesucht hat aber keiner bereit war, eine entsprechende Verschreibung vorzunehmen, obwohl die Voraussetzungen von § 6 vorliegen. Hierbei gibt es keine starre Grenze, wie viele Ärzte oder Ärztinnen eine Person bereits aufgesucht haben muss, bevor sie sich an die staatliche Stelle wenden kann. Vielmehr sind die individuellen Umstände zu berücksichtigen, die die suizidwillige Person der staatlichen Stelle gegenüber glaubhaft machen muss.

§ 8 soll damit sicherstellen, dass auch Betroffene, denen – aus welchen Gründen auch immer – keine Möglichkeit zu einer ärztlichen Verschreibung zuteilwird, nicht auf individuell als inhuman und nicht würdevoll empfundene Suizidmöglichkeiten verwiesen sind.

Es müssen sodann die gleichen Voraussetzungen wie bei einer Verschreibung nach § 6 eingehalten werden. Das heißt, dass sowohl der freie Wille im Sinne des § 3 festgestellt werden muss (vgl. § 6 Absatz 1) als auch eine medizinische Aufklärung entsprechend § 6 Absatz 2 zu erfolgen hat und die vorherige Beratung samt der vorgesehenen Fristen (vgl. § 6 Absatz 3 und 4) durchgeführt worden sein muss. Dass diese Voraussetzungen vorliegen, muss entweder ein Arzt oder eine Ärztin oder bei der zuständigen Stelle angestelltes gleichermaßen qualifiziertes Fachpersonal bestätigen. Entsprechend muss das Fachpersonal über hinreichende medizinische Kenntnisse verfügen.

§ 8 stellt insofern einen staatlichen Rückfallanker dar, der absichert, dass es suizidwilligen Personen möglich ist, auch dann auf würdevollem Weg ein Mittel zur Selbsttötung zu erlangen, wenn die Voraussetzungen des § 6 vorliegen, aber keinen aufgesuchten Arzt oder keine aufgesuchte Ärztin eine Verschreibung vornehmen möchte, wozu aufgrund § 2 Absatz 2 auch niemand verpflichtet ist.

Wichtig ist hervorzuheben, dass die Bediensteten der staatlichen Stelle keine Entscheidung darüber treffen, ob sie den entsprechenden Entschluss der Betroffenen ethisch billigen oder nicht. Der Staat erteilt keine Erlaubnis zur Selbsttötung – diese Position kann und darf er nicht einnehmen – sondern allein die Erlaubnis (vergleichbar mit einer ärztlichen Verschreibung), das Mittel zu erwerben, sofern die entsprechenden Voraussetzungen vorliegen.

Zu § 9 (Verordnungsermächtigung)

Durch § 9 wird die Bundesregierung (und qua Subdelegation das Bundesministerium für Gesundheit und das Bundesministerium der Justiz) ermächtigt, mit Zustimmung des Bundesrates per Rechtsverordnung nähere Regelungen zu einzelnen Unteraspekten des Verfahrens und der Zugänglichmachung des zur Selbsttötung einzunehmenden Mittels zu regeln.

Darunter fällt insbesondere auch die sichere Aufbewahrung der zu verschreibenden Arznei-/Betäubungsmittel. Im Bereich des Betäubungsmittelgesetzes ist in diesem Kontext z.B. auf bereits bestehende (und mit Sanktionen bewehrte) Regelungen hinzuweisen. Eine Höchstdauer der Aufbewahrung ergibt sich aus § 16 BtMG (Vernichtungspflicht, wenn nicht mehr verkehrsfähig bzw. Rückgabe auf Aufforderung des BfArM; Verstoß ist als Ordnungswidrigkeit gemäß § 32 Absatz 1 Nummer 10 BtMG sanktioniert). Die unerlaubte Weitergabe ist nach § 29 Absatz 1 Nummer 1 BtMG strafbar. Spezielle Verschreibungs- und Aufbewahrungsregelungen bei Menschen in Alten-/Pflegeheimen, Hospizen, ambulanter Palliativversorgung gibt es in § 5c BtMVV. Die Prüfung, ob und welche weiteren Regelungen zur sicheren Aufbewahrung von zum Suizidzweck verschriebenen Arznei-/Betäubungsmittel erforderlich sind, bleibt dem Verordnungsgeber überlassen.

Per Rechtsverordnung können ferner gewisse Meldepflichten errichtet werden, beispielsweise mit Blick auf vorgenommene Verschreibungen, ausgegebene Arznei-/Betäubungsmittel und in der Folge vorgenommene Selbsttötungen. Zudem können so durch Rechtsverordnung die notwendigen Regelungen für den Meldeschein im Todesfall an die neuen gesetzlichen Rahmenbedingungen durch den vorliegenden Entwurf angepasst werden. Die Verordnung soll zudem den Umfang der Dokumentationspflichten und der entsprechenden Weiterleitung für die in §10 vorgesehene Evaluation regeln. Dazu sollte u.a. die Festlegung gehören, anhand welcher Unterlagen evaluiert wird sowie welche Informationen in welcher Form Beratungsstellen und verschreibende Ärzte und Ärztinnen bzw. die nach Landesrecht zuständigen Stellen bereitzustellen haben (jeweils unter Wahrung des Datenschutzes/der Anonymität der Beratenen und Suizidenten).

Weiterhin müssen im Rahmen der Suizidhilfe tätige Ärzte und Ärztinnen, wenn sie suizidwillige Personen aufklären und Arznei-/Betäubungsmittel verschreiben, nicht unentgeltlich tätig sein. Es dürfte aber im Ergebnis sinnvoll sein, die mögliche Vergütung der Höhe nach zu begrenzen.

Gleichzeitig besteht auch die Möglichkeit, die Zulassung organisierter Angebote von Suizidhilfe von einer Zuverlässigkeitsprüfung abhängig zu machen und die Prävention gegen die Etablierung rein auf Gewinnstreben ausgerichteter Angebote per Rechtsverordnung zu regeln.

Zu § 10 (Berichtswesen; Evaluation)

Das Thema der Suizidhilfe ist von solch hoher Relevanz, dass es über den einmaligen Gesetzgebungsakt hinaus ständiger parlamentarischer Begleitung bedarf. Die Vorschrift sieht eine umfassende Evaluierung des Gesetzes vor. Der Zeitraum ist dabei so bemessen, dass einerseits hinreichend Zeit für die Ermittlung aller Fakten und die wissenschaftliche Auswertung bleibt und andererseits nicht zu viel Zeit vergeht, damit der Gesetzgeber die gewonnenen Erkenntnisse zum Anlass für eine eigene Bewertung nehmen kann.

Der Evaluationsbericht soll allgemeine Entwicklungen im Bereich der Suizidhilfe aufzeigen, damit der Gesetzgeber die kurzfristigen und langfristigen Auswirkungen der gesetzlichen Regelung fortlaufend betrachten kann. Die Bundesregierung hat erstmals spätestens nach drei Jahren nach Inkrafttreten dieses Gesetzes und sodann alle drei Jahre die Wirksamkeit dieses Gesetzes umfassend zu evaluieren. Sofern sich Entwicklungen hin zu der Entstehung und Etablierung rein auf Gewinnstreben orientierter Angebote erkennen lassen, ist staatlicherseits im Einklang mit den Erwägungen des Bundesverfassungsgerichts umgehend nachzusteuern.

Die entsprechende Erhebung der Daten, welche die Grundlage der Evaluation bilden, soll sich auf die Dokumentation und die Meldungen stützen. Der entsprechende Umfang und der zwingend erforderliche Datenschutz sind in der Rechtsverordnung zu regeln.

Zu § 11 Übergangsvorschrift

Der Aufbau und die Anerkennung der Beratungsstelleninfrastruktur wird eine gewisse Zeit in Anspruch nehmen. Um die Umsetzung von Sterbewünschen nach Inkrafttreten dieses Gesetzes gewährleisten zu können und Betroffene nicht allein zu lassen, treten Ärzte und Ärztinnen für einen Zeitraum von bis zu zwei Jahren an die Stelle der Beratungsstellen. Zwar ist ohnehin im Gesetz vorgesehen, dass sich Ärzte und Ärztinnen als Beratungsstellen anerkennen lassen können. Im Übergangszeitraum entfällt allerdings die Pflicht zur staatlichen Anerkennung. Gleichwohl sind die Ärzte und Ärztinnen an die Anforderungen, die § 4 für eine Beratung normiert, gebunden. Durch die Verweisung auf § 4 ist klar (siehe § 4 Absatz 4), dass Ärzte und Ärztinnen, die hier übergangsweise beraten, den jeweils Beratenen später kein Arznei-/Betäubungsmittel zum Suizidzweck verschreiben dürfen.

Zu Artikel 2 (Änderung des Betäubungsmittelgesetzes)

Aufgrund der Implementierung des prozeduralen Sicherungskonzeptes bedarf es einer konsistenten Anpassung des Betäubungsmittelrechts. Das Betäubungsmittelgesetz wird entsprechend den Erwägungen dieses Gesetzes angepasst, sodass eine Verschreibung eines Betäubungsmittels zum Zwecke der Selbsttötung auch in Kongruenz mit den betäubungsmittelrechtlichen Regelungen steht, sofern die Voraussetzungen dieses Gesetzes vorliegen.

Die Änderungen in § 4 und 12 sind erforderlich, da das BtMG bislang das Konstrukt einer verschreibungsgleichen Erlaubnis, wie § 8 Suizidhilfegesetz es vorsieht, nicht kennt.

Die Änderung in § 13 Absatz 1 erfolgt, um klarzustellen, dass die Verschreibung eines Betäubungsmittels zum Zwecke der Selbsttötung erlaubt ist. Die in Anlage III bezeichneten Betäubungsmittel (verkehrsfähige und verschreibungsfähige Betäubungsmittel) dürfen nur von Ärzten und nur dann verschrieben werden, wenn ihre Anwendung impliziert ist. Bislang war gesetzlich nicht explizit geregelt, dass die Anwendung auch dann begründet ist, wenn ein Arzt oder eine Ärztin das Betäubungsmittel einem Patienten oder einer Patientin zum Zwecke der Selbsttötung verschreibt. Nach der einschlägigen Kommentarliteratur ist die Anwendung begründet, wenn nach anerkannten Regeln der ärztlichen Wissenschaft eine Indikation für die Anwendung des Betäubungsmittels besteht, also das Mittel im Rahmen einer medizinischen Behandlung zu therapeutischen Zwecken eingesetzt werden soll und als Heilmittel für das Leiden von Patienten und Patientinnen geeignet ist. Bislang war nicht abschließend geklärt, ob die Verschreibung zum Zwecke eines freiverantwortlichen Suizids die Anwendung begründet. Zum Teil wurde dies verneint (vgl. Patzak/Volkmer/Fabricius/Patzak, 10. Aufl. 2022, BtMG § 13 Rn. 17; WD-Gutachten - WD 9 - 3000 - 020/20). Diese Unsicherheit soll durch die Änderung in § 13 Absatz 1 BtMG und den expliziten Verweis auf das Suizidhilfegesetz beseitigt werden. Durch den Verweis auf das Suizidhilfegesetz wird also klargestellt, dass eine Verschreibung zum Zwecke der Selbsttötung im Rahmen des Verfahrens des Suizidhilfegesetzes erfolgen darf.

Konkret in Bezug auf Natrium-Pentobarbital, welches in anderen Ländern (wie der Schweiz, Belgien und den Niederlanden) primär als Mittel zur Selbsttötung verwendet wird und z.B. in Österreich explizit als

geeignetes Präparat im Gesetzestext genannt wird, bedeutet dies, dass eine Verschreibung durch einen Arzt oder eine Ärztin gem. § 13 Absatz 1 BtMG möglich ist. Denn Barbital ist in Anlage III zu § 1 BtMG (verkehrsfähige und verschreibungsfähige Betäubungsmittel) genannt.

§ 2 Absatz 1 der Betäubungsmittelverschreibungsverordnung (BtMVV) bestätigt dies, da dort festgelegt ist, dass ein Arzt oder eine Ärztin für einen Patienten oder eine Patientin die in Anlage III des Betäubungsmittelgesetzes bezeichneten Betäubungsmittel außer Alfentanil, Cocain, Etorphin, Remifentanil und Sufentanil verschreiben darf. Barbital ist bei diesen Ausnahmen nicht genannt.

Die Änderung in § 13 Absatz 2 erfolgt, damit auch im Falle einer Erlaubniserteilung nach § 8 die Abgabe durch die Apotheken möglich ist. Auch hier wird explizit auf das Suizidhilfegesetz Bezug genommen.

Die Änderung in § 29 wird vorgenommen, um den Bezug der Strafvorschrift zu konkretisieren.

Zu Artikel 3 (Änderung des Strafgesetzbuches)

Die gemäß § 4 beratende Person kann je nach konkreter Ausgestaltung der Beratung einer suizidwilligen Person Einblicke in die Intim- und Privatsphäre, den Gesundheitszustand und die soziale Vernetzung der beratenen Person erhalten. Es ist davon auszugehen, dass sich die beratenen Menschen, die sterben möchten, ihrem Gegenüber öffnen. Dies kann jedoch nur gewährleistet werden, wenn die beratene Person weiß, dass der Inhalt der Beratung nicht an unberechtigte Dritte oder die Öffentlichkeit gelangt. Eine Verschwiegenheitsverpflichtung der beratenden Person über die in der Beratung gewonnenen Informationen ist deshalb unerlässlich.

Zu Artikel 4 (Änderung des Arzneimittelgesetzes)

Die Änderung in § 48 Abs.1 AMG sichert den Gleichlauf von ärztlicher Verschreibung nach § 6 und 7 und Erlaubniserteilung nach § 8 auch in arzneimittelrechtlicher Hinsicht. Die Änderung des § 81 AMG erfolgt als zusätzliche Klarstellung.

Zu Artikel 5 (Änderung des Heilmittelwerbegesetzes)

Durch die Ausweitung des Anwendungsbereichs des HWG auf Suizidhilfe finden die Verbote und Vorgaben des HWG auf Publikumswerbung für jegliche Form der Suizidhilfe Anwendung. So ist insbesondere eine irreführende Werbung für Suizidhilfe (§ 3 HWG) verboten. Zudem sind die Vorgaben für die Publikumswerbung in § 11 Absatz 1 HWG einzuhalten. Diese Ausweitung des Anwendungsbereichs soll gewährleisten, dass für Suizidhilfe nicht in irreführender, unsachlicher oder anstößiger Form geworben werden kann. Im Gegensatz zu den Regelungen in den Berufsordnungen der Landesärztekammern, die sich ausschließlich an Ärzte und Ärztinnen richten, gelten die Vorgaben des HWG für jedermann.

Zu Artikel 6 (Inkrafttreten)

Die Vorschrift regelt das Inkrafttreten des Gesetzes.

Deutscher Bundestag
Ausschuss f. Gesundheit

Ausschussdrucksache
20(14)38

23.05.2022

Bundesrechnungshof

Bericht nach § 88 Absatz 2 BHO

an den Haushaltsausschuss des Deutschen Bundestages,

Ausschuss für Arbeit und Soziales des Deutschen Bundestages und

Ausschuss für Gesundheit des Deutschen Bundestages

Demografische Entwicklung: Finanzrisiken des Bundes aus seiner Beteiligung an der Finanzierung der Sozialversicherungen

Dieser Bericht enthält das vom Bundesrechnungshof abschließend im Sinne des § 96 Absatz 4 BHO festgestellte Prüfungsergebnis. Die Entscheidung über eine Weitergabe bleibt dem Bundesrechnungshof vorbehalten.

Gz.: IX 4 - 2020 - 0523 Potsdam, den 23. Mai 2022

Dieser Bericht des Bundesrechnungshofes ist urheberrechtlich geschützt.
Eine Veröffentlichung ist nicht zulässig.

Bevölkerungsentwicklung führt zu Finanzrisiken für die Sozialversicherungen und den Bund

Die Menschen in Deutschland leben immer länger, die Bevölkerung schrumpft und der Anteil der Arbeitenden sinkt. Dadurch steigen die Ausgaben der Sozialversicherungen stärker als die Beitragseinnahmen. Der Bund gleicht dieses Defizit immer stärker durch Steuermittel aus. Er muss daher Maßnahmen ergreifen, die eine dauerhafte und für alle Beteiligten tragfähige Finanzierung der Sozialversicherungen sicherstellen.

Worum geht es?

Noch stärker als bereits in der Vergangenheit wird sich die Bevölkerungsstruktur in Deutschland in den nächsten Jahrzehnten verändern. Dadurch steigen die Ausgaben für die Krankenversicherung, Rentenversicherung und Pflegeversicherung (Sozialversicherungen) schneller als deren Einnahmen. Die Beiträge von Arbeitnehmer- und Arbeitgeberseite müssten deshalb steigen. Die finanzielle Unterstützung des Bundes für diese Sozialversicherungen belief sich schon im Jahr 2021 auf über 120 Mrd. Euro und könnte sich bis zum Jahr 2060 verdoppeln. Je mehr Steuermittel in die Sozialversicherungssysteme fließen, umso weniger werden künftige Regierungen in der Lage sein, anstehende Aufgaben aus dem jeweiligen Bundeshaushalt zu finanzieren.

Was ist zu tun?

Der Bund muss die Finanzierung der Sozialversicherungen langfristig sichern und die Lasten auf alle Generationen gerecht verteilen. Die hierfür in Betracht kommenden Maßnahmen müssen auf ihre finanziellen Auswirkungen hin analysiert, ergebnisoffen diskutiert und anschließend politisch bewertet werden. Der Bund sollte die Höhe der Beiträge für die Sozialversicherungen und seine finanzielle Unterstützung langfristig planen. Die Bundesregierung ist demgegenüber der Auffassung, dass solch langfristige Planungen mit zu großen Unsicherheiten behaftet sind und verweist auf den aktuellen Koalitionsvertrag.

Was ist das Ziel?

Das System muss an die sich verändernde Altersstruktur der Bevölkerung angepasst und die finanzielle Unterstützung durch den Bund begrenzt werden. Nur so bleibt der finanzielle Handlungsspielraum des Bundes auch in der Zukunft erhalten. Die hier vorgelegte Analyse der Auswirkungen ausgewählter Reformansätze soll die Befassung der Politik mit diesem wichtigen Thema auf breiterer Basis anstoßen.

Inhaltsverzeichnis

0	Zusammenfassung	9
1	Demografische Entwicklung als Finanzrisiko für den Bundeshaushalt	14
1.1	Auswirkungen der demografischen Entwicklung auf die Sozialversicherung	14
1.2	Risiken für den Bundeshaushalt durch die Auswirkungen der demografischen Entwicklung auf die Sozialversicherung	15
1.3	Mangelnde Transparenz	17
1.4	Sozialversicherungspolitik nach Kassenlage	18
2	Annahmen zur Demografie, zum Arbeitsmarkt und zur gesamtwirtschaftlichen Entwicklung	21
2.1	Demografische Annahmen der Bevölkerungsvorausberechnung bis zum Jahr 2060	21
2.2	Annahmen zur Arbeitsmarktentwicklung	25
2.3	Annahmen zur gesamtwirtschaftlichen Entwicklung	26
3	Gesetzliche Rentenversicherung	26
3.1	Einnahmen	27
3.1.1	Beitragseinnahmen	27
3.1.2	Bundeszuschüsse	27
3.2	Ausgaben	28
3.2.1	Rentenausgaben	28
3.2.2	Rentenformel und Rentenanpassungsformel	29
3.2.3	Rentenniveau	30
3.3	Beitragssatz	30
3.4	Abhängigkeit der gesetzlichen Rentenversicherung von der demografischen Entwicklung	31
3.5	Entwicklung bis zum Jahr 2060	32

3.5.1	Annahmen für die Modellrechnungen	32
3.5.2	Ergebnisse der Modellrechnungen	33
4	Gesetzliche Krankenversicherung	37
4.1	Einnahmen	37
4.1.1	Beitragseinnahmen	38
4.1.2	Bundeszuschuss	39
4.2	Ausgaben	39
4.2.1	Ausgabenentwicklung seit dem Jahr 2010	39
4.2.2	Ursachen für Ausgabensteigerungen	39
4.2.3	Ausgabenanstieg durch die Änderung der Altersstruktur	40
4.2.4	Ausgabenanstieg durch den medizinisch-technischen Fortschritt	41
4.2.5	Zuzahlungen	42
4.3	Abhängigkeit der gesetzlichen Krankenversicherung von der demografischen Entwicklung	43
4.4	Entwicklung bis zum Jahr 2060	44
4.4.1	Annahmen für die Modellrechnungen	44
4.4.2	Ergebnisse der Modellrechnungen	45
5	Soziale Pflegeversicherung	47
5.1	Einnahmen	47
5.2	Ausgaben	48
5.2.1	Pflegebedürftigkeit und Anzahl der Pflegebedürftigen	48
5.2.2	Leistungspauschalen und Leistungsausgaben	48
5.3	Beitragssatz	49
5.4	Pflegevorsorgefonds	49
5.5	Abhängigkeit der sozialen Pflegeversicherung von der demografischen Entwicklung	50
5.6	Entwicklung bis zum Jahr 2060	52
5.6.1	Annahmen für die Modellrechnungen	52
5.6.2	Ergebnisse der Modellrechnungen	53

6	Finanzielle Gesamtsituation der Sozialversicherung und deren Risiken für den Bund	56
6.1	Projektion für die betrachteten Sozialversicherungszweige	56
6.2	Ausgabenentwicklung bis zum Jahr 2060	56
6.3	Entwicklung der Sozialversicherungsbeiträge	58
6.4	Mittel aus dem Bundeshaushalt	59
6.5	Gesamte öffentliche Finanzen und Tragfähigkeitskennziffern	61
6.6	Fiskalische Tragfähigkeit der öffentlichen Haushalte	64
6.7	Auswirkungen von Alternativvarianten	66
6.8	Bewertung der Ergebnisse des Basisszenarios und mögliche Lösungsansätze	68
7	Reformen in der Alterssicherung	70
7.1	Stellschrauben und Instrumente in der gesetzlichen Rentenversicherung	70
7.2	Stellschraube Rentnerquotient	70
7.2.1	Instrument - Anhebung des gesetzlichen Renteneintrittsalters	70
7.2.2	Instrument - Ausweitung des Versichertenkreises	73
7.3	Stellschraube Rentenniveau	77
7.3.1	Instrument - Rentenanpassungsformel	77
7.3.2	Instrument - Rentenniveau-Haltelinie	77
7.3.3	Instrument - Reaktivierung des Nachholfaktors	79
7.3.4	Instrument - Anhebung der Beitragsbemessungsgrenze	81
7.4	Stellschraube Beitragssatz	81
7.4.1	Instrument - Beitragssatzhaltelinie	81
7.5	Ausbau der zusätzlichen Altersvorsorge	82
8	Reformen in der gesetzlichen Krankenversicherung	85

8.1 Stellschrauben und Instrumente in der gesetzlichen Krankenversicherung 85

8.2 Stellschraube - Versichertenstruktur 85

8.2.1 Instrument - Erweiterung des Versichertenkreises um die Selbstständigen 86
8.2.2 Instrument – Erweiterung des Versichertenkreises um Beamtinnen und Beamte 86

8.3 Stellschraube durchschnittliche beitragspflichtige Einnahmen 87

8.3.1 Instrument – Ausweitung der beitragspflichtigen Einnahmearten 87
8.3.2 Instrument – Erhöhung der Beitragsbemessungsgrenze 88
8.3.3 Instrument – Beitragszuschläge bei vorzeitigem Eintritt in die Rente 88

8.4 Stellschraube Ausgaben je Versicherten 89

8.4.1 Instrument – Begrenzung der Ausgaben 89
8.4.2 Instrument – Zuzahlungen 89

9 Reformen in der Pflegeversicherung 90

9.1 Stellschrauben und Instrumente in der sozialen Pflegeversicherung 90

9.2 Stellschraube Leistungsbeziehendenquotient 91

9.2.1 Instrument - Ausweitung des Versichertenkreises 91
9.2.2 Instrument - Reduktion der Wahrscheinlichkeit, pflegebedürftig zu werden 91

9.3 Stellschraube Leistungsniveau 92

9.3.1 Instrument - Erhöhung des beitragspflichtigen Einkommens bestimmter Personengruppen 92
9.3.2 Instrument – Änderung der Leistungspauschalen 92

9.4 Ausbau des Pflegevorsorgefonds 93

10 Kombinierte Effekte der einzelnen Reformelemente 95

10.1 Auswirkungen der Reformelemente auf die gesetzliche Rentenversicherung 96

10.2	Auswirkungen der Reformelemente auf die gesetzliche Krankenversicherung und die soziale Pflegeversicherung	99
10.3	Auswirkungen auf den Gesamtsozialversicherungsbeitragssatz und den Bundeshaushalt	100
11	Würdigung und Empfehlung	103
11.1	Reformen wegen demografischer Situation unausweichlich	103
11.2	Exemplarische Reformszenarien verbessern die finanzielle Lage	104
11.3	Sofortiges Handeln der politisch Verantwortlichen nötig	107
12	Gemeinsame Stellungnahme des Bundesministeriums für Arbeit und Soziales, des Bundesministeriums für Gesundheit sowie des Bundesministeriums der Finanzen	109
12.1	Allgemeine Vorbemerkungen und Anmerkungen zu den Simulationsberechnungen	109
12.2	Stellungnahme zu den Vorausberechnungen	111
12.3	Finanzielle Gesamtsituation der Sozialversicherung und deren Risiken für den Bund	112
12.4	Stellungnahme zu den Reformszenarien	112
13	Abschließende Würdigung des Bundesrechnungshofes	115
13.1	Allgemeine Vorbemerkungen und Anmerkungen zu den Simulationsberechnungen	115
13.2	Abschließende Würdigung zu den Vorausberechnungen	116
13.3	Abschließende Würdigung zu den Reformszenarien	118

Abkürzungsverzeichnis

B

BAS *Bundesamtes für Soziale Sicherung*
BBG *Beitragsbemessungsgrenze*
BIP *Bruttoinlandprodukte*
BMAS *Bundesministerium für Arbeit und Soziales*
BMF *Bundesministerium der Finanzen*
BMG *Bundesministerium für Gesundheit*

G

GKV *gesetzlichen Krankenversicherung*
GRV *gesetzlichen Rentenversicherung*
GVWG *Gesetz zur Weiterentwicklung der Gesundheitsversorgung, Gesundheitsversorgungsweiterentwicklungsgesetz*

K

KVdR *Krankenversicherung der Rentner*

P

PSG *Pflegestärkungsgesetz*

S

SPV *sozialen Pflegeversicherung*

0 Zusammenfassung

Der Bundesrechnungshof hat auf der Grundlage der Bevölkerungsprojektion bis zum Jahr 2060 die finanziellen Auswirkungen der demografischen Entwicklung auf die Sozialversicherungszweige und den Bund projizieren lassen. Auf dieser Basis hat er analysiert, wie sich ausgewählte Reformvorschläge auswirken. Die Analyse umfasst die Ausgaben der gesetzlichen Rentenversicherung (GRV), der gesetzlichen Krankenversicherung (GKV) und der sozialen Pflegeversicherung (SPV). Dabei wird ein Gesamteffekt etwaiger Reformen simuliert und werden Wechselwirkungen zwischen einzelnen Sozialversicherungszweigen erfasst. Das Bundesministerium für Arbeit und Soziales (BMAS), das Bundesministerium für Gesundheit (BMG) sowie das Bundesministerium der Finanzen (BMF) haben zu dem Bericht am 18. Januar 2022 Stellung genommen.

0.1 Analyse der finanziellen Auswirkungen der demografischen Entwicklung und von Reformvorschlägen auf die Sozialversicherungszweige und den Bund

Die durchschnittliche Lebenserwartung der Bevölkerung in Deutschland bei der Geburt hat sich innerhalb von rund 150 Jahren auf derzeit 78,4 Jahre für Männer und 83,2 Jahre für Frauen mehr als verdoppelt. Zudem nimmt die Anzahl der über 80-Jährigen, bei denen ein erhöhtes Pflegerisiko besteht, stark zu. Gleichzeitig verharrte die durchschnittliche Geburtenrate lange Zeit auf einem niedrigen Niveau von 1,4 Kindern je Frau und erreichte erst zuletzt knapp 1,6 Kinder je Frau.

Aufgrund der demografischen Entwicklung ist zu erwarten, dass die Ausgaben der GRV, der GKV und der SPV stark ansteigen. Diese Sozialversicherungen werden im Umlagesystem finanziert. Die erwerbstätige Bevölkerung trägt mit ihren Beitragszahlungen in erheblichem Umfang die laufenden Ausgaben dieser Sozialversicherungszweige. Daneben wird bereits heute ein bedeutender Anteil der Ausgaben vom Bund, also aus Mitteln der Steuerzahler, finanziert. Eine Kapitaldeckung zur Finanzierung dieser Ausgaben ist kaum vorhanden. Zu erwarten ist, dass die erwerbstätige Bevölkerung schrumpft und die Anzahl der Personen im Seniorenalter steigt, insbesondere durch den Renteneintritt der geburtenstarken Jahrgänge in den nächsten Jahren (sog. „Babyboomer“, also die im Zeitraum der Jahre 1955 bis 1970 Geborenen). Diese Entwicklung stellt die sozialen Sicherungssysteme vor enorme finanzielle Herausforderungen.

Diese Herausforderungen wurden bislang nicht ausreichend beachtet. Die Fehlentwicklungen betreffen die folgenden Bereiche:

- Die Sozialversicherungszweige sind nicht ausreichend auf die demografische Entwicklung vorbereitet. Dies birgt Risiken für den Bundeshaushalt.
- Die Finanzierung der Sozialversicherungszweige und deren Verpflichtungen sind nicht transparent.
- Die Sozialversicherungspolitik ist nicht langfristig, sondern kurzfristig ausgerichtet (Politik nach Kassenlage).

Der Bundesrechnungshof nimmt die absehbaren demografischen Veränderungen und die sich daraus ergebenden finanziellen Risiken für die Sozialversicherungszweige und den Bundeshaushalt zum Anlass, Maßnahmen zur Reduktion der finanziellen Risiken abzuwägen und Vorschläge zu analysieren. Der Bundesrechnungshof hält dabei am Modell des Generationenvertrages und am Solidarprinzip fest. Die darauf beruhenden Umlagesysteme haben wirtschaftliche Krisen überstanden und auch in schwierigen Zeiten ein angemessenes Versorgungsniveau gewährleistet. Die untersuchten Maßnahmen sollen eine andere Lastenverteilung auf Personen im Seniorenalter, auf die erwerbstätige Bevölkerung sowie zwischen den Generationen bewirken.

Der Bundesrechnungshof hat für seine Projektion der finanziellen Entwicklung der Sozialversicherungszweige und des Bundeshaushaltes moderate Annahmen zu Demografie, Arbeitsmarkt und zur gesamtwirtschaftlichen Entwicklung zugrunde gelegt. Die Berechnungen für die Projektion hat Prof. Dr. Werding von der Ruhr-Universität Bochum auf der Grundlage der Vorgaben des Bundesrechnungshofes erstellt.

Die demografische Entwicklung wird daran deutlich, dass im Jahr 2020 noch einer Person im Seniorenalter mehr als drei Personen im Erwerbsalter gegenüberstanden. Im Jahr 2060 kommen auf eine Person im Seniorenalter voraussichtlich nur noch zwei Personen im Erwerbsalter.

Auf der Grundlage dieser Annahmen steigen die Gesamtausgaben der Sozialversicherungszweige GRV, GKV und SPV, unter Berücksichtigung u. a. der Arbeitslosenversicherung, von 21,8 % des Bruttoinlandproduktes (BIP) im Jahr 2019 auf 29,3 % des BIP im Jahr 2060.

In den letzten Jahren lag der Gesamtsozialversicherungsbeitragssatz (Summe der Beitragssätze für die GRV, GKV, SPV und die Arbeitslosenversicherung) bei knapp 40 %. Auf der Basis der projizierten Ausgaben müsste der Gesamtsozialversicherungsbeitragssatz bei derzeitiger Gesetzeslage bis zum Jahr 2060 kontinuierlich auf 53,3 % steigen. Die Sozialversicherungszweige erhalten dauerhafte Zuschüsse aus dem Bundeshaushalt u. a. für versicherungsfremde Leistungen. In der Projektion steigen diese Bundesmittel von rund 121 Mrd. Euro im Jahr 2019 auf 454 Mrd. Euro bis zum Jahr 2060 an. Das entspricht einem Anstieg der Bundesmittel von 3,5 % des BIP im Jahr 2019 auf voraussichtlich 4,9 % bis zum Jahr 2060.

Bliebe der Gesamtsozialversicherungsbeitragssatz dauerhaft auf dem derzeitigen Niveau von 40 % eingefroren und würde der Bund die steigenden Ausgaben alleine finanzieren, würden die Bundesmittel für die Sozialversicherungen bei aktueller Rechtslage von 3,5 % des BIP auf 10,1 % des BIP steigen. Ein Vergleich verdeutlicht das Ausmaß dieses Ausgabenanstieges: Im Jahr 2019 hatte der gesamte Bundeshaushalt einen Anteil von 13,0 % am BIP. Wenn der Bund diese Zahlungen an die Sozialversicherungszweige durch zusätzliche Verschuldung finanzieren würde, stiege diese auf 141,6 % des BIP.

Ausgehend von dieser prognostizierten Entwicklung der finanziellen Situation der Sozialversicherungszweige bzw. des Bundes untersuchte der Bundesrechnungshof die finanziellen Auswirkungen folgender Reformelemente:

- Einführung einer am Anstieg der Lebenserwartung orientierten dynamischen Regelaltersgrenze in der GRV,
- Einbeziehung aller noch nicht obligatorisch versicherten Selbstständigen (unter 40 Jahren) in die GRV,
- Einhaltung eines Mindestrentenniveaus in der GRV von 44 %, 46 % und 48 %,
- Einführung einer obligatorischen, ergänzenden Altersvorsorge für alle sozialversicherungspflichtig Beschäftigten (mit einer jährlichen Sparquote von 3 % der Bruttoentgelte),
- Wiedereinsetzung des „Nachholfaktors“ bei Rentenanpassungen,
- Erhöhung der Zuzahlungen Versicherter zu Leistungen der GKV,
- Erhebung höherer Beiträge von Frührentnern in GKV und SPV zur Kompensation von Beitragsausfällen durch vorzeitige Renteneintritte,
- Einbeziehung neu ernannter Beamter in GKV und SPV und
- Ausbau des Pflegevorsorgefonds zur Stabilisierung des SPV-Beitragssatzes (ab dem Jahr 2022).

Werden alle dargestellten Reformelemente kombiniert, wird der Anstieg des Gesamtsozialversicherungsbeitragssatzes um 4,5 Prozentpunkte gedämpft. Er läge im Jahr 2060 bei 48,8 % und nicht bei 53,3 %. Die Gesamtausgaben in den Sozialversicherungen, unter Berücksichtigung u. a. der Arbeitslosenversicherung, verringern sich moderat bis zum Jahr 2060 auf 28,8 % des BIP anstatt auf 29,3 % ohne die Reformen.

Auch die Verschuldung des Bundes fiele durch die Reformen niedriger aus, wenn der Gesamtsozialversicherungsbeitragssatz auf 40 % eingefroren bliebe und der Bund die höheren Ausgaben durch eine höhere Verschuldung finanzierte. Mit den aufgezeigten Reformen würde die Verschuldung des Bundes bis zum Jahr 2060 nur auf 128 % des BIP ansteigen statt auf 141,6 %.

Veränderungen der Annahmen wie niedrigere Geburtenrate, höhere Lebenserwartung oder höheres Produktivitätswachstum haben nur begrenzte Auswirkungen auf die Höhe des Gesamtsozialversicherungsbeitragssatzes, des Bundeszuschusses oder auf die Höhe der Verschuldung des Bundes.

Die Sozialversicherungszweige finanzieren steigende Ausgaben aktuell im Wesentlichen durch steigende Sozialversicherungsbeiträge und steigende Bundesmittel. Dadurch bringen sie Einnahmen und Ausgaben zum Ausgleich. Hierbei besteht jedoch die Gefahr, dass zukünftige Generationen von Beitragszahlenden und Steuerzahlenden zu stark belastet werden, die wirtschaftliche Entwicklung gefährdet wird und der Bund an die Grenzen seiner finanziellen Leistungsfähigkeit stößt. Deshalb sollte der Gesetzgeber handeln.

Der Bundesrechnungshof vertritt aufgrund seiner Analyse die Auffassung, dass die politisch Verantwortlichen nicht umhinkommen werden, die verlautbarten politischen Ziele Beitragssatzstabilität, Erhaltung des derzeitigen Versorgungsniveaus sowie keine Steuererhöhungen und keine Neuverschuldung aufzugeben. Von diesen selbst angelegten Fesseln sollten sie sich befreien. Nur durch eine Erhöhung der Sozialversicherungsbeiträge und durch Leistungsverminderungen flankiert durch Bundeszuschüsse können nach Auffassung des Bundesrechnungshofes die Sozialversicherungen zukunftsfest gemacht werden.

Erforderlich ist ein politisches Programm für alle Sozialversicherungszweige, in dem Maßnahmen und deren finanzielle Auswirkungen aufgeführt sind. Bei der Auswahl dieser Maßnahmen wird darüber entschieden, in welchem Umfang Leistungen gekürzt, Sozialversicherungsbeiträge erhöht und die finanziellen Lasten durch eine höhere Verschuldung des Bundes auf kommende Generationen abgewälzt werden. Welche Maßnahmen umgesetzt werden sollen, müssen die politisch Verantwortlichen entscheiden.

0.2 Stellungnahme der Bundesministerien

Das BMAS, das BMG sowie das BMF (Bundesministerien) führten in ihrer gemeinsamen Stellungnahme aus, dass Projektionen für sehr lange Zeiträume (hier bis zum Jahr 2060) auf Basis des geltenden Rechtes keine Prognosen darstellten, sondern nur Anhaltspunkte für im Zeitablauf auftretenden politischen Handlungsbedarf geben könnten. Zu den Simulationsrechnungen wiesen die Bundesministerien darauf hin, dass bei derart langen Projektionszeiträumen selbst vermeintlich sichere Vorausberechnungen zur Demografie großen Unsicherheiten unterlägen. Aus den Modellrechnungen könne auf Anhaltspunkte für Handlungsbedarf geschlossen werden, der unter den getroffenen unterschiedlichen Annahmen im Projektionszeitraum entstehen würde. Ungeachtet dessen lieferten die im Bericht aufgezeigten Berechnungen grundsätzlich keine neuen Erkenntnisse, die über die bekannten Studien mit ebensolchen Berechnungen hinausgingen.

Die Bundesministerien führen in ihrer Stellungnahme aus, dass zu einigen der vom Bundesrechnungshof vorgebrachten Reformmaßnahmen die Parteien der Regierungskoalition mit dem am 7. Dezember 2021 unterzeichneten Koalitionsvertrag bereits konkrete Pläne vorgelegt hätten. Die übrigen Vorschläge lehnten die Bundesministerien aus unterschiedlichen Gründen ab.

0.3 Abschließende Würdigung

Der Bundesrechnungshof stimmt mit den Bundesministerien darin überein, dass die Berechnungen bis zum Jahr 2060 keine Prognosen darstellen. Vielmehr handelt es sich um langfristige Vorausberechnungen (Projektionen) unter Berücksichtigung bestimmter Annahmen zur wirtschaftlichen und demografischen Entwicklung. Der Bundesrechnungshof hat dabei die Modellannahmen so gestaltet, dass die Simulationsergebnisse ein mittleres Szenario darstellen.

Da das gesamte Sozialversicherungssystem und auch der Bundeshaushalt berücksichtigt werden, kann ein Gesamteffekt der Reformen simuliert und können Wechselwirkungen zwischen den Sozialversicherungszweigen erfasst werden. So können Möglichkeiten und Grenzen von Reformen der Sozialversicherungszweige aufgezeigt werden. Ein solch übergreifender Ansatz wurde in anderen Studien bisher kaum verfolgt.

Der Bundesrechnungshof hatte bei der Analyse der Reformszenarien auch das Ziel, zu versuchen, die mit der demografischen Entwicklung einhergehenden finanziellen Belastungen gleichmäßiger auf die Generationen zu verteilen.

Soweit die Bundesministerien bestimmte Reformmaßnahmen ablehnen, gilt es zu bedenken, dass ohne solche Maßnahmen die künftige Beitragsbelastung und die Verschuldung des Bundes höher sind und damit künftige Generationen stärker belastet werden. Soweit die Bundesministerien bei bestimmten Reformoptionen darauf verweisen, dass ähnliche Maßnahmen im Koalitionsvertrag vom 7. Dezember 2021 bereits geplant seien, bleibt abzuwarten, ob und wie die Vorhaben in einem Gesetz umgesetzt werden. Hier wird sich auch zeigen, ob bei solchen Gesetzesvorhaben dargelegt wird, in welchem Umfang sich diese Vorhaben langfristig auf die finanzielle Entwicklung in den Sozialversicherungszweigen und die Verschuldung des Bundes auswirken.

Ein Gesamtkonzept, wie die Sozialversicherungen zukunftsfest gemacht werden sollen, bietet der Koalitionsvertrag indes nicht. Der Bundesrechnungshof bleibt daher bei seiner Kernforderung, die Sozialversicherungen an die sich verändernde Altersstruktur der Bevölkerung anzupassen. Dazu sollte zunächst in einer Bestandsanalyse die Situation der Sozialversicherungszweige transparent dargelegt werden. Gesetzesvorhaben sollten eine Gesetzesfolgenabschätzung für einen langen Zeitraum enthalten. Zudem sollten die politisch Verantwortlichen auf der Basis der transparenten Bestandsanalyse ein politisches Programm für alle Sozialversicherungszweige erarbeiten, in dem die Maßnahmen und deren finanzielle Auswirkungen aufgeführt sind. Es ist nach der Auffassung des Bundesrechnungshofes die Aufgabe der politisch Verantwortlichen darzulegen, welche Maßnahmen ergriffen werden sollen, um die steigenden finanziellen Lasten in den Sozialversicherungszweigen auch in Zukunft finanzieren zu können und welche Personengruppen und Generationen dadurch belastet werden.

1 Demografische Entwicklung als Finanzrisiko für den Bundeshaushalt

1.1 Auswirkungen der demografischen Entwicklung auf die Sozialversicherung

Die Veränderung der Lebensumstände und der medizinisch-technische Fortschritt ermöglichen es, dass das Durchschnittsalter der Bevölkerung in Deutschland ansteigt. Lag die durchschnittliche Lebenserwartung bei der Geburt in den Jahren 1871/1881 noch bei 35,6 Jahren für Männer und 38,4 Jahren für Frauen, beträgt derzeit 78,4 Jahre für Männer und 83,2 Jahre für Frauen.[1] Der medizinisch-technische Fortschritt führt dazu, dass insbesondere die Älteren ein gesünderes Leben im Ruhestand führen können. Zudem nimmt die Anzahl der über 80-Jährigen stark zu. Bei ihnen besteht das größte Pflegerisiko. Damit steigt auch die Anzahl der Pflegebedürftigen. Gleichzeitig verharrte die Geburtenrate lange Zeit auf einem niedrigen Niveau von 1,4 Kinder je Frau. Erst zuletzt stieg sie auf rund 1,6 Kinder je Frau an.[2]

Diese Entwicklungen führen dazu, dass die Ausgaben der GRV, der GKV und der SPV stark ansteigen werden. Diese Sozialversicherungszweige werden im Umlagesystem finanziert. Die erwerbstätige Bevölkerung trägt mit ihren Beitragszahlungen in erheblichem Umfang die Ausgaben dieser Sozialversicherungszweige[3] für die noch nicht erwerbstätigen Kinder und für die nicht mehr erwerbstätigen Personen im Seniorenalter. Die jährlichen Einnahmen der Sozialversicherungszweige dienen fast ausschließlich der Deckung der laufenden Ausgaben. Eine Kapitaldeckung zur Finanzierung der Ausgaben ist so gut wie nicht vorhanden. Zu erwarten ist, dass die erwerbstätige Bevölkerung schrumpft und die Anzahl der Personen im Rentenalter steigt, insbesondere durch den Renteneintritt der geburtenstarken Jahrgänge in den nächsten Jahren (sog. „Babyboomer“, die im Zeitraum der Jahre 1955 bis 1970 Geborenen). Diese Entwicklung stellt die sozialen Sicherungssysteme vor enorme finanzielle Herausforderungen.

Diese Herausforderungen wurden bislang nicht ausreichend beachtet. Fehlentwicklungen betreffen die folgenden Bereiche:

- Die Sozialversicherungszweige sind nicht ausreichend auf die demografische Entwicklung vorbereitet. Dies birgt Risiken für den Bundeshaushalt.

---

1 Statistisches Bundesamt, Bevölkerung im Wandel – Annahmen und Ergebnisse der 14. koordinierten Bevölkerungsvorausberechnung, Wiesbaden 2019, Seite 36.

2 Statistisches Bundesamt, a. a. O., Seite 29.

3 Die Sozialversicherungsbeiträge lagen im Jahr 2021 bei rund 40 % des Bruttoarbeitsentgelts (14,6 % GKV; durchschnittlich 1,3 % Zusatzbeitrag GKV; 3,05 % für die SPV; 2,4 % für die Arbeitslosenversicherung; 18,6 % für die GRV) und werden im Wesentlichen je zur Hälfte von den Arbeitgebern und den Arbeitnehmern gezahlt.

- Die Finanzierung der Sozialversicherungszweige und deren Verpflichtungen sind nicht transparent.
- Die Sozialversicherungspolitik ist nicht langfristig, sondern kurzfristig ausgerichtet (Politik nach Kassenlage).

1.2 Risiken für den Bundeshaushalt durch die Auswirkungen der demografischen Entwicklung auf die Sozialversicherung

Die absehbaren demografischen Veränderungen haben finanzielle Folgen in den Sozialversicherungszweigen und dem Bundeshaushalt. Der Bundesrechnungshof nimmt dies zum Anlass, die finanziellen Risiken für die GRV, die GKV und die SPV zu analysieren und Vorschläge zur Verringerung dieser Risiken abzuwägen. Dabei ist zu berücksichtigen, dass sich Veränderungen bei einem Sozialversicherungszweig teilweise auf die Finanzen der anderen Sozialversicherungszweige auswirken. So führt zum Beispiel eine Erhöhung der Renten dazu, dass höhere Beiträge an die GKV und SPV entrichtet werden. Es ist daher erforderlich, die genannten Sozialversicherungszweige insgesamt zu betrachten, um die finanziellen Risiken zu analysieren.

Der Bundesrechnungshof geht davon aus, dass die demografischen Risiken der Sozialversicherung für Landwirtschaft, Forsten und Gartenbau, der Arbeitslosenversicherung sowie der Unfallversicherungsträger (gewerbliche Berufsgenossenschaften und der Unfallversicherungsträger der öffentlichen Hand) finanziell deutlich geringer sind.[4] Er berücksichtigt sie deshalb nicht bei seiner Analyse der Risiken für das Sozialversicherungssystem.

Grundsätzlich werden die betrachteten Sozialversicherungszweige durch Beiträge von Arbeitnehmern und Arbeitgebern je zur Hälfte finanziert (paritätische Finanzierung). Modellrechnungen wie z. B. der vom BMF in Auftrag gegebene Tragfähigkeitsbericht zeigen, dass durch die demografischen Veränderungen die Ausgaben des Sozialversicherungssystems und damit die Sozialversicherungsbeiträge stark ansteigen müssten. Werden diese Ausgabensteigerungen nicht durch höhere Sozialversicherungsbeiträge oder Leistungskürzungen ausgeglichen, müsste der Bund die Einnahmedefizite des Sozialversicherungssystems durch weitere Zuschüsse decken.[5]

Der Bundesrechnungshof hat die finanzielle Entwicklung der Sozialversicherungszweige und des Bundeshaushaltes bis zum Jahr 2060 projizieren lassen. Er legt dabei ähnliche Annahmen wie der Tragfähigkeitsbericht des BMF zugrunde. Da eine Projektion über einen derart langen Zeitraum mit vielen Unwägbarkeiten verbunden ist, verwendet er eine mittlere

[4] Stiftung Marktwirtschaft: Ehrbarer Staat? Die Generationenbilanz, Argumente zu Marktwirtschaft und Politik Nr. 142, September 2018, Seite 12.

[5] Bundesministerium der Finanzen, Fünfter Bericht zur Tragfähigkeit der öffentlichen Finanzen, Jahr 2020, Seite 30.

Variante, deren Annahmen zwischen denen der optimistischen und der pessimistischen Variante im Tragfähigkeitsbericht liegen. Der Bundesrechnungshof hält diese mittlere Variante für angemessen, um die voraussichtliche Entwicklung zu projizieren und Lösungsmöglichkeiten zu analysieren.[6]

Der Bundesrechnungshof geht davon aus, dass die absehbar höheren Ausgaben durch die demografischen Effekte nicht vollständig über höhere Beiträge ausgeglichen werden können. Dies würde die Wettbewerbsfähigkeit der Wirtschaft stark beeinträchtigen und könnte dazu führen, dass Betriebe abwandern und Arbeitsplätze verloren gehen. Auch eine höhere Belastung der Arbeitnehmerinnen und Arbeitnehmer durch steigende Sozialversicherungsbeiträge ist nur begrenzt umsetzbar.

Insgesamt werden derzeit in den Sozialversicherungszweigen die demografischen Lasten vor allem von der erwerbstätigen Bevölkerung als Beitragszahlende und Steuerzahlende sowie von zukünftigen Generationen über Schulden getragen. Es ist davon auszugehen, dass der Bund einen Teil der Ausgabensteigerung über den Haushalt finanzieren wird. Um die Mittel hierfür aufzubringen, kommen Einnahmesteigerungen (Steuern, Abgaben und Gebühren), Umschichtungen im Haushalt oder eine Erhöhung der Verschuldung in Frage. Der Bundesrechnungshof erkennt in den Ausgabensteigerungen sowie der gedämpften Einnahmeentwicklung (z. B. durch den Rückgang der Erwerbsbevölkerung) in den Sozialversicherungszweigen in jedem Fall ein erhebliches Risiko für den Bundeshaushalt.

Der Bundesrechnungshof hält bei dieser Betrachtung am Generationenvertrag fest. Denn das auf dem Generationenvertrag beruhende Umlagesystem ist dazu geeignet, wirtschaftliche Krisen zu überdauern und auch in schwierigen Zeiten ein angemessenes Versorgungsniveau zu gewährleisten. Dies hat gerade die jüngere Vergangenheit gezeigt, als in der Finanzkrise und der Corona-Pandemie das umlagefinanzierte Sozialversicherungssystem stabilisierend gewirkt und Sicherheit gegeben hat.

Der Bundesrechnungshof hält auch das Solidarprinzip in der GRV, GKV und SPV für unabdingbar. Beiträge sollen unabhängig von Alter, Geschlecht, Lebenserwartung und Gesundheitsrisiko bzw. Pflegerisiko erhoben werden. Versicherte in der GKV sollen nach dem Bedarfsprinzip Anspruch auf Leistungen haben.

Der Bundesrechnungshof erörtert auf dieser Grundlage Lösungsmöglichkeiten, durch die zum einen die finanziellen Risiken für den Bund und für die Sozialversicherungszweige vermindert werden können. Die Lösungsmöglichkeiten orientieren sich am bisherigen Sozialversicherungssystem. Grundlegende Veränderungen, wie z. B. eine Bürgerversicherung oder ein System mit Pauschalbeiträgen („Kopfpauschale“) in der GKV bzw. der SPV, werden nicht einbezogen. Zum anderen soll eine andere Lastenverteilung auf Personen im Seniorenalter, auf die erwerbstätige Bevölkerung, die den größten Teil der Steuern und Beiträge trägt, sowie zwischen den Generationen erreicht werden. Der Bundesrechnungshof beschränkt sich

[6] Die getroffenen Annahmen zur Projektion der Bevölkerung sowie den wirtschaftlichen und gesellschaftlichen Entwicklungen sind unten in Tnr. 2 dargestellt.

deshalb im Wesentlichen auf Lösungsmöglichkeiten, durch die der Bund oder die kommende, schrumpfende erwerbstätige Bevölkerung entlastet wird. Dies kann insbesondere durch folgende Maßnahmen erreicht werden:

- Intergenerativer Lastentransfer: Heute Vorsorge treffen, um die Finanzierungslast für die kommende, schrumpfende Gruppe der Erwerbstätigen zu vermindern (z. B. Fondsmodelle oder private Vorsorge).
- Stärkere finanzielle Beteiligung von Personen im Seniorenalter bei Vermeidung finanzieller Überlastung (z. B. ein etwas geringeres Rentenniveau oder Anhebung der Regelaltersgrenze).
- Beitragssatzglättung: Minderung der Beitragsbelastung in der Phase, in der die demografische Entwicklung zu einem starken Anstieg der Beitragssätze führt (z. B. durch Einbeziehung von weiteren Versichertengruppen, die Beiträge zahlen, aber Leistungen im Wesentlichen erst später in Anspruch nehmen).

Soweit die darauf beruhenden Maßnahmen zu sozialen Verwerfungen führen (zu hohe Beiträge oder zu geringe Leistungen), sind in diesem Zusammenhang auch Aspekte des sozialen Ausgleiches zu berücksichtigen. Maßnahmen, die dazu dienen, aus verteilungspolitischen Erwägungen die Beitragslast innerhalb einer Gruppe umzuverteilen (z. B. Erhöhung der Lasten für Besserverdienende und Entlastung von Geringverdienern im gleichen Ausmaß), stehen dagegen nicht im Zentrum der Analyse des Bundesrechnungshofes.

Der Bundesrechnungshof schlägt hier mögliche Maßnahmen vor (vgl. Tnr. 7 bis 10). Mit welchen Maßnahmen die genannten Ziele erreicht und in welchem Umfang die finanziellen Risiken für den Bund konkret gemindert werden sollen, entscheiden die politisch Verantwortlichen.

BMAS, BMG sowie BMF haben zu dem Bericht am 18. Januar 2022 Stellung genommen.

1.3 Mangelnde Transparenz

Die GRV, die GKV und die SPV werden auch heute schon dauerhaft durch Zahlungen des Bundes mitfinanziert. Der Bund sieht für das Haushaltsjahr 2021 bei einem Haushaltsvolumen von 498,6 Mrd. Euro[7] Zuschüsse in Höhe von 106,2 Mrd. Euro an die GRV vor[8]. Die GKV erhält grundsätzlich einen Zuschuss in Höhe von 14,5 Mrd. Euro, der sich in den Jahren 2020

[7] § 1 Absatz 1 des Gesetzes über die Feststellung des Bundeshaushaltsplanes für das Haushaltsjahr 2021 (Haushaltsgesetz 2021) vom 21. Dezember 2020 (BGBl. I Seite 3208).

[8] Gesetz über die Feststellung des Bundeshaushaltsplanes für das Haushaltsjahr 2021 (Haushaltsgesetz 2021) vom 21. Dezember 2020 (BGBl. I Seite 3208), Kapitel 1102, Titelgruppe 01.

bis 2022 für pandemiebedingte Sonderausgaben erhöht.[9] Die SPV erhält einen dauerhaften Bundeszuschuss von 1 Mrd. Euro jährlich.[10]

Der Bund finanziert mit seinen Zuschüssen an diese Sozialversicherungszweige auch pauschal versicherungsfremde Leistungen. Was versicherungsfremde Leistungen und wie hoch diese sind, lässt der Gesetzgeber offen. Damit ist unklar, was genau mit den Zuschüssen finanziert wird und ob sie ausreichen oder zu hoch sind. Da die Zuschüsse die versicherungsfremden Leistungen pauschal abgelten, steht der Gesetzgeber bei einer Ausweitung der versicherungsfremden Leistungen nicht in der Pflicht, die Bundeszuschüsse zu erhöhen.

Nicht transparent ausgewiesen ist auch die Höhe der Ansprüche der Versicherten, die im Sozialversicherungssystem durch die Zahlung von Beiträgen oder durch ein Versichertenverhältnis entstehen. Der Gesetzgeber hat dies für die Vermögensrechnung der Sozialversicherungsträger nicht vorgeschrieben. Die Versicherten in der GRV erwerben durch die Rentenversicherungsbeiträge einen individuellen, eigentumsähnlichen Anspruch auf Zahlung z. B. einer Altersrente. Die Versicherten der GKV erwerben für die Dauer des Versicherungsverhältnisses Leistungsansprüche im Krankheitsfall. In der SPV erwerben die Versicherten durch den Versichertenstatus einen Anspruch auf die Gewährung der Pflegeleistungen.

Da diese Ansprüche innerhalb der Sozialversicherungszweige wegen der grundsätzlichen Umlagefinanzierung nur geringfügig über Kapitalanlagen gedeckt sind, steht diesen Ansprüchen insoweit kein nennenswertes Vermögen gegenüber.

1.4 Sozialversicherungspolitik nach Kassenlage

Die Gesetzgebung im Bereich der Sozialversicherung umfasst seit der Jahrtausendwende im Wesentlichen drei Phasen.

Phase 1: Anfang des Jahrtausends war das politische Handeln von der anhaltenden wirtschaftlichen Stagnation, begleitet von hoher Arbeitslosigkeit und den damit einhergehenden finanziellen Problemen der Sozialversicherungen geleitet. Aber auch die absehbare Veränderung der demografischen Situation und deren Auswirkung auf die Sozialversicherungszweige fanden Beachtung. Die Bundesregierung beschloss im Rahmen der „Agenda 2010“ verschiedene Maßnahmen im Bereich der GRV und GKV. Zudem setzte sie eine Kommission (Rürup-Kommission) ein, die die demografische Situation analysierte, sich mit den daraus resultierenden finanziellen Risiken und Vorschlägen zur Verringerung der finanziellen Auswirkungen befasste und im Jahr 2003 ihre Ergebnisse präsentierte.[11] Das BMF ließ erstmalig im

[9] § 221 SGB V.

[10] Gesetz zur Weiterentwicklung der Gesundheitsversorgung (Gesundheitsversorgungsweiterentwicklungsgesetzes – GVWG) vom 11. Juli 2021 (BGBl. I Seite 2754).

[11] Bundesministerium für Gesundheit und Soziale Sicherung, Nachhaltigkeit in der Finanzierung der sozialen Sicherungssysteme, Berlin 2003.

Jahr 2005 ein Gutachten zur Tragfähigkeit der öffentlichen Finanzen[12] erstellen, das seitdem alle vier Jahre fortgeschrieben wird.

Der Gesetzgeber beschloss im Bereich der Alterssicherung im Jahr 2001 das Altersvermögensgesetz und das Altersvermögensergänzungsgesetz[13] sowie im Jahr 2004 – auf Grundlage der Ergebnisse der Rürup-Kommission – das RV-Nachhaltigkeitsgesetz.[14] Im Jahr 2006 folgte das RV-Altersgrenzenanpassungsgesetz, das ab dem Jahr 2012 eine schrittweise Anhebung der Regelaltersgrenze von 65 auf 67 Jahre vorsah.[15] Der Gesetzgeber reagierte auf die sich wandelnden demografischen, aber auch ökonomischen und gesellschaftlichen Rahmenbedingungen und wollte die Grundlagen für eine generationengerechte Rente schaffen.[16]

In der GKV hat der Gesetzgeber mehrere Reformgesetze beschlossen: GKV-Gesundheitsreform 2000, Arzneimittelausgaben-Begrenzungsgesetz 2002, Fallpauschalengesetz 2003, Beitragssatzsicherungsgesetz 2003 und das GKV-Modernisierungsgesetz 2004. Letzteres sah z. B. grundsätzlich bei allen Leistungen Zuzahlungen und die Praxisgebühr vor. Aus dem Leistungskatalog gestrichen wurden Sterbegeld, Entbindungsgeld, nicht medizinisch notwendige Sterilisation, künstliche Befruchtung (hälftige Kostenübernahme nur noch in bestimmten Fällen), Sehhilfen/Brillen, Fahrkosten zur ambulanten Behandlung. Für Zahnersatz führte der Gesetzgeber einen Beitrag von 0,5 % der beitragspflichtigen Einnahmen ein, den die Mitglieder ohne Arbeitgeberbeteiligung selbst zahlen mussten. Das Gesetz zur Stärkung des Wettbewerbes in der GKV führte u. a. den Gesundheitsfonds ein und einen Bundeszuschuss zur pauschalen Abdeckung von versicherungsfremden Leistungen.[17]

Phase 2: Durch den anschließenden wirtschaftlichen Aufschwung stiegen die Einkommen, die Arbeitslosigkeit sank und der Anteil der Erwerbstätigen in der Bevölkerung nahm zu. Dadurch nahmen die Sozialversicherungszweige deutlich mehr ein. Insbesondere ab dem Jahr 2014 nutzte die Politik dies, um die Leistungen in den Sozialversicherungszweigen auszuweiten.[18] Für die GRV beschloss der Deutsche Bundestag u. a. die abschlagsfreie Rente mit 63 Jahren für bestimmte Versicherte, die Ausweitung der rentensteigernden Anrechnung von Kindererziehungszeiten für vor dem Jahr 1992 geborene Kinder („Mütterrente I+II“) und

[12] Bundesministerium der Finanzen, Bericht zur Tragfähigkeit der öffentlichen Finanzen, Berlin 2005.

[13] Gesetz zur Reform der gesetzlichen Rentenversicherung und zur Förderung eines kapitalgedeckten Altersvermögens (Altersvermögensgesetz – AVmG) vom 26. Juni 2001 (BGBl. I S. 1310).

[14] Gesetz zur Sicherung der nachhaltigen Finanzierungsgrundlagen der gesetzlichen Rentenversicherung (RV-Nachhaltigkeitsgesetz) vom 21. Juli 2004 (BGBl. I S. 1791).

[15] Gesetz zur Anpassung der Regelaltersrente an die demografische Entwicklung und Stärkung der Finanzierungsgrundlagen der gesetzlichen Rentenversicherung (RV-Altersgrenzenanpassungsgesetz) vom 20. April 2007 (BGBl. I S. 554).

[16] Deutscher Bundestag, Entwurf eines Gesetzes zur Anpassung der Regelaltersgrenze an die demografische Entwicklung und zur Stärkung der Finanzierungsgrundlagen der gesetzlichen Rentenversicherung (RV-Altersgrenzenanpassungsgesetz), Drucksache 16/3794 vom 12. Dezember 2006, Seite 1.

[17] Gesetz zur Stärkung des Wettbewerbs in der gesetzlichen Krankenversicherung (GKV-Wettbewerbsstärkungsgesetz -GKV-WSG) vom 26. März 2007, (BGBl. I S. 378).

[18] Siehe hierzu auch: Bundesministerium der Finanzen, Fünfter Bericht zur Tragfähigkeit der öffentlichen Finanzen, Jahr 2020, Seiten 21, 22.

Verbesserungen bei Erwerbsminderungsrenten.[19] Zudem beschloss der Gesetzgeber, den Rentenwert (Ost) an den Rentenwert (West) anzugleichen.[20]

In der GKV schaffte der Gesetzgeber im Jahr 2013 die Praxisgebühr ab. Das GKV-Versichertenentlastungsgesetz regelte u. a., dass Arbeitgeber und Versicherte die Beiträge zur GKV wieder zu gleichen Teilen zahlen müssen.[21] Mit dem GKV-Betriebsrentenfreibetragsgesetz reduzierte der Gesetzgeber die Beitragslast für Pflichtversicherte der GKV, die eine Betriebsrente beziehen, indem er einen Freibetrag einführte.[22]

In der SPV verbesserte der Gesetzgeber mit den Pflegestärkungsgesetzen (PSG) die Leistungen für Pflegebedürftige. So führte der Gesetzgeber z. B. mit dem Zweiten PSG einen neuen Pflegebedürftigkeitsbegriff mit fünf Pflegegraden und ein neues Begutachtungssystem ein.[23] Durch diese Änderung stieg die Zahl der als pflegebedürftig eingestuften Personen um rund 35 %.[24]

Phase 3: Im Jahr 2020 führte die Corona-Pandemie zu einem drastischen wirtschaftlichen Einbruch.[25] Für die Sozialversicherungszweige kam es dadurch zu deutlich geringeren Einnahmen (Kurzarbeit, Arbeitsplatzabbau usw.). Diese Einnahmeausfälle werden teilweise durch den Abbau der Rücklagen einzelner Sozialversicherungszweige und durch höhere Zahlungen des Bundes aufgefangen. So erhielt die SPV im Jahr 2020 erstmalig Leistungen des Bundes von 1,8 Mrd. Euro für durch die Corona-Pandemie verursachte Belastungen.[26] Zum Ausgleich pandemiebedingter Belastungen erhielt die GKV einen zusätzlichen Bundeszuschuss von 3,5 Mrd. Euro im Jahr 2020, 5 Mrd. Euro im Jahr 2021 und 14 Mrd. Euro im Jahr 2022.[27]

Der Bund finanziert seine Mehrausgaben für die Sozialversicherungszweige über zusätzliche Schulden. Damit soll die Wirtschaft stabilisiert und der Gesamtsozialversicherungsbeitragssatz auf 40 % der beitragspflichtigen Bruttoarbeitsentgelte begrenzt werden („Sozialgarantie“).

---

[19] Gesetz über Leistungsverbesserungen in der gesetzlichen Rentenversicherung (RV-Leistungsverbesserungsgesetz) vom 23. Juni 2014 (BGBl. I S.787).

[20] Gesetz über den Abschluss der Rentenüberleitung (Rentenüberleitungs-Abschlussgesetz) vom 17. Juli 2017 (BGBl. I S.2575).

[21] Gesetz zur Beitragsentlastung der Versicherten in der gesetzlichen Krankenversicherung (GKV-Versichertenentlastungsgesetz – GKV-VEG) vom 11. Dezember 2018 (BGBl. I S. 2387).

[22] Gesetz zur Einführung eines Freibetrages in der gesetzlichen Krankenversicherung zur Förderung der betrieblichen Altersvorsorge (GKV-Betriebsrentenfreibetragsgesetz – GKV-BRG) vom 21. Dezember 2019 (BGBl. I S. 2913).

[23] Zweites Gesetz zur Stärkung der pflegerischen Versorgung und zur Änderung weiterer Vorschriften (Zweites Pflegestärkungsgesetz – PSG II) vom 21. Dezember 2015 (BGBl. I S. 2424).

[24] Bundesministerium der Finanzen, Fünfter Bericht zur Tragfähigkeit der öffentlichen Finanzen, Jahr 2020, Seite 22.

[25] Statistisches Bundesamt, Pressemitteilung 020 vom 14. Januar 2021.

[26] Gesetz über die Feststellung eines Zweiten Nachtrages zum Bundeshaushaltsplan für das Haushaltsjahr 2020 (Zweites Nachtragshaushaltsgesetz 2020) vom 14. Juli 2020 (BGBl. I Seite 1669).

[27] Weitere 300 Mio. Euro zahlt der Bund jeweils in den Jahren 2021 und 2022 an den Gesundheitsfonds zum Ausgleich von Mehrausgaben aufgrund der erweiterten Kinderkrankengeldregelung nach § 45 Absatz 2a SGB V.

Insgesamt zeigt sich, dass das politische Handeln eher von der kurzfristigen finanziellen Situation der Sozialversicherungszweige geprägt ist, als dass langfristige Risiken aufgrund der demografischen Entwicklung eine Rolle spielen.

2 Annahmen zur Demografie, zum Arbeitsmarkt und zur gesamtwirtschaftlichen Entwicklung

Der Bundesrechnungshof hat für seine Projektion der finanziellen Entwicklung der Sozialversicherungszweige und des Bundeshaushaltes moderate Annahmen zur Demografie, zum Arbeitsmarkt und zur gesamtwirtschaftlichen Entwicklung zugrunde gelegt. Die Berechnungen für die Projektion hat Prof. Dr. Werding von der Ruhr-Universität Bochum erstellt. Dabei verwendete er das gleiche Modell – allerdings mit teilweise abweichenden Modellannahmen und -spezifikationen –, das auch dem Tragfähigkeitsbericht im Auftrag des BMF zugrunde lag.[28] Die Berechnungen fanden im Jahr 2021 statt, Datengrundlage sind Zahlen aus den Jahren 2018 und 2019. Im Modell sind die Jahre 2020 bis 2060 projiziert worden.[29] Dies kann in den Jahren 2020 und 2021 zu leichten Abweichungen von den tatsächlich eingetretenen Werten führen.

2.1 Demografische Annahmen der Bevölkerungsvorausberechnung bis zum Jahr 2060

Das Statistische Bundesamt projiziert in seiner Bevölkerungsvorausberechnung die Entwicklung der Bevölkerung für Deutschland bis zum Jahr 2060. Die Projektion beruht auf bestimmten Annahmen zur Geburtenhäufigkeit, zur Lebenserwartung und zum Wanderungssaldo (Differenz zwischen Zuzügen nach und Fortzügen aus Deutschland).

Um die Unwägbarkeiten bei der Projektion der Bevölkerungsentwicklung abbilden zu können, umfasst die Bevölkerungsvorausberechnung neun Hauptvarianten, zwölf weitere Varianten sowie neun Modellrechnungen. Die Hauptvarianten sehen bei den Annahmen zur Geburtenhäufigkeit, zur Lebenserwartung und zum Wanderungssaldo bestimmte Spannbreiten vor (vgl. Tabelle 1).

[28] Werding, Martin; Läpple, Benjamin: Finanzrisiken für den Bund durch die demographische Entwicklung in der Sozialversicherung: Reformszenarien, Studie im Auftrag des Bundesrechnungshofes, Bochum Mai 2021.

[29] Spätere Entwicklungen, verursacht durch die Pandemie und aktuell durch die Ukraine-Krise, sind daher nicht berücksichtigt.

Tabelle 1

Annahmen zur Bevölkerungsprojektion bis zum Jahr 2060

Spannbreite der demografischen Komponenten der Bevölkerungsprojektion bis zum Jahr 2060		
Demografische Komponente	untere Annahme	obere Annahme
Geburtenhäufigkeit	1,4 Kinder je Frau	1,7 Kinder je Frau
Lebenserwartung	Jungen: 82,5 Jahre Mädchen: 86,4 Jahre	Jungen: 86,2 Jahre Mädchen: 89,6 Jahre
Wanderungssaldo	durchschnittlich 147 000 Personen pro Jahr	durchschnittlich 311 000 Personen pro Jahr

Quelle: Statistisches Bundesamt, Bevölkerung im Wandel – Annahmen und Ergebnisse der 14. koordinierten Bevölkerungsvorausberechnung, Wiesbaden 2019, Seiten 13, 16.

Der Bundesrechnungshof legt die mittlere Variante zugrunde. Diese geht von einer stabilen Geburtenhäufigkeit (1,55 Kinder je Frau), einem moderaten Anstieg der Lebenserwartung bei Geburt (Jungen: 84,4 Jahre, Mädchen: 88,1 Jahre) und einem moderaten positiven Wanderungssaldo (durchschnittlich 221 000 Personen pro Jahr) aus.

Die Bevölkerung würde bei dieser Variante von rund 83 Millionen Menschen im Jahr 2018 bis zum Jahr 2024 auf 83,7 Millionen Menschen zunehmen und dann bis zum Jahr 2060 auf 78,2 Millionen Menschen schrumpfen.[30]

Für die Finanzierung der Sozialversicherungen und die Risiken für den Bundeshaushalt ist jedoch insbesondere die Altersstruktur der Bevölkerung in den Jahren bis 2060 von Bedeutung. Entscheidend ist das zahlenmäßige Verhältnis der Menschen im Erwerbsalter von 20 bis 66 Jahre zu den Menschen im Seniorenalter ab 67 Jahren. Die Menschen im Erwerbsalter tragen die finanzielle Hauptlast der Sozialversicherungszweige. Die Menschen im Seniorenalter verursachen den Großteil der Ausgaben.

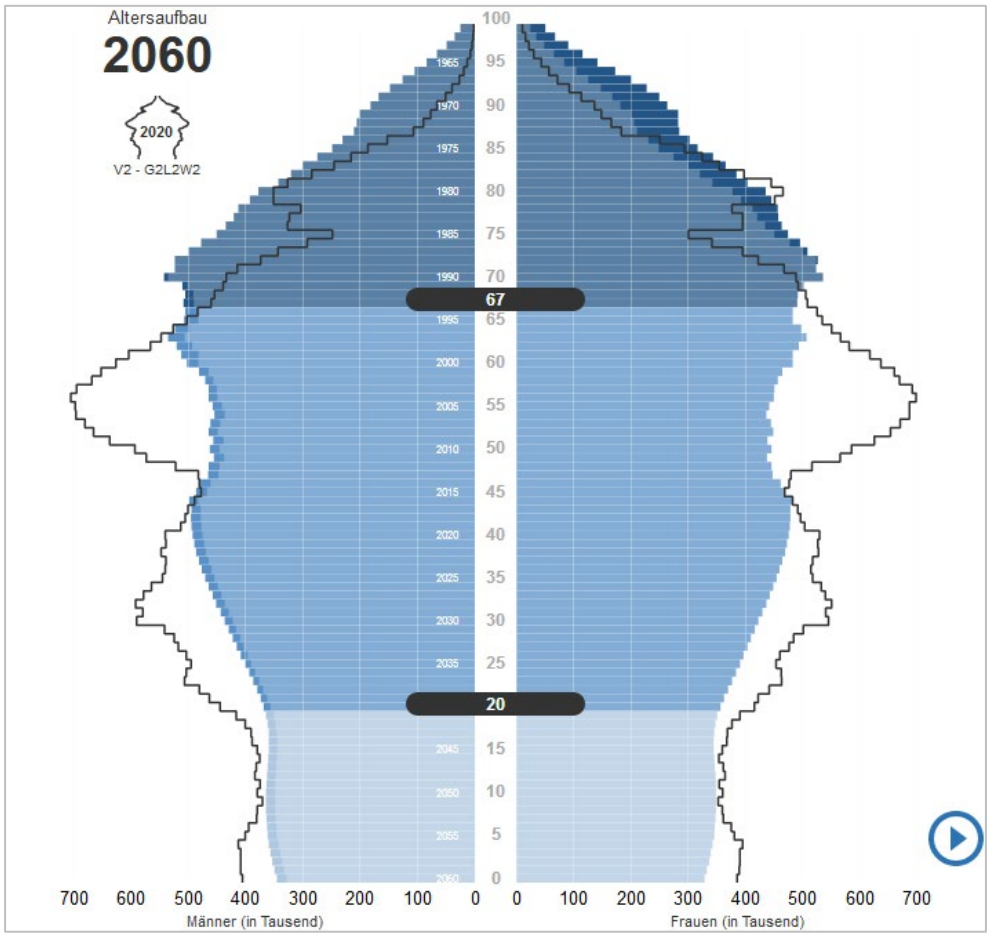
Die Bevölkerungspyramide des Jahres 2020 im Vergleich mit der Bevölkerungspyramide des Jahres 2060 (vgl. Abbildung 1) verdeutlicht anschaulich, wie die Erwerbsbevölkerung (Alter 20 bis 66 Jahre) abnimmt und die Anzahl der Menschen im Seniorenalter (Alter über 67 Jahre) zunimmt. Besonders deutlich sticht dabei die Generation der „Babyboomer“ hervor.

[30] Statistisches Bundesamt, a. a. O., Seiten 16, 17.

Abbildung 1

Die Anzahl der Personen über 67 Jahren steigt bis zum Jahr 2060 deutlich an, während die Anzahl der unter 67-Jährigen deutlich sinkt

Ein Vergleich der Alterspyramiden des Jahres 2020 mit der des Jahres 2060 (Vorausberechnung) verdeutlicht den Anstieg der über 67-Jährigen im Vergleich zu den unter 67-Jährigen.

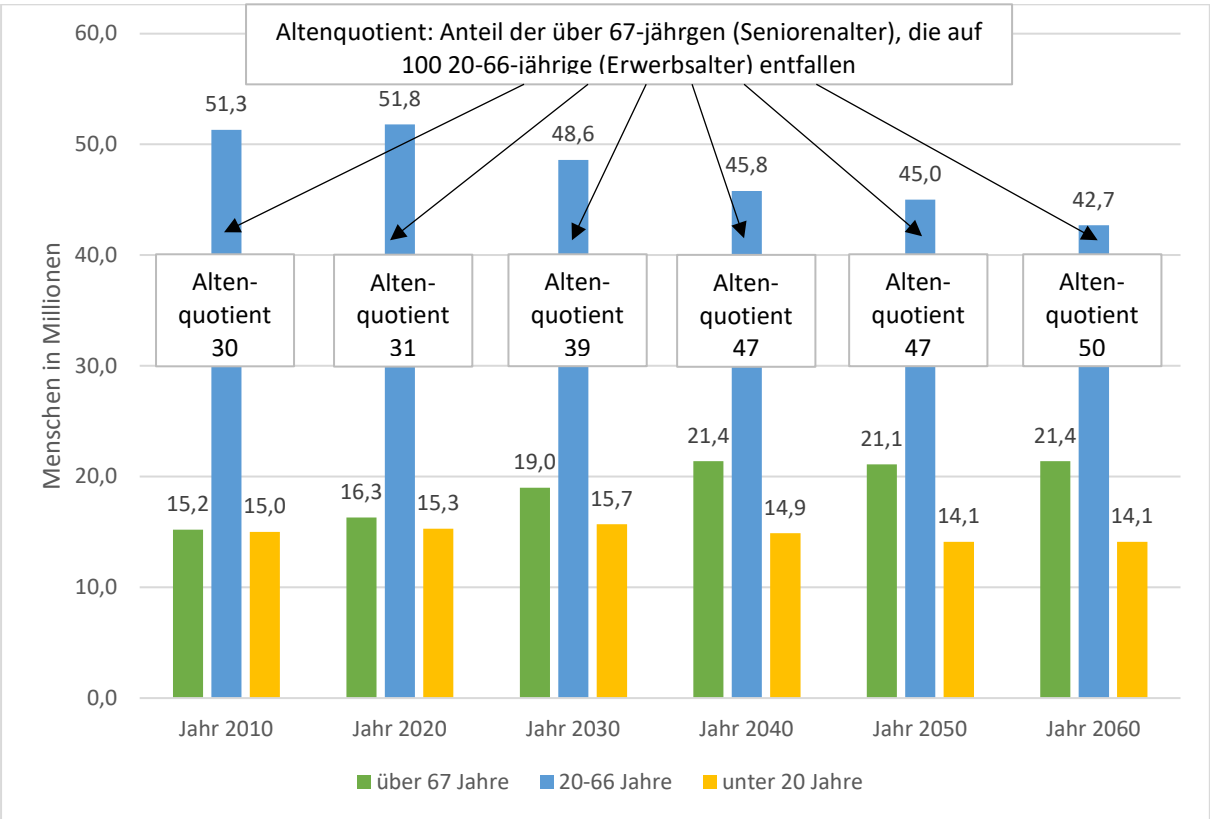


Quelle: Dynamische Bevölkerungspyramide des Statistischen Bundesamtes.

Abbildung 2

Der Altenquotient steigt bis zum Jahr 2060 stark an

Im Jahr 2020 kamen auf 31 Personen im Seniorenalter 100 Personen im Erwerbsalter. Das Verhältnis verschlechtert sich in Zukunft und erreicht im Jahr 2060 voraussichtlich 50 zu 100.



Grafik: Bundesrechnungshof.

Quelle für Grundzahlen: Statistisches Bundesamt, Bevölkerung im Wandel – Annahmen und Ergebnisse der 14. koordinierten Bevölkerungsvorausberechnung, Wiesbaden 2019, Tabellenband Bevölkerung Deutschlands bis 2060 (Hauptvarianten 1 bis 9), Variante G2-L2-W2.

Die Anzahl der Personen in der Bevölkerungsgruppe im Seniorenalter steigt im Zeitraum von 2010 bis 2060 von 15,2 auf 21,4 Millionen Personen an. Gleichzeitig schrumpft die Bevölkerungsgruppe im Erwerbsalter von über 51 Millionen Personen auf 42,7 Millionen Personen im Jahr 2060 (vgl. Abbildung 2).

Diese Veränderung der Altersstruktur der Bevölkerung verdeutlicht der sogenannte Altenquotient. Diese Kennzahl setzt die Größe der Bevölkerungsgruppe im Seniorenalter ins Verhältnis zur Größe der Bevölkerungsgruppe im erwerbsfähigen Alter (bezogen auf 100 Personen). Im Jahr 2020 ist der Altenquotient mit 31 im Vergleich zum Jahr 2010 relativ stabil. Das heißt, auf 100 Personen im Erwerbsalter kamen 31 Personen im Seniorenalter. Im Jahr 2030 steigt der Altenquotient voraussichtlich deutlich auf 39, im Jahr 2040 nochmals stark auf 47 an. Anschließend verlangsamt sich der Anstieg. Im Jahr 2060 kommen auf 100 Personen im Erwerbsalter 50 Personen im Seniorenalter (vgl. Abbildung 2).

Die Projektion legt nahe, dass die wesentlichen finanziellen Herausforderungen durch die demografische Entwicklung für die betrachteten Sozialversicherungszweige im Zeitraum zwischen den Jahren 2030 und 2040 liegen werden.

Insbesondere für die SPV ist wichtig, wie sich die Anzahl der über 80-Jährigen entwickelt. Denn ab einem Alter von 80 Jahren steigt die Wahrscheinlichkeit stark, pflegebedürftig zu werden. So hat sich die Anzahl der über 80-jährigen Personen bereits im Jahr 2020 im Vergleich zum Jahr 2010 deutlich um 40 % auf 5,9 Millionen Personen erhöht. Im Vergleich zum Jahr 2010 nimmt die Anzahl der über 80-jährigen Personen bis zum Jahr 2040 voraussichtlich um 83 % und bis zum Jahr 2050 um 131 % auf 9,7 Millionen Personen zu. Bis zum Jahr 2060 sinkt sie wieder etwas auf voraussichtlich 8,8 Millionen Personen.

2.2 Annahmen zur Arbeitsmarktentwicklung

Für die finanzielle Situation der Sozialversicherungen sind die Erwerbsbeteiligung und die Beschäftigung der Bevölkerung bedeutend. Die Anzahl der Erwerbspersonen wird im Simulationsmodell anhand von nach Geschlecht und Alter differenzierten Erwerbsquoten aus dem Mikrozensus bestimmt, die auf die Bevölkerung im Alter von 15 bis 74 Jahren angewendet werden. Diese Quoten wurden mit Hilfe eines von der OECD entwickelten Ansatzes so fortgeschrieben, dass laufende Trends aus der Vergangenheit weiter anhalten. Die Anzahl der Erwerbspersonen sinkt demnach von rund 46,5 Millionen im Jahr 2019 auf 39,8 Millionen Personen im Jahr 2060. Der für die Zukunft absehbare Rückgang der Anzahl der Erwerbspersonen fällt damit – trotz einer angenommenen Steigerung der Erwerbsbeteiligung – deutlich ausgeprägter aus als der Rückgang der Gesamtbevölkerung. Zurückzuführen ist dies auf die Änderung der Altersstruktur, ausgedrückt in dem deutlichen Anstieg des Altenquotienten.

Die zukünftige Entwicklung der Erwerbstätigkeit bzw. Erwerbslosigkeit der Erwerbspersonen wird in dem Modell nicht durch explizite Annahmen gesteuert. Die Erwerbslosenquote reagiert vielmehr auf variierende Belastungen des Faktors Arbeit mit Sozialbeiträgen und Steuern. Dadurch wächst sie von 3,9 % im Jahr 2020 auf 4,8 % im Jahr 2060. Die Anzahl der Erwerbstätigen sinkt von 45,3 Millionen im Jahr 2019 auf 38,0 Millionen Personen im Jahr 2060.

Während diese Betrachtung für die gesamtwirtschaftliche Entwicklung von Bedeutung ist, sind für die Finanzierung der Sozialversicherungszweige sowohl die Zahl der sozialversicherungspflichtig Beschäftigten, als auch die Zahl der registrierten Arbeitslosen maßgeblich. Letztere wird aus der Erwerbslosenzahl auf Basis des langjährigen Durchschnittes der Relation der Erwerbs- und Arbeitslosenquote berechnet. Die Arbeitslosenquote steigt in dem Szenario von 5,0 % im Jahr 2020 auf 7,8 % im Jahr 2060.

Bei den sozialversicherungspflichtig Beschäftigten wird davon ausgegangen, dass deren Anteil an den Erwerbstätigen auf der Basis eines mehrjährigen Durchschnittswertes im Zeitablauf konstant bleibt. Vor dem Hintergrund der demografischen Entwicklung geht die Anzahl

der sozialversicherungspflichtig Beschäftigten von rund 33,4 Millionen Personen im Jahr 2019 auf rund 27,7 Millionen Personen im Jahr 2060 zurück.

2.3 Annahmen zur gesamtwirtschaftlichen Entwicklung

Im Projektionsmodell ergibt sich das BIP aus der Kombination der Produktionsfaktoren Arbeit, verfügbarer Kapitalstock und der sogenannten totalen Faktorproduktivität, die den Stand des technischen Fortschrittes misst. Die Wachstumsrate des BIP berechnet sich dann aus den Wachstumsraten dieser drei Faktoren.

Die Wachstumsrate des Faktors Arbeit bestimmt sich durch die oben dargestellte Projektion der Anzahl der Erwerbstätigen. Der Kapitalstock wird auf der Basis einer konstanten (Brutto) Investitionsquote in Höhe des mehrjährigen Durchschnittswertes von 21,2 % fortgeschrieben. Für das künftige Wachstum der Faktorproduktivität wird der langjährige Durchschnittswert von 0,7 % pro Jahr angesetzt.

Aus den Berechnungen ergibt sich, dass Arbeitsproduktivität und Löhne – nach dem Ende der derzeitigen starken, krisenbedingten Schwankungen bis zum Jahr 2022/2023 – im gesamten Simulationszeitraum in einem engen Band mit realen Wachstumsraten zwischen 1,3 % und 1,5 % pro Jahr zunehmen werden. Wegen des anhaltenden Rückganges der Anzahl der Erwerbstätigen resultieren daraus zukünftige reale Wachstumsraten des BIP, die im Durchschnitt knapp unter 1,0 % pro Jahr liegen.

Zur Berechnung nominaler Werte wird in den Simulationen unterstellt, dass die Inflationsrate bis zum Jahr 2060 konstant 1,5 % pro Jahr beträgt. Bedeutsam ist diese Annahme für die Berechnung von Kennziffern für die öffentlichen Finanzen, weil Zinszahlungen auf die jeweilige Staatsschuld auf Basis von Nominalwerten bestimmt werden müssen. Wichtig ist für diese Berechnungen auch der dafür fällige Zinssatz, der in den letzten Jahren deutlich gesunken ist und über dessen zukünftige Entwicklung große Unsicherheit herrscht. Vereinfachend wird unterstellt, dass der durchschnittliche Nominalzins auf alle umlaufenden Staatsschuldtitel des Bundes von zuletzt - 0,3 % pro Jahr im Jahr 2019 langsam wieder linear auf 3,5 % pro Jahr im Jahr 2060 steigt. Der Realzins erreicht dann langfristig einen Satz von 2,0 % pro Jahr. Er bleibt damit voraussichtlich dauerhaft unter seinem langjährigen Durchschnitt aus der Zeit vor der Finanz- und Wirtschaftskrise von 3,0 % pro Jahr.

3 Gesetzliche Rentenversicherung

Die GRV besteht aus der allgemeinen Rentenversicherung und der knappschaftlichen Rentenversicherung. Sie ist im Umlagesystem organisiert. Dies bedeutet, dass die Einnahmen

eines Jahres im gleichen Jahr zur Finanzierung der Leistungen verwendet werden. Es wird – bis auf die sogenannte Nachhaltigkeitsrücklage[31] – kein Kapitalstock gebildet.

3.1 Einnahmen

3.1.1 Beitragseinnahmen

Die wichtigste Einnahmequelle der GRV sind die Beitragseinnahmen. Beitragszahlende sind alle versicherungspflichtig Beschäftigten. Zudem sind einige Selbstständige wie bestimmte Handwerker, Privatlehrer, Hebammen und arbeitnehmerähnliche Selbstständige versicherungspflichtig.

Die **Höhe des Beitrages** ergibt sich, indem man den Beitragssatz mit der Beitragsgrundlage multipliziert. Beitragsgrundlage bei den Beschäftigten ist das Bruttoarbeitsentgelt, maximal bis zur sogenannten Beitragsbemessungsgrenze.[32] Arbeitgeber und Arbeitnehmer zahlen den Beitrag je zur Hälfte. Die Beitragsgrundlage bei den Selbstständigen ist grundsätzlich entweder ihr tatsächliches Arbeitseinkommen (i. d. R. der steuerliche Gewinn) oder die sogenannte Bezugsgröße[33].

Beiträge werden auch für Arbeitslose (durch die Bundesagentur für Arbeit auf Grundlage von 80 % des vormals verdienten Bruttoentgeltes), für Pflegepersonen (durch die Pflegeversicherungen) und für Kindererziehungszeiten (vom Bund) gezahlt.

Die Beitragseinnahmen der allgemeinen Rentenversicherung machten 252,2 Mrd. Euro im Jahr 2020 aus. Darin enthalten waren die Beiträge des Bundes für Kindererziehungszeiten von rund 16,2 Mrd. Euro. Die Beitragszahlungen an die knappschaftliche Rentenversicherung beliefen sich im Jahr 2020 auf 0,5 Mrd. Euro, sodass die Beitragseinnahmen der GRV insgesamt bei 252,7 Mrd. Euro lagen.

3.1.2 Bundeszuschüsse

Neben den Beitragseinnahmen sind die Zahlungen des Bundes die Hauptfinanzierungsquelle der GRV. Die allgemeine Rentenversicherung erhält drei Bundeszuschüsse:

---

[31] Die Nachhaltigkeitsrücklage soll Einnahme- und Ausgabeschwankungen ausgleichen und damit häufige Beitragssatzschwankungen verhindern sowie die Liquidität der Rentenversicherung sichern. Aus ihr werden Defizite gedeckt. Überschüsse werden der Nachhaltigkeitsrücklage zugeführt. Sie belief sich Ende des Jahres 2019 auf rund 40,5 Mrd. Euro und machte damit 1,79 Monatsausgaben aus.

[32] Die Beitragsbemessungsgrenze beträgt im Jahr 2020 in den alten Bundesländern 6 900 Euro und im Beitrittsgebiet 6 450 Euro.

[33] Die Bezugsgröße nach § 18 Absatz 1 SGB IV beträgt im Jahr 2020 in Westdeutschland 3 185 Euro monatlich bzw. 38 220 Euro jährlich und im Beitrittsgebiet 3 010 Euro bzw. 36 120 Euro. In den ersten drei Kalenderjahren nach Geschäftsaufnahme kann als Beitragsgrundlage auch die halbe Bezugsgröße gewählt werden.

- den allgemeinen Bundeszuschuss,
- den zusätzlichen Bundeszuschuss und
- den Erhöhungsbetrag zum zusätzlichen Bundeszuschuss.

Der allgemeine Bundeszuschuss belief sich auf rund 48,2 Mrd. Euro im Jahr 2020. Er wird jährlich gemäß der Lohnentwicklung und der Beitragssatzentwicklung[34] fortgeschrieben. Der zusätzliche Bundeszuschuss wurde im Jahr 1998 eingeführt. Er finanzierte sich durch eine Erhöhung des allgemeinen Umsatzsteuersatzes von 15 % auf 16 %. Deshalb wird er gemäß der Entwicklung des Umsatzsteueraufkommens fortgeschrieben. Der Erhöhungsbetrag zum zusätzlichen Bundeszuschuss wurde im Zusammenhang mit der ökologischen Steuerreform eingeführt. Er wird jährlich gemäß der Entwicklung der Summe der Bruttolöhne und -gehälter angepasst. Der zusätzliche Bundeszuschuss belief sich im Jahr 2020 inklusive Erhöhungsbetrag auf rund 27,1 Mrd. Euro. Zudem deckt der Bund das Defizit der knappschaftlichen Rentenversicherung in Form einer sogenannten Bundesbeteiligung. Diese belief sich auf 5,2 Mrd. Euro im Jahr 2020.

3.2 Ausgaben

3.2.1 Rentenausgaben

Die GRV zahlt

- Altersrenten,
- Erwerbsminderungsrenten,
- Hinterbliebenenrenten

und gewährt Leistungen zur Teilhabe, um die Erwerbsfähigkeit wiederherzustellen („Reha-Leistungen“).

Die Gesamtausgaben der GRV im Jahr 2020 betrugen 338,3 Mrd. Euro. Auf die Rentenausgaben entfielen 303,7 Mrd. Euro. Die Ausgaben für Leistungen zur Teilhabe beliefen sich im Jahr 2020 auf 6,9 Mrd. Euro. Zusätzlich zu den Renten zahlt die GRV den hälftigen Beitrag zur Krankenversicherung der Rentner (KVdR). Die Ausgaben hierfür beliefen sich im Jahr 2020 auf 23,1 Mrd. Euro.

[34] Maßgeblich ist die Entwicklung des sog. fiktiven Beitragssatzes. Dies ist derjenige Beitragssatz, der erforderlich wäre, wenn es den zusätzlichen Bundeszuschuss inklusive Erhöhungsbetrag nicht gäbe (§ 213 Absatz 2 SGB VI).

3.2.2 Rentenformel und Rentenanpassungsformel

Mit ihren Beiträgen erwerben die Versicherten Rentenansprüche. Diese werden in Entgeltpunkten ausgedrückt. Die Anzahl der in einem Jahr durch Beiträge erworbenen **Entgeltpunkte** ergibt sich aus dem Verhältnis des individuellen beitragspflichtigen Jahresentgeltes zum Durchschnittsentgelt der Rentenversicherung.[35] Damit besteht ein Zusammenhang zwischen dem beitragspflichtigen Einkommen bzw. den geleisteten Beiträgen und den Rentenansprüchen, der als **Äquivalenzprinzip** bezeichnet wird.[36]

Die Rentenhöhe bemisst sich nach der sogenannten **Rentenformel.** Durch sie werden die Entgeltpunkte mit einem Wert in Euro (sog. aktueller Rentenwert) versehen und etwaige Rentenabschläge (z. B. wegen eines Renteneintrittes vor der Regelaltersgrenze[37]) berücksichtigt.

Bei den jährlichen Rentenanpassungen zum 1. Juli eines jeden Jahres wird der aktuelle Rentenwert mit dem Rentenanpassungsfaktor fortgeschrieben, sodass sich entsprechend die Monatsrente der Rentnerinnen und Rentner erhöht. Die sogenannte **Rentenanpassungsformel** lautet vereinfacht:

Rentenanpassungsfaktor = Lohnfaktor x Beitragssatzfaktor x Nachhaltigkeitsfaktor

Dabei beschreibt der *Lohnfaktor* grundsätzlich die Entwicklung der Bruttoentgelte der Beitragszahlenden.[38] Der *Beitragssatzfaktor* berücksichtigt Beitragssatzänderungen.[39] Er ist kleiner als eins, wenn der Beitragssatz im Vorjahr gestiegen ist. Durch den Nachhaltigkeitsfaktor beeinflusst die demografische Entwicklung die Rentenhöhe: Ist die Relation der Anzahl der Rentenbeziehenden zur Anzahl der Beitragszahlenden im Vorjahr gestiegen, ist der *Nachhaltigkeitsfaktor* kleiner als eins, die Rentenanpassung fällt also geringer aus.

Damit entwickeln sich die Renten grundsätzlich mit den Löhnen der Beschäftigten. Die Rentensteigerungen bleiben aber immer dann hinter den Lohnsteigerungen zurück, wenn der

---

[35] Einen Entgeltpunkt erhält man, wenn man in einem Jahr genau das Durchschnittsentgelt der Rentenversicherung verdient und darauf Beiträge entrichtet. Hat man ein geringeres oder höheres Einkommen, erhält man proportional weniger oder mehr Entgeltpunkte.

[36] In der gesetzlichen Rentenversicherung gilt eine besondere Form des Äquivalenzprinzips, die sogenannte Teilhabeäquivalenz. Danach entspricht die relative Einkommensposition in der Rentenphase der relativen Einkommensposition in der Beitragsphase.

[37] Jeder Monat des Rentenbezugs vor der Regelaltersgrenze wird grundsätzlich mit einem Abschlag von 0,3 % versehen.

[38] Maßgeblich ist grundsätzlich die Entwicklung der Bruttolöhne und -gehälter gemäß den volkswirtschaftlichen Gesamtrechnungen. Diese wird durch einen Faktor ergänzt, der die Entwicklung der beitragspflichtigen Entgelte je Beschäftigten widerspiegelt.

[39] Bis zur Rentenanpassung 2013 war in dem Beitragssatzfaktor auch die sogenannte „Riester-Treppe“ relevant. Mit ihr wurden in pauschalierter Form die allmählich ansteigenden Aufwendungen der Beschäftigten für die private Altersvorsorge von 0,5 % des Bruttolohns auf 4 % (Altersvorsorgeanteil) abgebildet. Dieser höhere Altersvorsorgeaufwand der Beschäftigten sollte sich bei den Rentnerinnen und Rentnern in niedrigeren Rentenzuwachsraten niederschlagen.

Beitragssatz zur gesetzlichen Rentenversicherung im Vorjahr gestiegen ist und/oder sich das Verhältnis von Rentenbeziehenden zu Beitragszahlenden erhöht hat.

3.2.3 Rentenniveau

Das maßgebliche Rentenniveau im Sozialgesetzbuch ist das sogenannte **Sicherungsniveau vor Steuern**.[40] Es ist definiert als die Relation der verfügbaren Standardrente zum verfügbaren Durchschnittsentgelt. Die verfügbare Standardrente ist dabei die Standardrente (Rente nach 45 Beitragsjahren gemäß dem Durchschnittsentgelt) abzüglich der darauf entfallenden Sozialabgaben (Kranken- und Pflegeversicherung). Vom Durchschnittsentgelt sind ebenfalls die durchschnittlichen Sozialabgaben (Kranken-, Pflege-, Renten- und Arbeitslosenversicherung) und Altersvorsorge abzuziehen.

Bleiben die Rentenerhöhungen wegen der Wirkungen des Nachhaltigkeitsfaktors oder des Beitragssatzfaktors in der Rentenanpassungsformel hinter den Lohnsteigerungen zurück, sinkt das Rentenniveau.[41]

Das Rentenniveau belief sich im Jahr 2019 auf 48,2 % des verfügbaren Durchschnittsentgeltes. Bis zum Jahr 2025 gilt für das Rentenniveau eine Untergrenze von 48,0 % (sog. **Rentenniveau-Haltelinie**): Würde nach der Rentenanpassung gemäß Rentenanpassungsformel ein Rentenniveau unter 48,0 % erreicht, wird die Rentenanpassungsrate abweichend von der Rentenanpassungsformel so festgelegt, dass ein Rentenniveau von 48,0 % realisiert wird. Die Rentenanpassung fällt entsprechend höher aus, als durch die Rentenanpassungsformel vorgegeben.

3.3 Beitragssatz

Der Beitragssatz wird jährlich für das kommende Jahr grundsätzlich so festgelegt, dass die voraussichtlichen Einnahmen die voraussichtlichen Ausgaben unter Einbeziehung der Nachhaltigkeitsrücklage decken. Er wird angepasst, wenn die Nachhaltigkeitsrücklage ansonsten die Obergrenze von 1,5 Monatsausgaben überschreitet (Beitragssatzsenkung) oder die Untergrenze von 0,2 Monatsausgaben unterschreiten würde (Beitragssatzerhöhung).[42]

Der Beitragssatz beträgt derzeit 18,6 %. Für den Beitragssatz besteht bis zum Jahr 2025 eine Beitragssatzobergrenze von 20,0 % (sog. **Beitragssatz-Haltelinie**).[43] Diese wird durch die sogenannte **Beitragssatzgarantie** des Bundes gewährleistet: Sollte ein Beitragssatz über 20,0 %

[40] § 154 Absatz 3 Nr. 2 SGB VI.
[41] Grundsätzlich beschreibt das Rentenniveau eine Rentengröße in Relation zu einer Lohngröße.
[42] § 158 Absatz 1 SGB VI.
[43] Zudem besteht bis zum Jahr 2025 eine Beitragssatzuntergrenze von 18,6 %.

nötig sein, wird der zusätzliche Bundeszuschuss so erhöht, dass ein Beitragssatz von 20,0 % gehalten werden kann.[44]

3.4 Abhängigkeit der gesetzlichen Rentenversicherung von der demografischen Entwicklung

Wie die GRV von der demografischen Entwicklung abhängt und welche Stellschrauben für Reformmaßnahmen in Frage kommen, lässt sich aus der Arithmetik einer umlagefinanzierten Rentenversicherung ableiten. In einer stilisierten umlagefinanzierten Rentenversicherung entsprechen die Rentenausgaben in jedem Jahr den Beitragseinnahmen. Die Ausgaben ergeben sich aus der Durchschnittsrente multipliziert mit der Anzahl der Rentenbeziehenden. Die Beitragseinnahmen sind das Produkt aus dem Beitragssatz und dem durchschnittlichen beitragspflichtigen Einkommen aller Beitragszahlenden. Somit ergibt sich für den Beitragssatz folgender Zusammenhang:

$$Beitragssatz = \frac{Durchschnittsrente}{Durchschnittseinkommen} \times \frac{Rentnenbeziehende}{Beitragszahlende}$$

oder:

$$Beitragssatz = Rentenniveau \times Rentnerquotient$$

Der grundsätzliche Zusammenhang bleibt bestehen, wenn man einen Zuschuss aus Steuermitteln (Bundeszuschuss) der umlagefinanzierten Rentenversicherung berücksichtigt. Die Gleichung lautet dann:[45]

$$Beitragssatz = Rentenniveau \times Rentnerquotient - \frac{Bundeszuschuss}{Lohnsumme}$$

Anhand der Gleichungen wird der direkte Einfluss der demografischen Entwicklung auf den Beitragssatz deutlich. Durch die sinkenden Geburtenraten wird zum einen die Anzahl der Beitragszahlenden reduziert. Durch die längere Lebenserwartung steigt zum anderen tendenziell die Anzahl der Rentenbeziehenden. Die zu beobachtende „Alterung der Gesell-

[44] Nach dem zum 1. Januar 2019 in Kraft getretenen Gesetz über Leistungsverbesserungen und Stabilisierungen in der gesetzlichen Rentenversicherung (RV-Leistungsverbesserungs- und Stabilisierungsgesetz) ist bis zum Jahr 2025 der Beitragssatz in der allgemeinen Rentenversicherung auf 20,0 % begrenzt (§ 287 Absatz 1 SGB VI). Der Bund sichert die Beitragssatzobergrenze von 20,0 % (Haltelinie) zum einen durch Sonderzahlungen in den Jahren 2022 bis 2025 in Höhe von jeweils 500 Mio. Euro ab (§ 287a SGB VI). Reichen die Sonderzahlungen nicht aus, erhöht er zum anderen den zusätzlichen Bundeszuschuss, sodass ein Beitragssatz von 20,0 % gehalten werden kann.

[45] Die Lohnsumme ist dabei die Summe der beitragspflichtigen Einkommen. Diese ergibt sich, wenn man das durchschnittliche beitragspflichtige Einkommen je Beitragszahlenden mit der Anzahl der Beitragszahlenden multipliziert.

schaft“ erhöht somit den Rentnerquotienten, was tendenziell zu einem höheren Beitragssatz führt. Höhere Bundeszuschüsse würden den Beitragssatz stabilisieren.

Aufgrund dieser allgemein gültigen arithmetischen Zusammenhänge stehen die wichtigsten Kenngrößen, aber auch die Stellschrauben für Reformmaßnahmen in der GRV fest. Es sind:

- der Beitragssatz,
- das Rentenniveau,
- der Rentnerquotient und
- der Bundeszuschuss.

Maßnahmen zur Abfederung der demografischen Lasten bzw. zur finanziellen Stabilisierung der GRV müssen auf mindestens eine dieser Stellschrauben einwirken.

Wie die demografischen Lasten auf Rentenbeziehende, Beitragszahlende und den Bund und damit auf Steuerzahlende verteilt werden, bestimmt wesentlich die Rentenanpassungsformel (siehe Nr. 3.2.2).

So führen Mehrausgaben aufgrund der demografischen Entwicklung (z. B. wegen einer höheren Anzahl von Rentenbeziehenden) tendenziell zu einem höheren Beitragssatz und damit zur Beteiligung der Beitragszahlenden. Ein höherer Beitragssatz führt zu einem höheren Bundeszuschuss, womit der Bund, und damit letztlich die Steuerzahlenden, die Mehrausgaben mitfinanzieren. Beitragssatzerhöhungen haben über die Rentenanpassungsformel (Beitragssatzfaktor) aber auch geringere Rentenanpassungen zur Folge. Der höhere Rentnerquotient aufgrund der gestiegenen Anzahl der Rentenbeziehenden sorgt über die Rentenanpassungsformel (Nachhaltigkeitsfaktor) ebenfalls für geringere Rentensteigerungen. Somit sind über den Beitragssatzfaktor und den Nachhaltigkeitsfaktor auch alle Rentenbeziehenden an der Finanzierung der demografischen Lasten beteiligt. Damit haben die Bundeszuschüsse und die Rentenanpassungsformel eine wichtige „**Lastenverteilungsfunktion**“.

3.5 Entwicklung bis zum Jahr 2060

3.5.1 Annahmen für die Modellrechnungen

Grundlage für die Simulationsrechnung bis zum Jahr 2060 im Basisszenario ist der rentenrechtliche Status quo zu Beginn des Jahres 2021. Insbesondere wird angenommen, dass die doppelte Haltelinie – Rentenniveauuntergrenze von 48 % und Beitragssatzobergrenze von 20 % – nach dem Jahr 2025 nicht mehr gilt. Die Rentenanpassungen folgen somit der Rentenanpassungsformel, und der Beitragssatz kann über 20 % steigen.

Es wird zudem unterstellt, dass das durchschnittliche Renteneintrittsalter auch wegen der schrittweisen Anhebung der Regelaltersgrenze auf 67 Jahre von derzeit 64,1 Jahre auf 65,5 Jahre bis zum Jahr 2037 zunimmt. Danach bleibt es in etwa konstant. Aus den simu-

lierten Renteneintritten aktiver Versicherter lässt sich die Entwicklung der Zahl der Altersrentnerinnen und -rentner herleiten. Die Zahl der Erwerbsminderungsrentnerinnen und -rentner wird unter Berücksichtigung geschlechts- und altersspezifischer Zugangsraten und deren trendmäßiger Veränderungen abgeschätzt. Für Hinterbliebenenrenten wird unterstellt, dass die Anteile verheirateter Personen unter den aktiven Versicherten und Rentenbeziehenden im Zeitablauf konstant bleiben. Effekte der steigenden Erwerbsbeteiligung von Frauen für ihre Ansprüche auf Witwenrenten, die sich aus den geltenden Anrechnungsvorschriften ergeben, werden beachtet.

3.5.2 Ergebnisse der Modellrechnungen

3.5.2.1 Entwicklung des Rentnerquotienten

Der Rentnerquotient (Relation der Anzahl von Rentenbeziehenden zur Anzahl von Beitragszahlenden) wird – trotz der beschlossenen Anhebung des gesetzlichen Renteneintrittsalters auf 67 Jahre – aufgrund der demografischen Entwicklung stark ansteigen. So kommen im Jahr 2019 rund 76 Rentenbeziehende auf 100 Beitragszahlende. Im Jahr 2035 sind es 93 Rentenbeziehende, im Jahr 2050 über 99 und im Jahr 2060 müssen 100 Beitragszahlende für fast 103 Rentenbeziehende aufkommen (vgl. Abbildung 3).

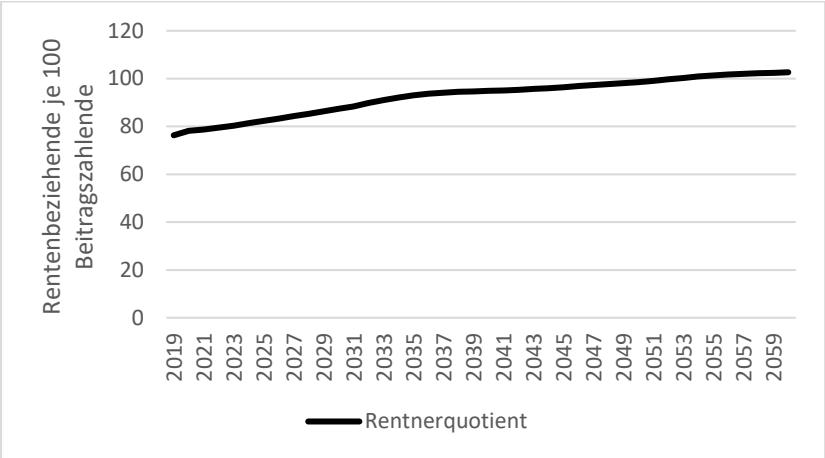
Der für den Nachhaltigkeitsfaktor relevante Äquivalenzrentnerquotient[46] wird vom Jahr 2019 bis zum Jahr 2035 um ein Viertel ansteigen. Finanzieren im Jahr 2019 noch 100 Äquivalenzbeitragszahler knapp 52 Äquivalenzrentner, stehen im Jahr 2035 rechnerisch 65 Äquivalenzrentner 100 Beitragszahlern gegenüber. Im Jahr 2060 sind es 80 Äquivalenzrentner.

[46] Der Äquivalenzrentnerquotient nach § 68 Absatz 4 SGB VI beschreibt die Relation von Äquivalenzrentnern zu Äquivalenzbeitragszahlern. Die Anzahl der Äquivalenzrentner wird ermittelt, indem die gesamten Rentenausgaben durch die Standardrente (Rente mit 45 Entgeltpunkten ohne Abschläge) desselben Kalenderjahres dividiert wird. Die Anzahl der Äquivalenzbeitragszahler wird ermittelt, indem das Gesamtvolumen der Beiträge aller in der allgemeinen Rentenversicherung versicherungspflichtig Beschäftigten, der geringfügig Beschäftigten und der Bezieher von Arbeitslosengeld eines Kalenderjahres durch den auf das Durchschnittsentgelt nach Anlage 1 des SGB VI entfallenden Beitrag (Durchschnittsentgelt mal Beitragssatz) dividiert wird.

Abbildung 3

Rentnerquotient steigt im Basisszenario stark an

Die Relation der Anzahl der Rentenbeziehenden je 100 Beitragszahlende (Rentnerquotient) steigt in den Jahren von 2019 bis 2060 aufgrund der demografischen Entwicklung von 76 auf 103 Rentenbeziehende je 100 Beitragszahler.



Grafik: Bundesrechnungshof.

Quelle für Grundzahlen: Werding, Läpple (2021).

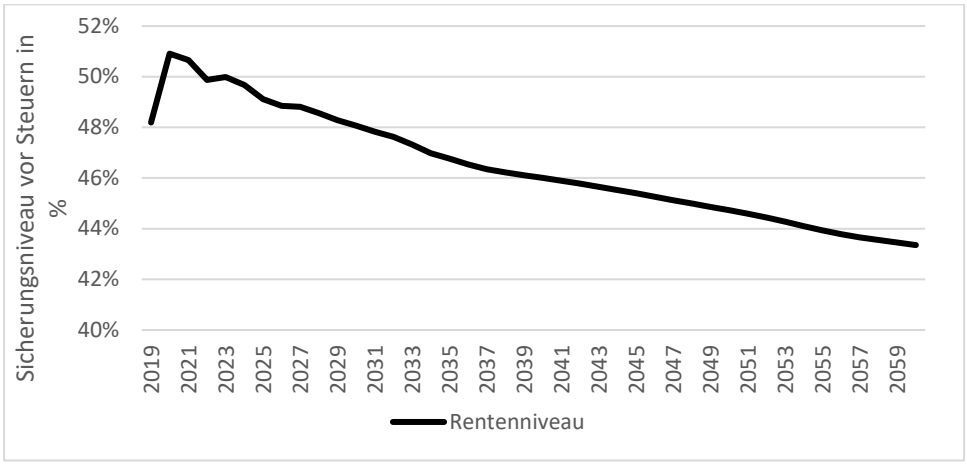
3.5.2.2 Entwicklung des Rentenniveaus

Das Rentenniveau wird zunächst ansteigen. Dies ist darauf zurückzuführen, dass die Löhne wegen der Corona-Pandemie im Jahr 2021 im Durchschnitt gesunken sind, die Renten aufgrund eines Rentensenkungsverbotes (sog. Rentengarantie) jedoch nicht sinken. Ab dem Jahr 2025 werden der Beitragssatzfaktor und der Nachhaltigkeitsfaktor in der Rentenanpassungsformel dafür sorgen, dass die Rentensteigerungen hinter den Lohnsteigerungen zurückbleiben und somit das Rentenniveau sinkt: auf 46,8 % im Jahr 2035, 44,7 % im Jahr 2050 und auf 43,4 % im Jahr 2060 (vgl. Abbildung 4).

Abbildung 4

Das Rentenniveau sinkt

Das Rentenniveau (Sicherungsniveau vor Steuern) sinkt bis zum Jahr 2060 unter 44 %. Die Abbildung zeigt das Rentenniveau in den Jahren 2019 bis 2060 in Prozent.



Grafik: Bundesrechnungshof.

Quelle für Grundzahlen: Werding, Läpple (2021).

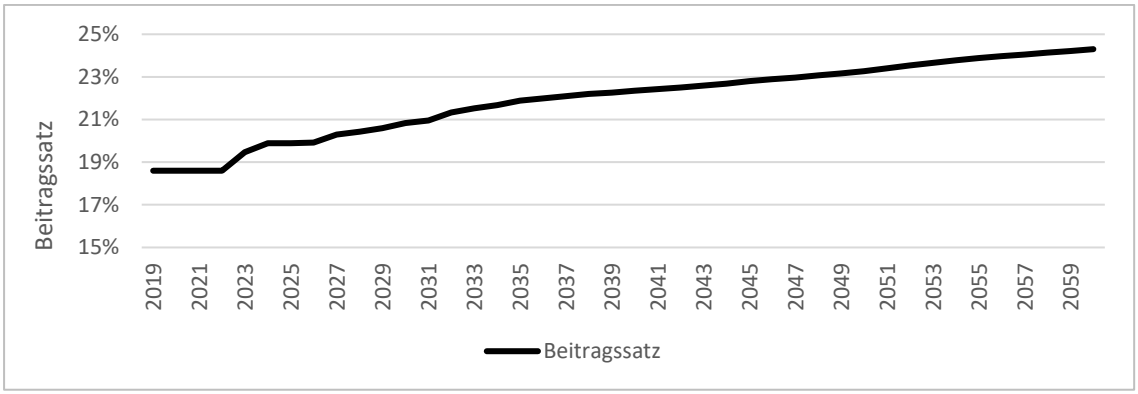
3.5.2.3 Entwicklung des Beitragssatzes

Aufgrund der zunehmenden Anzahl an Rentenbeziehenden, die aufgrund der zunehmenden Lebenserwartung länger Rente beziehen, wird der Beitragssatz – trotz sinkendem Rentenniveau – in Zukunft steigen, im Basisszenario auf 21,9 % im Jahr 2035, auf 23,3 % im Jahr 2050 und auf 24,3 % im Jahr 2060 (vgl. Abbildung 5). Die Beitragssatzobergrenze (Haltelinie) von 20,0 % greift allerdings bis zum Jahr 2025 nicht.

Abbildung 5

Der Beitragssatz steigt

Der Beitragssatz steigt bis zum Jahr 2060 auf über 24 %. Die Abbildung zeigt den Beitragssatz in den Jahren 2019 bis 2060 in Prozent.



Grafik: Bundesrechnungshof.

Quelle für Grundzahlen: Werding, Läpple (2021).

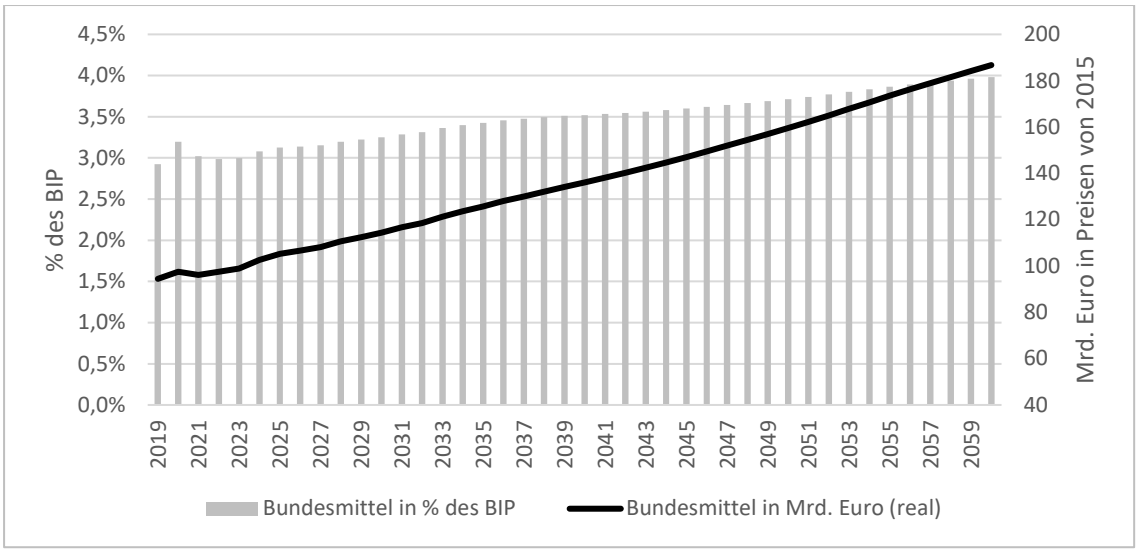
3.5.2.4 Entwicklung der Bundeszuschüsse

Die Mittel, die der Bund an die GRV zahlen muss (v. a. Bundeszuschüsse und Beitragszahlung für Kindererziehungszeiten), steigen kontinuierlich. In Preisen von 2015 erhöhen sie sich von 94,4 Mrd. Euro im Jahr 2019 auf 126 Mrd. Euro im Jahr 2035. Bis ins Jahr 2060 haben sich die Bundesmittel in Preisen von 2015 fast verdoppelt und liegen dann bei 187 Mrd. Euro. Als Prozentsatz des BIP berechnet erhöhen sich vom Jahr 2019 bis zum Jahr 2060 die Bundesmittel von 2,9 % auf 4,0 % des BIP und damit um mehr als ein Drittel (vgl. Abbildung 6).

Abbildung 6

Bundesmittel für die gesetzliche Rentenversicherung steigen stark an

Die Zahlungen des Bundes an die GRV (Bundesmittel) verdoppeln sich in Preisen von 2015 (real) auf 187 Mrd. Euro im Jahr 2060. Relativ zum BIP steigen die Bundesmittel um mehr als ein Drittel, von 2,9 % auf 4 % des BIP im Jahr 2060.



Grafik: Bundesrechnungshof.

Quelle für Grundzahlen: Werding, Läpple (2021).

4 Gesetzliche Krankenversicherung

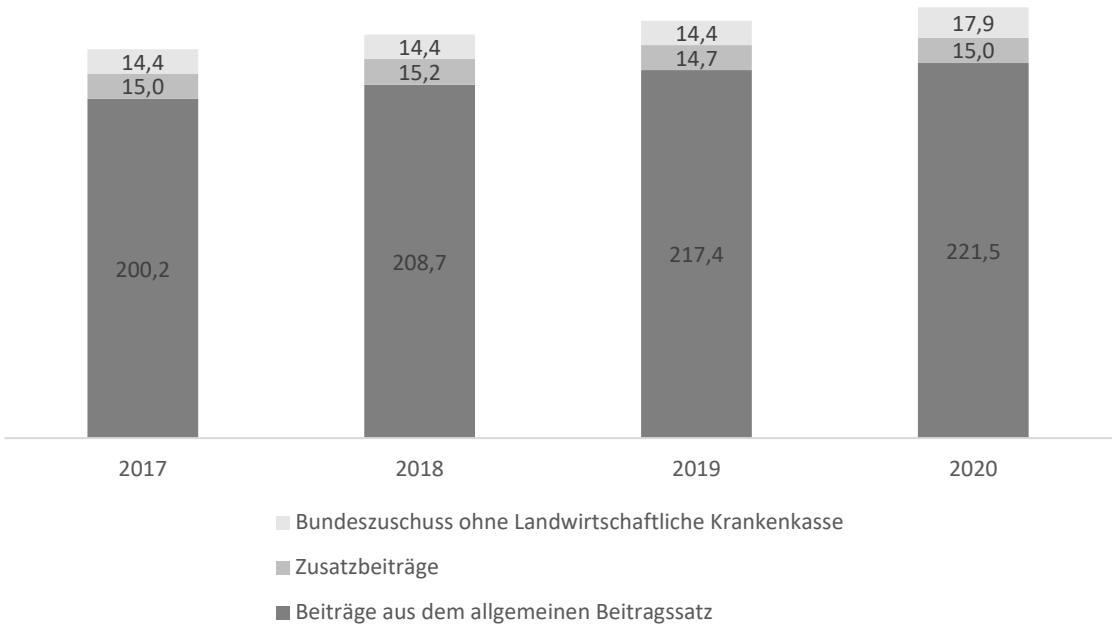
4.1 Einnahmen

Die wichtigsten Einnahmen der GKV sind die Beiträge und der Bundeszuschuss (vgl. Abbildung 7).

Abbildung 7

Die wichtigste Einnahmequelle der GKV sind die Beiträge - der Anteil des Bundeszuschusses ist vergleichsweise gering

Die Haupteinnahmequelle des Gesundheitsfonds[47] sind die Beträge aus dem allgemeinen Beitragssatz und Zusatzbeiträge. Der Bundeszuschuss ist vergleichsweise gering. Die Abbildung zeigt die Einnahmen in Mrd. Euro.



Quelle: https://www.bundesgesundheitsministerium.de/themen/krankenversicherung/zahlen-und-fakten-zur-krankenversicherung/finanzergebnisse.html, Endgültige Rechnungsergebnisse des Gesundheitsfonds 2017-2020, ohne sonstige Einnahmen, Darstellung Bundesrechnungshof.

[47] Seit 1. Januar 2009 fließen alle Beitragseinnahmen der einzelnen Krankenkassen und der Bundeszuschuss in den Gesundheitsfonds. Die gesetzlichen Krankenkassen erhalten aus dem Gesundheitsfonds einen festgelegten Anteil pro Monat. Damit werden ihre durchschnittlichen Leistungs- und Verwaltungsausgaben ausgeglichen. Diese Zuweisungen basieren auf der Versichertenanzahl und werden an das Morbiditätsrisiko der Versicherten der jeweiligen Krankenkasse angepasst (morbiditätsorientierter Risikostrukturausgleich).

4.1.1 Beitragseinnahmen

Der Versichertenkreis der GKV besteht aus den pflichtversicherten und freiwillig versicherten Mitgliedern, die jeweils Beiträge zahlen und aus den beitragsfrei mitversicherten Familienangehörigen.

Die **Höhe der Beiträge** errechnet sich grundsätzlich, indem man die beitragspflichtigen Einnahmen der Mitglieder mit dem Beitragssatz multipliziert. Seit dem 1. Januar 2011 ist der allgemeine Beitragssatz gesetzlich festgelegt.[48] Er liegt zurzeit bei 14,6 % und wird je zur Hälfte von Arbeitgebern einerseits und Arbeitnehmerinnen und Arbeitnehmern andererseits aufgebracht. Krankenkassen müssen außerdem einen einkommensabhängigen Zusatzbeitrag erheben, wenn die Zuweisungen von Finanzmitteln, die sie aus dem Gesundheitsfonds erhalten, nicht zur Deckung ihrer voraussichtlichen Ausgaben ausreichen.

Pflichtversichert sind grundsätzlich alle abhängig Beschäftigten.[49] Beamte und Selbstständige sind nicht pflichtversichert. Beitragspflichtige Einnahme der **Pflichtversicherten** ist das Arbeitsentgelt.

Bei abhängig Beschäftigten endet die Versicherungspflicht, wenn ihre beitragspflichtigen Einnahmen die sogenannte **Jahresarbeitsentgeltgrenze**, auch Versicherungspflichtgrenze genannt, überschreiten.[50] Sie liegt im Jahr 2021 bei 64 350 Euro. Personen mit beitragspflichtigen Einnahmen über dieser Grenze können sich freiwillig in der GKV versichern oder eine private Krankenversicherung (PKV) abschließen.[51] Beitragspflichtige Einnahmen für die **freiwillig Versicherten** sind nicht nur das Arbeitsentgelt. Vielmehr berücksichtigt die Beitragsbelastung die gesamte wirtschaftliche Leistungsfähigkeit des freiwilligen Mitgliedes.[52]

Einnahmen der Mitglieder sind grundsätzlich nur bis zur **Beitragsbemessungsgrenze** (BBG) beitragspflichtig.[53] Damit wird ein Höchstbeitrag eingeführt, der verhindern soll, dass der GKV-Beitrag im Einzelfall eine im Verhältnis zum Leistungsanspruch unvertretbare Höhe erreicht und der Gedanke der solidarischen Mittelaufbringung nach dem individuellen finanziellen Leistungsvermögen überstrapaziert wird.[54] Die BBG belief sich im Jahr 2021 auf 58 050 Euro jährlich.

---

[48] § 241 SGB V.

[49] § 7 Absatz 1 SGB IV.

[50] Die Jahresarbeitsentgeltgrenze wird jährlich nach Maßgabe der Lohnentwicklung angepasst: Sie ändert sich zum 1. Januar eines jeden Jahres in dem Verhältnis, in dem die Bruttolöhne und -gehälter je Arbeitnehmer im vergangenen Kalenderjahr zu den entsprechenden Bruttolöhnen und -gehältern im vorvergangenen Kalenderjahr stehen (§ 6 Absatz 6 SGB V).

[51] Grundsätzlich besteht eine allgemeine Pflicht zur Absicherung für den Krankheitsfall, *Peters*, in: Kasseler Kommentar Sozialversicherungsrecht, Stand: Dezember 2019, § 6 SGB V, Rdn. 10.

[52] Der Begriff der gesamten wirtschaftlichen Leistungsfähigkeit schließt das Vermögen des freiwillig Versicherten regelmäßig nicht ein. Der Beitragspflicht unterliegen lediglich die **Einnahmen**, die aus dem Vermögen erzielt werden.

[53] § 223 Absatz 3 SGB V begrenzt die der Beitragsbemessung zugrunde zu legenden beitragspflichtigen Einnahmen.

[54] Becker/Kingreen, SGB V Gesetzliche Krankenversicherung, 7. Auflage 2020, § 220 Rdn. 9.

Familienangehörige von Mitgliedern sind unter bestimmten Voraussetzungen beitragsfrei mitversichert. Die sogenannte **Familienversicherung** begründet eigene Leistungsrechte des mitversicherten Angehörigen.

Rentenbeziehende, die in der **Krankenversicherung der Rentner (KVdR**) pflichtversichert sind, zahlen in der Regel Krankenkassenbeiträge auf die gesetzliche Rente, Arbeitseinkommen und Versorgungsbezüge. Freiwillig gesetzlich versicherte Rentenbeziehende zahlen zusätzlich auf private Renten, Einkünfte aus Vermietung- und Verpachtung und Kapitalerträge Krankenkassenbeiträge bis zur Höhe der BBG.

Im Jahresdurchschnitt 2020 waren 57,2 Millionen Mitglieder der GKV und 16,1 Millionen Personen beitragsfrei mitversichert. Insgesamt gab es 73,3 Millionen GKV-Versicherte.[55]

4.1.2 Bundeszuschuss

Der Bund leistet einen Zuschuss an den Gesundheitsfonds zur pauschalen Abgeltung der Aufwendungen der Krankenkassen für gesamtgesellschaftliche Aufgaben (Bundeszuschuss). Er gehört zu den sonstigen Einnahmen. Seine Höhe ist gesetzlich festgelegt und beläuft sich auf 14,5 Mrd. Euro jährlich.[56] Pandemiebedingt wurde der Zuschuss in den Jahren 2020 bis 2022 erhöht.

4.2 Ausgaben

4.2.1 Ausgabenentwicklung seit dem Jahr 2010

Die GKV gewährt ihren Versicherten Sachleistungen nach dem Bedarfsprinzip und Krankengeld. Zugrunde liegt ein Leistungspaket, das dem jeweiligen medizinischen Kenntnisstand entspricht. Die Ausgaben lagen im Jahr 2010 bei 176,0 Mrd. Euro und im Jahr 2020 bei 262,9 Mrd. Euro. Der Anteil der GKV-Ausgaben am BIP ist seit dem Jahr 2010 von 6,7 % auf 7,8 % im Jahr 2020 gestiegen.

4.2.2 Ursachen für Ausgabensteigerungen

Die Gesundheitsausgaben für einen Versicherten sind u. a. abhängig vom sozioökonomischen Status, dem Eintritt einer chronischen Krankheit oder dem Auftreten von Multimorbidität.[57] Die Ausgabenentwicklung der GKV wird beeinflusst von

---

[55] https://www.bundesgesundheitsministerium.de/fileadmin/Dateien/3_Downloads/Statistiken/GKV/Mitglieder_Versicherte/KM1_JD_2020_CPS_bf.pdf.

[56] § 221 SGB V.

[57] Bertelsmann Stiftung: Zukünftige Entwicklung der GKV-Finanzierung, 2019, Seite 16.

- der Altersstruktur der Versichertengemeinschaft,
- der Erhöhung der Vergütung der Leistungserbringer (Preiserhöhung),
- dem medizinisch-technischen Fortschritt,
- Leistungsausweitungen bzw. Leistungskürzungen,
- Präferenzverschiebungen und
- der Anzahl der in Anspruch genommenen Versicherungsleistungen.

Nachfolgend werden wesentliche Einflussfaktoren näher erläutert.

4.2.3 Ausgabenanstieg durch die Änderung der Altersstruktur

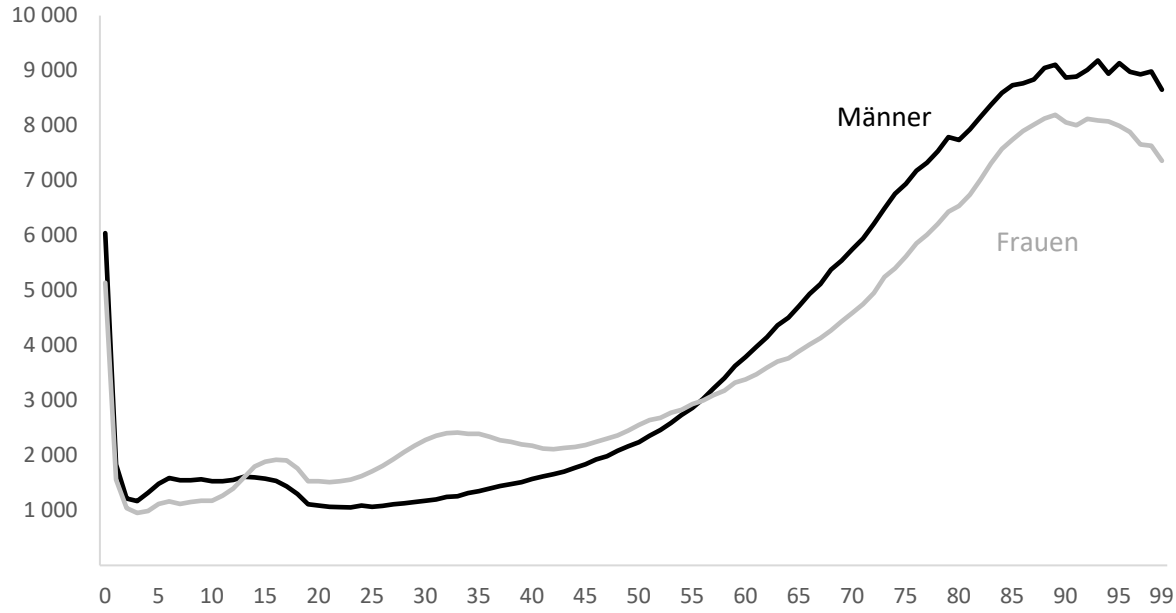
Bedeutend für die Ausgabenentwicklung ist insbesondere die Altersstruktur der Versicherten. Denn die durchschnittlichen Ausgaben je Versicherten steigen mit zunehmendem Alter an. Die altersspezifischen Ausgabenprofile zeigen, welche durchschnittlichen Ausgaben für einen Versicherten in einem bestimmten Alter entstehen (vgl. Abbildung 8). Ein 50-jähriger Mann z. B. verursachte im Jahr 2019 durchschnittliche Ausgaben von 2 242 Euro. Die Ausgaben eines 90-jährigen Mannes betrugen dagegen durchschnittlich 8 875 Euro. Eine 50-jährige Frau verursachte im Jahr 2019 durchschnittliche Ausgaben von 2 553 Euro. Die Ausgaben für 90-jährige Frauen lagen bei durchschnittlich 8 058 Euro.[58] Verändert sich die Altersstruktur derart, dass die Anzahl der Versicherten im höheren Lebensalter zunimmt, steigen die GKV-Ausgaben.

[58] Bundesamt für Soziale Sicherung, Ausgabenprofile 2017.

Abbildung 8

Durchschnittliche alters- und geschlechtsspezifische Ausgaben je Versicherten steigen im Alter an

Die Ausgaben der Versicherten (ohne Krankengeld) liegen zwischen dem zweiten und fünfundfünfzigsten Lebensjahr durchschnittlich unter 3 000 Euro je Versicherten und steigen mit zunehmendem Alter bis auf fast 9 000 Euro an. Die Abbildung zeigt die Ausgaben je Versicherten in Euro in einem bestimmten Alter für Männer und Frauen im Jahr 2019.



Grafik: Bundesrechnungshof.

Quelle: Bundesamt für Soziale Sicherung.

4.2.4 Ausgabenanstieg durch den medizinisch-technischen Fortschritt

Auch durch den medizinisch-technischen Fortschritt werden die Ausgaben der GKV beeinflusst. Der medizinisch-technische Fortschritt bezeichnet neue und verbesserte Diagnose- und Behandlungsmethoden. Durch diese können Krankheiten besser und schneller erkannt und behandelt werden, sodass die Lebensqualität und die Lebenserwartung steigen. Viele Ausprägungen des medizinischen Fortschritts erweitern die bestehenden Behandlungsmöglichkeiten. Nur in wenigen Fällen ersetzen sie alte Behandlungs- und Diagnosemethoden.[59]

Der medizinisch-technische Fortschritt wird häufig als Ausgabentreiber in der GKV gesehen. Es ist jedoch schwierig, den Effekt des medizinisch-technischen Fortschrittes auf die Ausgaben zu beziffern. Häufig werden in Studien die Effekte des medizinisch-technischen Fortschrittes mit anderen ausgabensteigernden Faktoren wie veränderte politische Rahmen-

[59] Gabler, Versicherungslexikon, https://www.versicherungsmagazin.de/lexikon/medizinischer-fortschritt-1945926.html.

bedingungen vermischt (z. B. gesetzliche Änderungen des Leistungsumfanges).[60] Medizinische Innovationen können auch positive gesamtwirtschaftliche Auswirkungen haben und Wachstum und Beschäftigung begünstigen.[61]

4.2.5 Zuzahlungen

Zuzahlungen in der GKV sind eine Form der direkten finanziellen Selbstbeteiligung der Versicherten an den Ausgaben für ihre individuell in Anspruch genommenen Gesundheitsleistungen. Sie unterliegen damit nicht dem Prinzip der solidarischen Finanzierung. Zuzahlungen entlasten die GKV von Ausgaben, da sie den Leistungserbringern in Höhe der Zuzahlungen weniger vergüten müssen. Die Zuzahlungspflicht besteht zum Beispiel bei Krankenhausbehandlung, stationären Vorsorge- und Rehabilitationsleistungen, Heil- und Hilfsmitteln, Fahrkosten, Arznei- und Verbandmitteln, Zahnersatz, Haushaltshilfen und häuslicher Krankenpflege. Grundsätzlich wird durch den Leistungserbringer vom Versicherten eine Zuzahlung von 10 % der Kosten der Leistung erhoben; höchstens allerdings zehn Euro, mindestens fünf Euro oder der tatsächliche Preis (z. B. von Arzneimitteln). Die Zuzahlungen für Klinikaufenthalte liegen bei zehn Euro pro Tag, maximal für 28 Tage im Jahr. Kinder und Jugendliche bis zum vollendeten 18. Lebensjahr sind grundsätzlich von Zuzahlungen befreit. Zuzahlungen sind nur bis zu einer individuellen Belastungsgrenze[62] zu tragen. Ist die Grenze erreicht, wird der Versicherte von weiteren Zuzahlungen befreit.

Die Zuzahlungen beliefen sich im Jahr 2019 auf 4,2 Mrd. Euro bzw. etwa 1,8 % der gesamten Leistungsausgaben der GKV.

Neben diesem Finanzierungseffekt sollen Zuzahlungen einen Lenkungseffekt haben: So sollen sie ein wirtschaftliches Inanspruchnahmeverhalten und mehr Eigenverantwortung der Versicherten fördern. Skeptiker sehen Zuzahlungen jedoch als Finanzierungsinstrument zu Lasten der Kranken. Soweit sie sozial verträglich wären, würde die gewünschte Änderung des Inanspruchnahmeverhaltens ausbleiben. Werden hohe Zuzahlungen eingeführt, könne dies zu unerwünschten sozialen und gesundheitlichen Folgen führen.[63]

---

[60] Bundestagsdrucksache 18/4283, Bericht des Ausschusses für Bildung, Forschung und Technikfolgenabschätzung, Seite 60.

[61] Bundestagsdrucksache 18/4283, a.a.O., Seite 61.

[62] Die Belastungsgrenze für Zuzahlungen liegt bei 2 % der jährlichen Bruttoeinnahmen zum Lebensunterhalt, d. h., wer darüber hinaus Geld für Zuzahlungen ausgeben muss(te), kann sich für den Rest des Jahres von den Zuzahlungen befreien lassen. Für chronisch Kranke liegt die Belastungsgrenze bei 1 % der jährlichen Bruttoeinnahmen. Die persönliche Belastungsgrenze wird jedes Jahr neu ermittelt.

[63] AOK-Bundesverband, online unter https://www.aok-bv.de/lexikon/z/index_00046.html, Zugriff vom 6. Mai 2021.

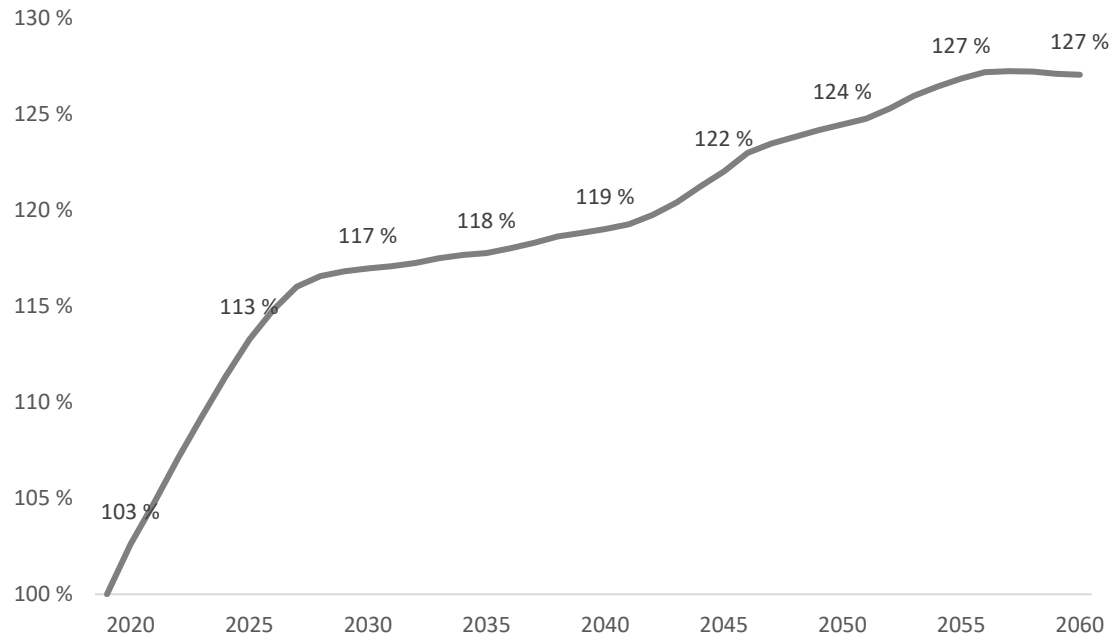
4.3 Abhängigkeit der gesetzlichen Krankenversicherung von der demografischen Entwicklung

Die Abhängigkeit der GKV von der demografischen Entwicklung zeigt sich am Nettoempfängerquotienten. Er spiegelt vergleichbar mit dem Rentnerquotienten in der GRV die Relation der Anzahl der Nettoempfänger zur Anzahl der Nettozahler in der GKV wider. Nettozahler entrichten mehr Beiträge als sie Ausgaben verursachen. Verglichen mit dem Jahr 2019 steigt der Nettoempfängerquotient um 27 % bis zum Jahr 2060 (vgl. Abbildung 9).

Abbildung 9

Der Nettoempfängerquotient steigt stark

Der Nettoempfängerquotient ist die Relation der Anzahl der Nettoempfänger zur Anzahl der Nettozahler. Er steigt um 27 % bis zum Jahr 2060. Die Abbildung zeigt die Entwicklung, wobei der Wert des Jahres 2019 auf 100 % festgelegt wird.



Grafik: Bundesrechnungshof.

Quelle für Grundzahlen: Werding, Läpple (2021).

Dabei ist zu beachten, dass alle Versicherten laufend Leistungen dieses Systems in Anspruch nehmen können und zugleich alle Mitglieder laufend Beiträge entrichten. Mitglieder im Alter von 20 bis 55 Jahren stellen derzeit die Nettozahler der GKV dar, während ältere und jüngere Mitglieder sowie alle mitversicherten Familienangehörigen als Nettoempfänger einzustufen sind.

Zu diesem Effekt der demografischen Entwicklung auf die **Versichertenstruktur** kommt hinzu, dass die demografische Entwicklung die durchschnittlichen Ausgaben je Versicherten und die durchschnittlichen beitragspflichtigen Einkommen der Mitglieder beeinflusst.

Höhere durchschnittliche Ausgaben und gedämpft steigende durchschnittliche beitragspflichtige Einkommen tragen zu einem Beitragssatzanstieg bei.

So werden aufgrund der Alterung mehr Mitglieder Rentenbeziehende sein. Die Renteneinkommen sind im Durchschnitt niedriger als die Lohneinkommen der Mitglieder, sodass die **durchschnittlichen beitragspflichtigen Einkommen** aufgrund der demografischen Entwicklung nicht so stark steigen.

Die **durchschnittlichen Ausgaben** steigen aufgrund der demografischen Entwicklung stärker, weil durch die Alterung der Anteil der älteren Versicherten zunimmt, die höhere Leistungsausgaben verursachen (vgl. Abbildung 8).

Die demografische Entwicklung könnte zudem auch die altersspezifischen Ausgaben selbst beeinflussen, also die Form der altersspezifischen Ausgabenprofile in Abbildung 8. So besagt die Kompressionsthese, dass mit steigender Lebenserwartung die Zahl der gesunden Jahre steigt. Die Menschen altern relativ gesund und verursachen erst als Hochbetagte in ihren letzten Lebensjahren hohe Ausgaben.[64] Bildlich gesprochen würde sich die Kurve der altersspezifischen Ausgaben in Abbildung 8 nach rechts verschieben. Die steigende Lebenserwartung würde folglich nur in geringem Umfang zu Mehrausgaben führen. Hingegen besagt die Medikalisierungsthese, dass mit der Alterung der Bevölkerung die Zahl der chronisch Kranken und Multimorbiden zunimmt. Folglich steigen auch die Ausgaben für die älteren Versicherten relativ stärker als für die jüngeren Versicherten.[65] Bildlich gesprochen kommt es zu einer steileren Kurve der altersspezifischen Ausgaben für die älteren Jahrgänge (Versteilerung). Die Zunahme der Lebenserwartung würde zu einer deutlichen Steigerung der Ausgaben führen.

Empirisch zeigt sich bisher keine eindeutige Tendenz zugunsten einer der beiden Theorien.

4.4 Entwicklung bis zum Jahr 2060

4.4.1 Annahmen für die Modellrechnungen

Projektionen der zukünftigen Ausgaben in der GKV werden maßgeblich durch den Leistungskatalog bestimmt. Der Leistungskatalog kann reduziert oder ausgeweitet werden. Die zukünftigen Ausgaben und damit auch der kostendeckende Beitragssatz unterliegen daher großen Unsicherheiten.

Für die Modellprojektionen sind Annahmen zur Entwicklung der alters- und geschlechtsspezifischen durchschnittlichen Ausgaben erforderlich. Dazu werden zunächst die Ausgaben-

[64] https://www.aok-bv.de/lexikon/k/index_06430.html.
[65] https://www.aok-bv.de/lexikon/m/index_06431.html.

profile des Bundesamtes für Soziale Sicherung (BAS) herangezogen. Diese alters- und geschlechtsspezifischen Pro-Kopf-Ausgaben werden mit den jährlichen Wachstumsraten des Pro-Kopf-BIP[66] zuzüglich eines Aufschlages von 0,46 % fortgeschrieben.[67] Um diesen Prozentsatz lag die allgemeine Steigerungsrate für die gesamten Ausgabenprofile in den letzten 20 Jahren über der Wachstumsrate des Pro-Kopf-BIP. Es wird also für die Projektionen weder eine Versteilerung noch eine Kompression der Ausgabenprofile unterstellt. Die Anzahl der Versicherten und deren Altersstruktur ergibt sich aus der demografischen Entwicklung. In dem untersuchten Basisszenario[68] wird die Anzahl der Mitglieder und mitversicherten Angehörigen fortgeschrieben. Dabei werden Veränderungen des Erwerbsverhaltens berücksichtigt. Die Entwicklung der geschlechts- und altersspezifischen Anteile von Verheirateten bleibt hingegen unverändert. Die Bundeszuschüsse an die GKV werden proportional zu den Leistungsausgaben des Systems fortgeschrieben.

4.4.2 Ergebnisse der Modellrechnungen

Die zukünftigen Ausgaben der GKV sind abhängig von der Entwicklung der Ausgaben für einzelne Sachleistungen und von der Zusammensetzung des Leistungspaketes. Im Jahr 2019 betrugen die Ausgaben der GKV 7,2 % des BIP. Sie steigen im Basisszenario bis zum Jahr 2060 auf 10,9 % des BIP (vgl. Abbildung 10).[69]

[66] Werden die altersspezifischen Pro-Kopf-Ausgaben mit der Zuwachsrate des Pro-Kopf-BIP fortgeschrieben, bleiben die altersspezifischen Ausgabenprofile ausgedrückt als Anteil am Pro-Kopf-BIP konstant. Bei konstanten altersspezifischen Ausgabenprofilen sind Ausgabensteigerungen nur auf eine geänderte Besetzung der Altersklassen, also auf eine Änderung der Altersstruktur der Versicherten zurückzuführen. Diese Ausgabensteigerungen sind somit Folge der demografischen Entwicklung. Methodisch kann so der „demografische Effekt“ auf die Ausgabenentwicklung der GKV isoliert werden.

[67] Der Aufschlag auf die Zuwachsrate des Pro-Kopf-BIP soll grob die Auswirkungen des medizinisch-technischen Fortschritts auf die Pro-Kopf-Ausgaben abbilden.

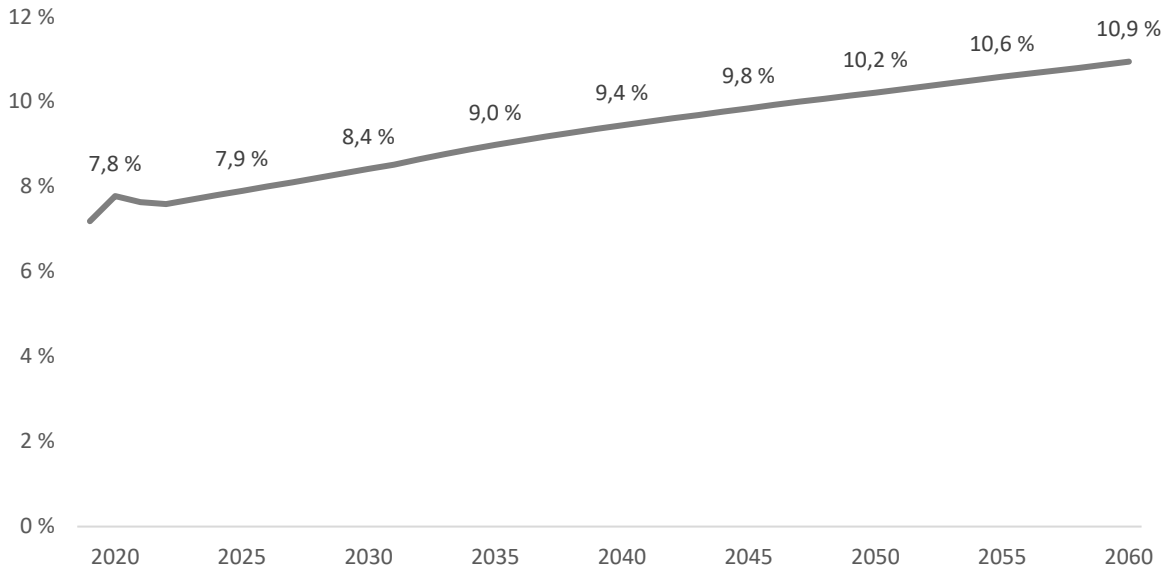
[68] Finanzrisiken für den Bund durch die demographische Entwicklung in der Sozialversicherung, Kurzstudie, Prof. Dr. Werding, Seiten 10, 13 und 14.

[69] Finanzrisiken für den Bund durch die demografische Entwicklung in der Sozialversicherung: Reformszenarien, Prof. Dr. Werding (2021).

Abbildung 10

Die Ausgaben in der gesetzlichen Krankenversicherung gemessen am Bruttoinlandsprodukt steigen deutlich an

Die Ausgaben gemessen am BIP steigen von 7,8 % auf 10,9 % des BIP. Die Abbildung zeigt die Ausgabenquote in % des BIP bis zum Jahr 2060.



Grafik: Bundesrechnungshof.

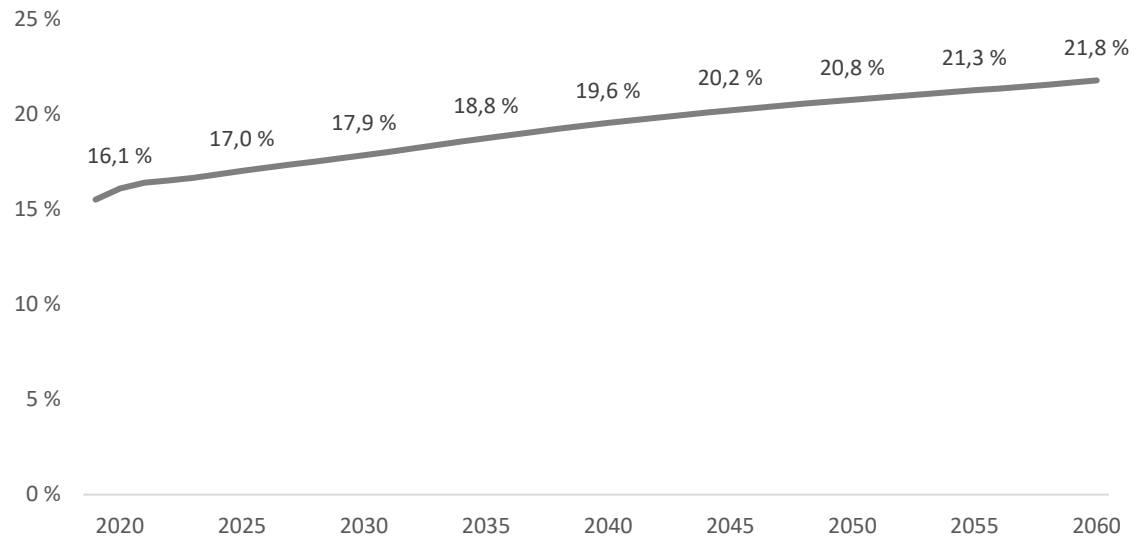
Quelle für Grundzahlen: Werding, Läpple (2021).

Die Änderung der Versichertenstruktur sowie die Entwicklung der durchschnittlichen Pro-Kopf-Ausgaben und der durchschnittlichen Pro-Kopf-Einkommen führen dazu, dass sich Ausgaben und Beitragsgrundlage auseinanderentwickeln. Die Folge ist ein deutlicher Beitragssatzanstieg in der GKV: Im Basisszenario auf 18,8 % im Jahr 2035, auf 20,2 % im Jahr 2045 und auf 21,8 % im Jahr 2060 (vgl. Abbildung 11).

Abbildung 11

Der durchschnittliche Beitragssatz in der gesetzlichen Krankenversicherung erhöht sich deutlich

Der Beitragssatz der GKV steigt von 16,1 % im Jahr 2020 auf 21,8 % im Jahr 2060. Die Abbildung zeigt den durchschnittlichen Beitragssatz in Prozent.



Grafik: Bundesrechnungshof.

Quelle für Grundzahlen: Werding, Läpple (2021).

Bundeszuschüsse an die GKV, die zur Finanzierung versicherungsfremder Leistungen dienen sollen, werden in der Projektion proportional zu den Leistungsausgaben des Systems fortgeschrieben. Während im Jahr 2019 0,4 % des BIP als Bundeszuschuss gezahlt wurden, steigt er bis zum Jahr 2060 auf 0,65 % des BIP an.

5 Soziale Pflegeversicherung

5.1 Einnahmen

Nach dem Prinzip „Pflegeversicherung folgt Krankenversicherung“ entspricht der Versichertenkreis der SPV grundsätzlich dem der GKV. Es gelten die gleichen Regeln zur Berechnung der Beiträge. Die Beitragsbemessungsgrenze und die Versicherungspflichtgrenze sind identisch.

Der hauptsächliche Unterschied zur GKV besteht auf der Einnahmenseite darin, dass die SPV bis zum Jahr 2019 keinen dauerhaften Bundeszuschuss erhielt. Im Gesundheitsversorgungs-

weiterentwicklungsgesetz (GVWG) ist jedoch ab dem Jahr 2022 ein pauschaler Bundeszuschuss für die SPV von jährlich 1 Mrd. Euro vorgesehen.[70]

Die Einnahmen beliefen sich im Jahr 2020 auf 50,6 Mrd. Euro. Davon waren 47,9 Mrd. Euro Beitragseinnahmen.[71]

5.2 Ausgaben

5.2.1 Pflegebedürftigkeit und Anzahl der Pflegebedürftigen

Die Ausgaben der SPV hängen von der Anzahl der Pflegebedürftigen und von der Höhe der Leistungen für diese Pflegebedürftigen ab. Die **Anzahl der Pflegebedürftigen** wird zum einen von der Morbidität der Versichertengemeinschaft bestimmt. Hier ist insbesondere die Altersstruktur entscheidend. Zum andern ist sie abhängig von den Kriterien für die Pflegebedürftigkeit, also der Frage, welche Anspruchsvoraussetzungen erfüllt sein müssen, damit eine versicherte Person Leistungen der Pflegeversicherung in Anspruch nehmen kann.

Mit dem zweiten Pflegestärkungsgesetz führte der Gesetzgeber einen neuen Pflegebedürftigkeitsbegriff[72] und ein anderes Verfahren zur Feststellung und Kategorisierung der Pflegebedürftigkeit ein. Die bis dahin geltenden drei Pflegestufen wurden durch fünf Pflegegrade ersetzt.

Im Jahresdurchschnitt 2020 gab es 4,0 Millionen Pflegebedürftige. Davon wurden 3,2 Millionen Personen ambulant und 0,9 Millionen Personen stationär gepflegt.[73]

5.2.2 Leistungspauschalen und Leistungsausgaben

Die Pflegeversicherung ist als Teilkaskoversicherung organisiert. Das heißt, die SPV erbringt in Abhängigkeit vom Grad der Pflegebedürftigkeit und von der Frage, ob es sich um eine ambulante oder stationäre Pflege handelt, bestimmte Leistungspauschalen. Ausgaben für die pflegebedürftigen Personen, die über die Pauschalen hinausgehen, müssen diese Personen oder ihre Angehörigen selbst tragen. Erst wenn die Einkommens- und Vermögensverhältnisse der Pflegebedürftigen und deren naher Angehöriger nicht ausreichen, um für die

---

[70] Laut § 61a SGB XI dienen diese Bundesmittel zur „pauschalen Beteiligung an den Aufwendungen der sozialen Pflegeversicherung“.

[71] Vgl. BMG: Pflegeversicherung, Zahlen und Fakten, https://www.bundesgesundheitsministerium.de/themen/pflege/pflegeversicherung-zahlen-und-fakten.html#c3291.

[72] Pflegebedürftig sind Personen, die gesundheitlich bedingte Beeinträchtigungen der Selbstständigkeit oder der Fähigkeiten aufweisen und deshalb der Hilfe durch Andere bedürfen. Es muss sich um Personen handeln, die körperliche, kognitive oder psychische Beeinträchtigungen oder gesundheitlich bedingte Belastungen oder Anforderungen nicht selbstständig kompensieren oder bewältigen können.

[73] Vgl. BMG: Pflegeversicherung, Zahlen und Fakten, a. a. O.

Pflegekosten aufzukommen, springt die zuständige Kommune im Rahmen der Sozialhilfe mit Leistungen zur Pflege ein.

Leistungen der Pflegeversicherung werden grundsätzlich nur an Personen mit den Pflegegraden 2 bis 5 erbracht. Pflegebedürftige mit Pflegegrad 1 können jedoch z. B. Beratungsleistungen oder Zuschüsse für Maßnahmen zur Verbesserung des allgemeinen Wohnumfeldes erhalten.

Die Leistungen in der ambulanten Pflege sind die Pflegesachleistungen und das Pflegegeld. In der stationären Pflege zahlt die Pflegekasse Pauschalbeträge für die pflegebedingten Aufwendungen einschließlich Aufwendungen für Betreuung und Aufwendungen für die medizinische Behandlungspflege. Die Pauschalen für diese Leistungen werden nach den Pflegegraden gestaffelt.

Die Höhe der Leistungspauschalen soll alle drei Jahre überprüft werden. Als Orientierungsgröße für die Fortschreibung dient die Inflationsrate der letzten drei Jahre. Dabei ist jedoch sicherzustellen, dass die Fortschreibungsrate nicht höher ist als die Bruttolohnentwicklung.[74]

Die Ausgaben der SPV beliefen sich im Jahr 2020 auf insgesamt 49,1 Mrd. Euro. Davon waren 45,6 Mrd. Euro Leistungsausgaben.[75] Die Leistungsausgaben je Pflegebedürftigen lagen im Jahr 2020 bei 12 143 Euro.

5.3 Beitragssatz

Der Beitragssatz zur Pflegeversicherung wird gesetzlich festgelegt und beträgt derzeit 3,05 % der beitragspflichtigen Einnahmen der Mitglieder. Kinderlose Mitglieder im Alter über 23 Jahren mussten bis zum Jahr 2021 einen Zuschlag auf den Beitragssatz von 0,25 % zahlen.[76] Dieser Zusatzbeitragssatz für Kinderlose beträgt 0,35 % im Jahr 2022.

5.4 Pflegevorsorgefonds

Mit dem ersten Pflegestärkungsgesetz wurde in der SPV ein Sondervermögen unter dem Namen „Vorsorgefonds der sozialen Pflegeversicherung“ errichtet.[77] Dieser Pflegevorsorgefonds wird von der Deutschen Bundesbank verwaltet und dient der langfristigen Beitragssatzstabilisierung.

[74] § 30 Absatz 1 SGB XI.
[75] Vgl. BMG: Pflegeversicherung, Zahlen und Fakten, a. a. O.
[76] § 55 SGB XI.
[77] § 131 SGB XI.

Dem Fonds werden ab dem Jahr 2015 bis zum Jahr 2033 jährlich 0,1 % der beitragspflichtigen Einkommenssumme des Vorjahres zugeführt. Ab dem Jahr 2035 kann der Pflegevorsorgefonds zur Dämpfung des Beitragssatzanstieges aufgelöst werden.

Der Fonds hatte Mitte des Jahres 2021 ein Volumen von rund 9 Mrd. Euro.[78]

5.5 Abhängigkeit der sozialen Pflegeversicherung von der demografischen Entwicklung

Anders als für die GRV ist für die SPV die Abhängigkeit von der demografischen Entwicklung nicht so offensichtlich. Dies liegt vor allem daran, dass Rentnerinnen und Rentner auch Beiträge an die SPV zahlen und jüngere Personen Leistungen in Anspruch nehmen können. Jedoch zeigen die altersspezifischen Pflegewahrscheinlichkeiten, dass das Risiko der Pflegebedürftigkeit vor allem im Alter ab 80 Jahren sehr hoch ist, während es für jüngere Menschen kaum vorhanden ist (vgl. Abbildung 12).[79] Daraus folgt, dass die Ausgaben in der SPV ansteigen, wenn die Anzahl der Menschen mit einer hohen Pflegewahrscheinlichkeit zunimmt. Genau dies ist in Deutschland der Fall: Die Anzahl der Personen im Alter von 80 Jahren und älter steigt nach der mittleren Variante der Bevölkerungsvorausberechnung des Statistischen Bundesamtes von 3,5 Millionen Personen im Jahr 2019 kontinuierlich auf 5,6 Millionen Personen im Jahr 2050, bevor sie bis zum Jahr 2060 auf 5,1 Millionen Personen sinkt.

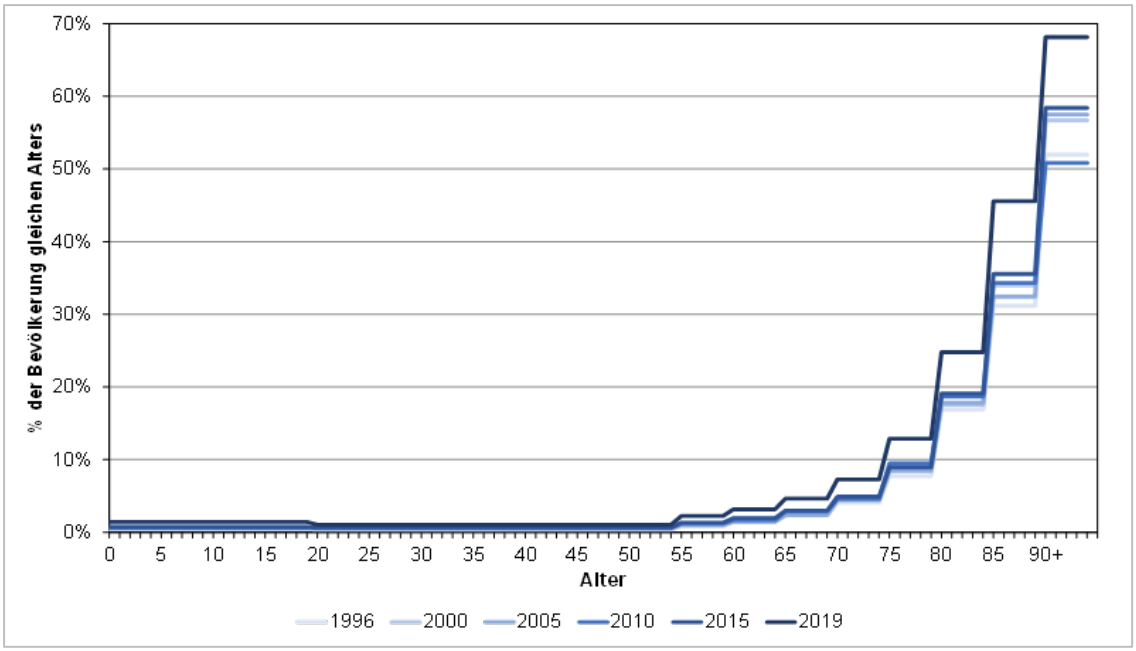
[78] Vgl. Deutsche Bundesbank, Monatsbericht September 2021, S. 10, Fußnote 10.

[79] Der Anstieg der Pflegewahrscheinlichkeiten im Jahr 2019 ist im Wesentlichen auf die Pflegereform des Jahres 2017 (Zweites Pflegestärkungsgesetz) zurückzuführen, in dem die Pflegegrade eingeführt wurden. Diese neue Definition der Pflegebedürftigkeit führte zu einer Ausweitung der Anzahl der Pflegebedürftigen.

Abbildung 12

Die Pflegewahrscheinlichkeit steigt mit dem Alter an

Insbesondere ab dem Alter von 80 Jahren ist die Wahrscheinlichkeit hoch, pflegebedürftig zu werden (Pflegewahrscheinlichkeit). Die Abbildung zeigt die altersspezifischen Pflegewahrscheinlichkeiten im Zeitraum von 1996 bis 2019 in Prozent der Bevölkerung in dieser Altersklasse.



Grafik: Bundesrechnungshof.

Quelle: Werding, Läpple (2021).

Auch für die SPV lassen sich die grundsätzlichen finanziellen Zusammenhänge anhand einer stilisierten umlagefinanzierten Pflegeversicherung darstellen. So entsprechen die Leistungsausgaben in jedem Jahr den Beitragseinnahmen. Die Ausgaben ergeben sich aus den durchschnittlichen Ausgaben je Pflegebedürftigen (Durchschnittsausgaben) multipliziert mit der Anzahl der Pflegebedürftigen (Leistungsbeziehende). Die Beitragseinnahmen sind das Produkt aus dem Beitragssatz und dem durchschnittlichen beitragspflichtigen Einkommen aller Beitragszahler (Durchschnittseinkommen). Somit ergibt sich für den Beitragssatz folgender Zusammenhang:

$$Beitragssatz = \frac{Durchschnittsausgaben}{Durchschnittseinkommen} \times \frac{Leistungsbeziehende}{Beitragszahlende}$$

oder:

$$Beitragssatz = Leistungsniveau \times Leistungsbeziehendenquotient$$

Die demografische Entwicklung beeinflusst den Leistungsbeziehendenquotienten, wenn die Anzahl der Leistungsbeziehenden aufgrund der Alterung steigt und die Anzahl der Beitrags-

zahlenden wegen der gesunkenen Geburtenraten zurückgeht. Die demografische Entwicklung kann jedoch auch das Leistungsniveau beeinflussen:

- wenn der Anteil der Personen unter den Pflegebedürftigen steigt, die besonders hohe Leistungsausgaben verursachen, und
- wenn der Anteil der Personen unter den Beitragszahlenden zunimmt, die ein niedriges beitragspflichtiges Einkommen haben.

Letzteres ist auch dann der Fall, wenn der Anteil der Rentnerinnen und Rentner an den Beitragszahlenden steigt, weil die Rentenbeziehenden im Durchschnitt ein geringeres beitragspflichtiges Einkommen haben als die erwerbstätigen Versicherten.

Aus dem oben genannten Zusammenhang ergeben sich die Kenngrößen und Stellschrauben für die SPV:

- Leistungsbeziehendenquotient,
- Leistungsniveau und
- Beitragssatz.

5.6 Entwicklung bis zum Jahr 2060

5.6.1 Annahmen für die Modellrechnungen

Grundlage für die Simulationsrechnung bis zum Jahr 2060 im Basisszenario ist der rechtliche Status quo im Bereich der SPV zu Beginn des Jahres 2021. Es wird angenommen, dass die **altersspezifischen Pflegewahrscheinlichkeiten** des Jahres 2019 (vgl. Abbildung 11) im Zeitverlauf konstant bleiben.

Für die **Fortschreibung der Leistungsausgaben je Pflegebedürftigen** wird ein Ansatz übernommen, der im jüngsten EU-Ageing Report für die Pflegeausgaben in Deutschland[80] entwickelt wurde. Danach werden die Sachleistungen der SPV zu zwei Dritteln mit der Wachstumsrate der Löhne und zu einem Drittel mit der Inflationsrate fortgeschrieben, während für Geldleistungen die umgekehrten Gewichte gelten. Damit stiegen die durchschnittlichen Leistungsausgaben weniger stark als die Bruttolöhne und -gehälter.[81]

[80] European Commission 2018, Seite 122 f.

[81] Für die angenommene Entwicklung der Leistungsausgaben je Pflegebedürftigen sind damit diejenigen Größen relevant, die auch für die gesetzlich vorgesehene Dynamisierung der Leistungspauschalen maßgeblich sind. So regelt § 30 Absatz 1 SGB XI, dass die Bundesregierung alle drei Jahre Notwendigkeit und Höhe einer Anpassung der Leistungen der Pflegeversicherung prüft. Als ein Orientierungswert für die Anpassungsnotwendigkeit dient die kumulierte Preisentwicklung in den letzten drei abgeschlossenen Kalenderjahren; dabei ist sicherzustellen, dass der Anstieg der Leistungsbeträge nicht höher ausfällt als die Bruttolohnentwicklung im gleichen Zeitraum.

5.6.2 Ergebnisse der Modellrechnungen

5.6.2.1 Entwicklung der Anzahl der Leistungsbeziehenden und des Leistungsempfängerquotienten

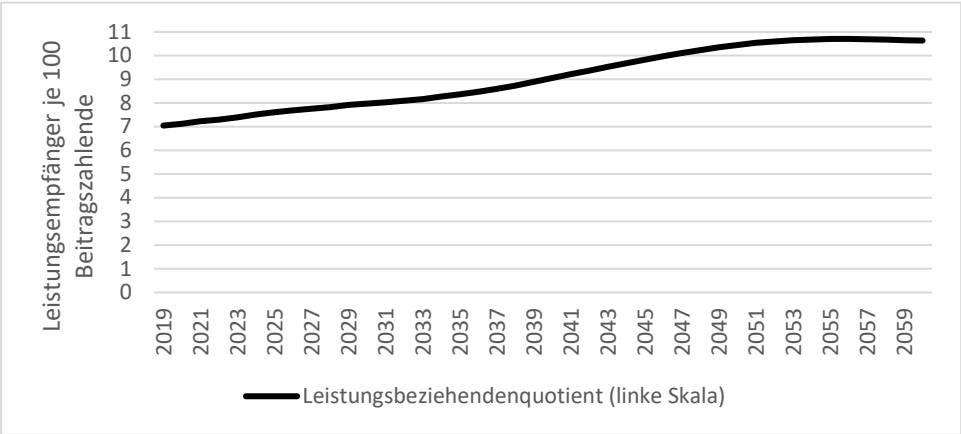
Die **Anzahl der Leistungsbeziehenden** erhöht sich von 4,0 Millionen im Jahr 2019 auf 4,7 Millionen Pflegebedürftige im Jahr 2035. Danach steigt die Zahl weiter an und erreicht mit 5,9 Millionen Personen im Jahr 2054 ihren Höhepunkt. Im Jahr 2060 wird die Zahl der Pflegebedürftigen wieder leicht auf 5,5 Millionen Pflegebedürftige sinken.

Der **Leistungsbeziehendenquotient** lag im Jahr 2019 bei 7,0. Das heißt, auf 100 Beitragszahlende kamen sieben Pflegebedürftige. Diese Zahl steigt kontinuierlich bis auf 10,7 im Jahr 2056 (+52 %) an und bleibt dann in etwa konstant. Im Jahr 2060 kommen somit auf 100 Beitragszahler fast 11 Pflegebedürftige (vgl. Abbildung 13).

Abbildung 13

Der Leistungsbeziehendenquotient steigt

Die Anzahl der Leistungsbeziehenden je 100 Beitragszahlende (Leistungsbeziehendenquotient) steigt wegen der demografischen Entwicklung in den Jahren 2019 bis 2060 von 7 auf fast 11 Leistungsbeziehende je 100 Beitragszahlende an.



Grafik: Bundesrechnungshof.

Quelle für Grundzahlen: Werding, Läpple (2021).

5.6.2.2 Entwicklung des Leistungsniveaus

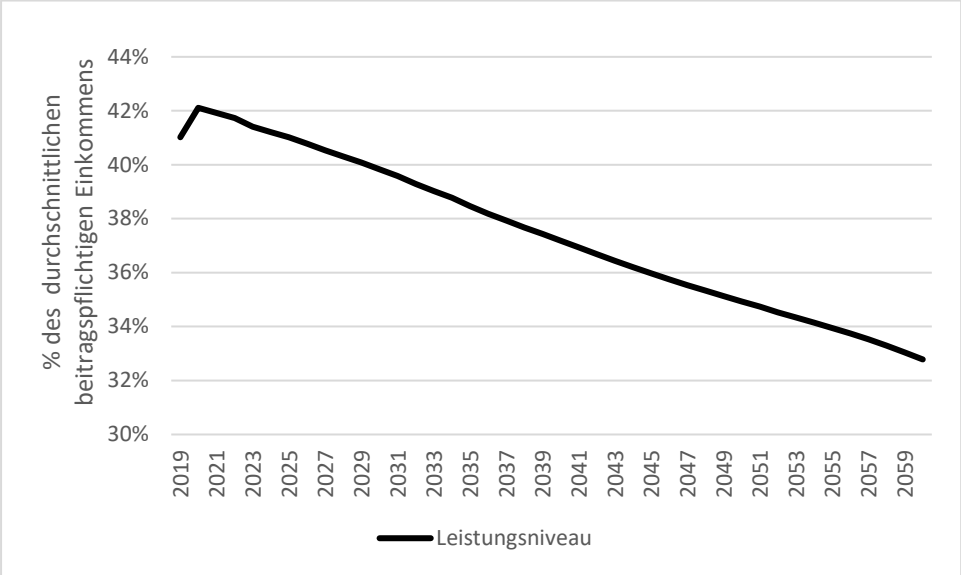
Das oben als Quotient aus den Ausgaben je Leistungsbeziehenden und dem durchschnittlichen beitragspflichtigen Einkommen definierte Leistungsniveau sinkt kontinuierlich von 41,0 % im Jahr 2019 auf 32,8 % im Jahr 2060 (Rückgang um 30 %). Dies bedeutet, dass die Ausgaben je Leistungsempfänger im Jahr 2019 rund 41 % des Durchschnittseinkommens ausmachten. Im Jahr 2060 sind es nur noch 32 % (vgl. Abbildung 14). Zurückzuführen ist dies

darauf, dass die Einkommen (annahmegemäß) stärker steigen als die Leistungspauschalen (vgl. Tnr. 5.6.1).

Abbildung 14

Das Leistungsniveau der sozialen Pflegeversicherung sinkt

Die Relation der Leistungsausgaben je Versicherten zu den durchschnittlichen beitragspflichtigen Einkommen (Leistungsniveau) sinkt.



Grafik: Bundesrechnungshof.

Quelle für Grundzahlen: Werding, Läpple (2021).

5.6.2.3 Entwicklung des Beitragssatzes

Die gegenläufige Entwicklung des Leistungsbeziehendenquotienten und des Leistungsniveaus schlägt sich in der Entwicklung des Beitragssatzes nieder (siehe Gleichung Tnr. 5.5).[82] Zunächst überwiegt der Anstieg des Leistungsbeziehendenquotienten, sodass der Beitragssatz als gewichteter Durchschnitt des Beitragssatzes für Mitglieder mit Kindern und des Beitragssatzes der Mitglieder ohne Kinder ansteigt. Dann heben sich die beiden Bestimmungsgrößen gegenseitig auf, sodass der Beitragssatz in etwa konstant bleibt (vgl. Abbildung 13). Ab dem Jahr 2040 steigt der Leistungsbeziehendenquotient deutlich stärker als das Leistungsniveau sinkt, sodass der Beitragssatz wieder zunimmt. Er erreicht im Jahr 2049 mit 3,7 % seinen höchsten Wert. Danach bleibt er in etwa auf diesem Niveau konstant.

Der Pflegevorsorgefonds erreicht im Jahr 2034 seinen Höchstwert von 34,6 Mrd. Euro. Seine Auflösung in den Jahren 2035 bis 2053 kann den Anstieg des Beitragssatzes nur leicht abbremsen (vgl. Abbildung 15). Der jährliche Beitragssatzeffekt in dieser Auflösungsphase

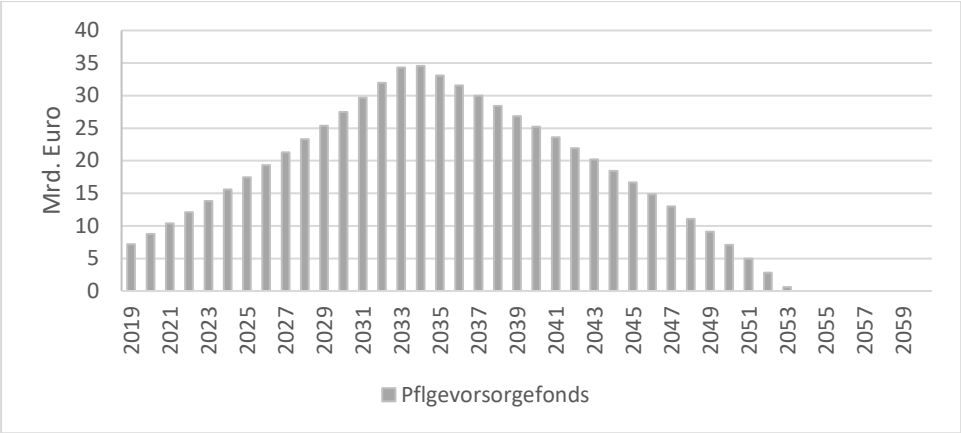
[82] Der Beitragssatz ergibt sich aus dem Produkt aus Leistungsniveau und Leistungsbeziehendenquotienten.

beträgt rund 0,06 Prozentpunkte. Im Jahr 2054 springt der Beitragssatz dann auf das Niveau, das er ohne den Fonds gehabt hätte (vgl. Abbildung 16).

Abbildung 15

Auf- und Abbau des Pflegevorsorgefonds

Der Pflegevorsorgefonds wächst auf fast 35 Mrd. Euro, bevor er ab dem Jahr 2035 abgebaut wird. Die Abbildung zeigt das Volumen des Pflegevorsorgefonds in Mrd. Euro in den Jahren 2019 bis 2060.



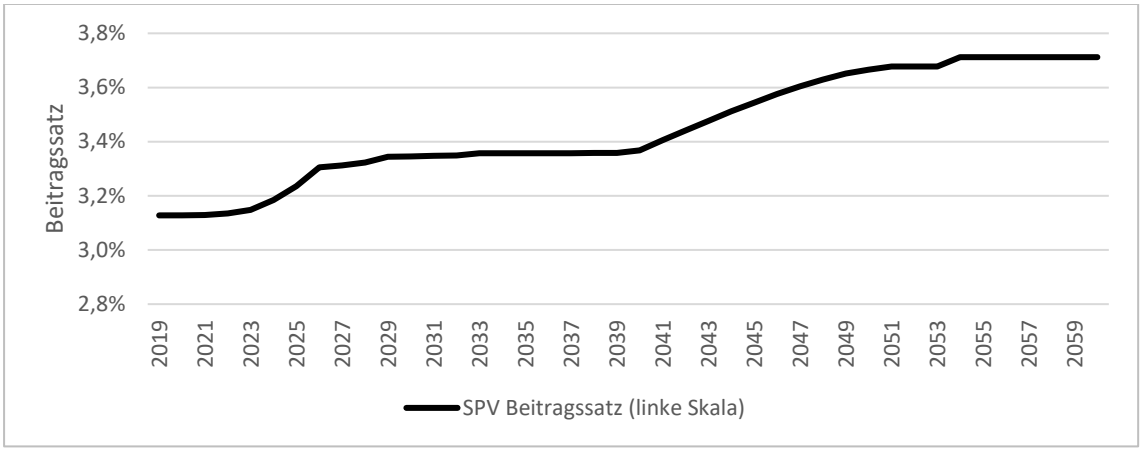
Grafik: Bundesrechnungshof.

Quelle für Grundzahlen: Werding, Läpple (2021).

Abbildung 16

Der Beitragssatz in der sozialen Pflegeversicherung steigt trotz Abbau des Pflegevorsorgefonds

Der Beitragssatz steigt bis auf 3,7 % im Jahr 2060. Die Abbildung zeigt den Beitragssatz in Prozent in den Jahren 2019 bis 2060.



Grafik: Bundesrechnungshof.

Quelle für Grundzahlen: Werding, Läpple (2021).

6 Finanzielle Gesamtsituation der Sozialversicherung und deren Risiken für den Bund

6.1 Projektion für die betrachteten Sozialversicherungszweige

Die demografische Entwicklung beeinflusst die finanzielle Situation aller betrachteten Sozialversicherungszweige. Durch deren unterschiedliche Ausrichtung differiert die Höhe der finanziellen Belastung und die Entwicklung im betrachteten Zeitraum bis zum Jahr 2060. Für das Ausmaß der finanziellen Belastung der Beitragszahlenden und des Bundes ist die Gesamtentwicklung entscheidend. Sie lässt sich insbesondere an der Ausgabenentwicklung für die Sozialversicherungszweige, der Entwicklung der Sozialversicherungsbeitragssätze und anhand möglicher Finanzrisiken für den Bund darstellen.

6.2 Ausgabenentwicklung bis zum Jahr 2060

Die Gesamtausgaben der GRV werden im Basisszenario von 9,4 % des BIP im Jahr 2019 bis zum Jahr 2060 auf 12,3 % des BIP ansteigen. Der reale Anstieg kann anhand der preisbereinigten Gesamtausgaben auf Grundlage der Preise des Jahres 2015 veranschaulicht werden. Danach werden die Gesamtausgaben in der GRV von 304,4 Mrd. Euro im Jahr 2019 auf 574,7 Mrd. Euro im Jahr 2060 steigen. Damit verdoppeln sich die preisbereinigten Ausgaben im Bereich der GRV nahezu.

In der GKV lagen die Ausgaben im Jahr 2019 bei 7,2 % des BIP. Nach der Projektion im Basisszenario werden diese Ausgaben bis zum Jahr 2060 auf 10,9 % des BIP steigen. Damit fällt der Anstieg prozentual stärker aus als in der GRV.

Die Ausgaben der SPV lagen im Jahr 2019 bei 1,3 % des BIP. Sie werden voraussichtlich bis zum Jahr 2060 auf 1,7 % ansteigen. Die Ausgaben für die Arbeitslosenversicherung[83] sowie für die Grundsicherung für Arbeitssuchende lagen im Jahr 2019 zusammen bei 2,2 % des BIP. Sie werden bis zum Jahr 2060 auf 3,2 % des BIP steigen.

Die Gesamtausgaben in den Sozialversicherungen (Summe der Beiträge für die GRV, GKV, SPV und die Arbeitslosenversicherung) steigen in der Projektion preisbereinigt von 560 Mrd. Euro im Jahr 2019 auf 1 108 Mrd. Euro im Jahr 2060 (in Preisen des Jahres 2015).

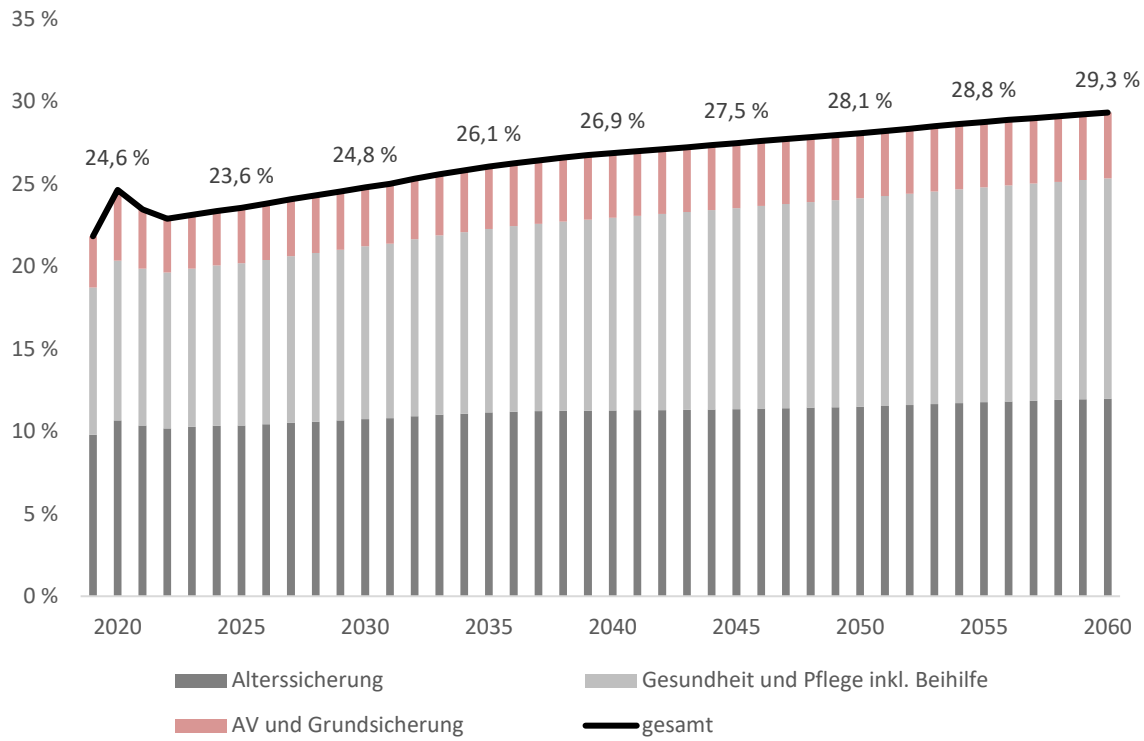
[83] Die Arbeitslosenversicherung ist nicht Gegenstand dieses Berichtes. Die Entwicklung wird lediglich zur Vervollständigung der Gesamtentwicklung in der Sozialversicherung dargestellt.

Die Gesamtausgaben für die Sozialversicherungen[84] beliefen sich im Jahr 2019 auf 21,8 % des BIP. Das entspricht 50 % der gesamtstaatlichen Primärausgaben[85]. Die Ausgabenquote wird nach dem Basisszenario bis zum Jahr 2060 auf 29,3 % des BIP um 7,5 Prozentpunkte ansteigen (vgl. Abbildung 17). Die Ausgaben werden sich damit deutlich erhöhen. Sie steigen zunächst kontinuierlich an. Nach dem Jahr 2035 wird der Anstieg leicht nachlassen.

Abbildung 17

Die projizierten Ausgaben im Zeitraum 2019 bis 2060 steigen deutlich

Die Gesamtausgaben für die Sozialversicherungen beliefen sich im Jahr 2019 auf 21,8 % des BIP. Die Ausgabenquote wird bis zum Jahr 2060 auf 29,3 % des BIP ansteigen. Die Abbildung zeigt die Ausgaben in Prozent des BIP.



Grafik: Bundesrechnungshof.

Quellen: DRV, BMG, BA, Statistisches Bundesamt, Werding, Läpple (2021).

[84] Dazu gehören neben den Ausgaben der Sozialversicherungen auch die projizierten Ausgaben für die Beamtenversorgung und die Beihilfen zu Gesundheits- und Pflegekosten. Diese lagen im Jahr 2019 bei 2,1 % des BIP und werden bis zum Jahr 2060 in diesem Szenario 2,8 % des BIP betragen. Enthalten sind auch die Ausgaben der Arbeitslosenversicherung sowie für die Grundsicherung für Arbeitsuchende. Sie lagen im Jahr 2019 bei 1,0 % bzw. 1,2 %, zusammengenommen bei 2,2 % des BIP. Diese Ausgaben sind nicht Gegenstand dieses Berichtes. Sie werden jedoch berücksichtigt, um die Gesamtentwicklung darzustellen.

[85] Die Primärausgaben errechnen sich, indem von den Ausgaben die Zinsausgaben abgezogen werden.

6.3 Entwicklung der Sozialversicherungsbeiträge

Die umlagefinanzierten Sozialversicherungen sind grundsätzlich so ausgestaltet, dass die steigenden Ausgaben durch steigende Sozialversicherungsbeiträge finanziert werden. Um höhere Lohnnebenkosten durch steigende Sozialversicherungsbeitragssätze zu verhindern, hatte die Regierungskoalition im Jahr 2020 beschlossen[86], im Rahmen einer „Sozialgarantie 2021“ den Gesamtsozialversicherungsbeitragssatz bei maximal 40 % zu stabilisieren. Darüber hinaus gehende Finanzbedarfe sollen bis zum Jahr 2021 aus dem Bundeshaushalt gedeckt werden.[87]

In den letzten Jahren lag der Gesamtsozialversicherungsbeitragssatz im Wesentlichen bei knapp 40 %. Auf der Basis der projizierten Ausgaben müsste der Gesamtsozialversicherungsbeitragssatz bis zum Jahr 2060 kontinuierlich auf 53,3 % steigen (vgl. Abbildung 18). Der Anstieg um gut 13 Prozentpunkte vollzieht sich kontinuierlich, mit einer allmählichen Verlangsamung nach dem Jahr 2035.

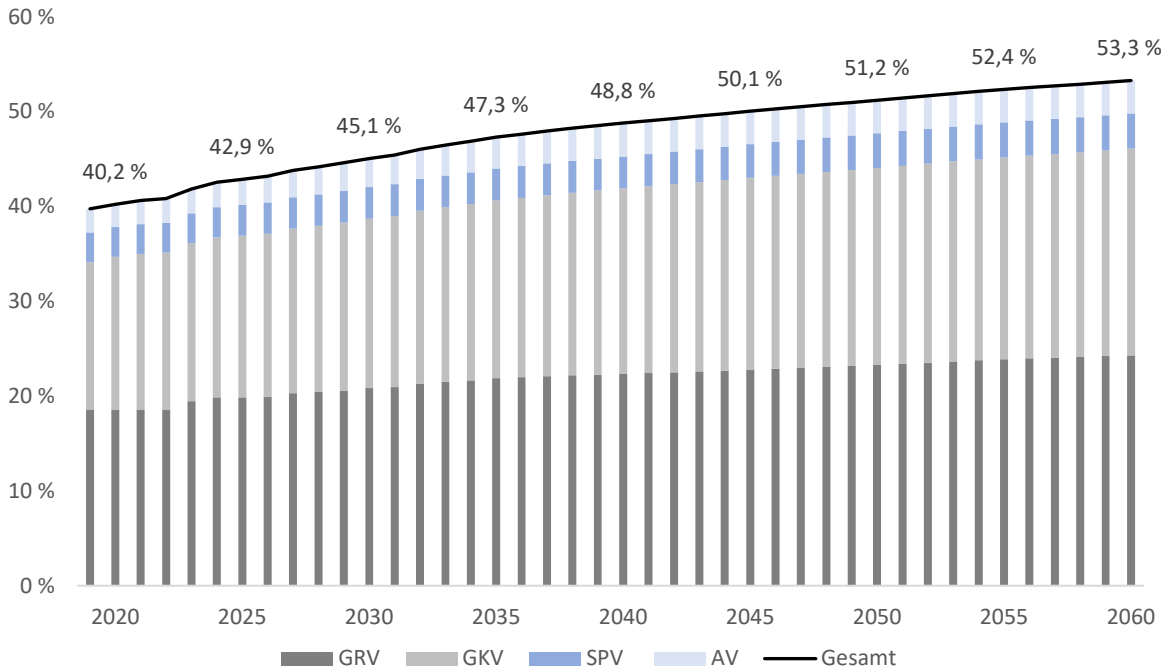
[86] Ergebnis Koalitionsausschuss vom 3. Juni 2020; Corona-Folgen bekämpfen, Wohlstand sichern, Zukunftsfähigkeit stärken, Seite 2.

[87] Der Finanzbedarf wurde für das Jahr 2020 auf 5,3 Mrd. Euro geschätzt. Vgl. Ergebnis Koalitionsausschuss vom 3. Juni 2020 a.a.O., Seite 2.

Abbildung 18

Die Sozialversicherungsbeitragssätze steigen in den Jahren 2019 bis 2060 deutlich an

In den letzten Jahren lag der Gesamtsozialversicherungsbeitragssatz bei knapp 40 %. Auf der Basis der projizierten Ausgaben müsste der Gesamtsozialversicherungsbeitragssatz bis zum Jahr 2060 kontinuierlich auf 53,3 % steigen. Die Abbildung zeigt die Entwicklung und die Zusammensetzung des Gesamtsozialversicherungsbeitragssatzes in Prozent.



Grafik: Bundesrechnungshof.

Quellen: DRV, BMG, BA, Statistisches Bundesamt, Werding, Läpple (2021).

6.4 Mittel aus dem Bundeshaushalt

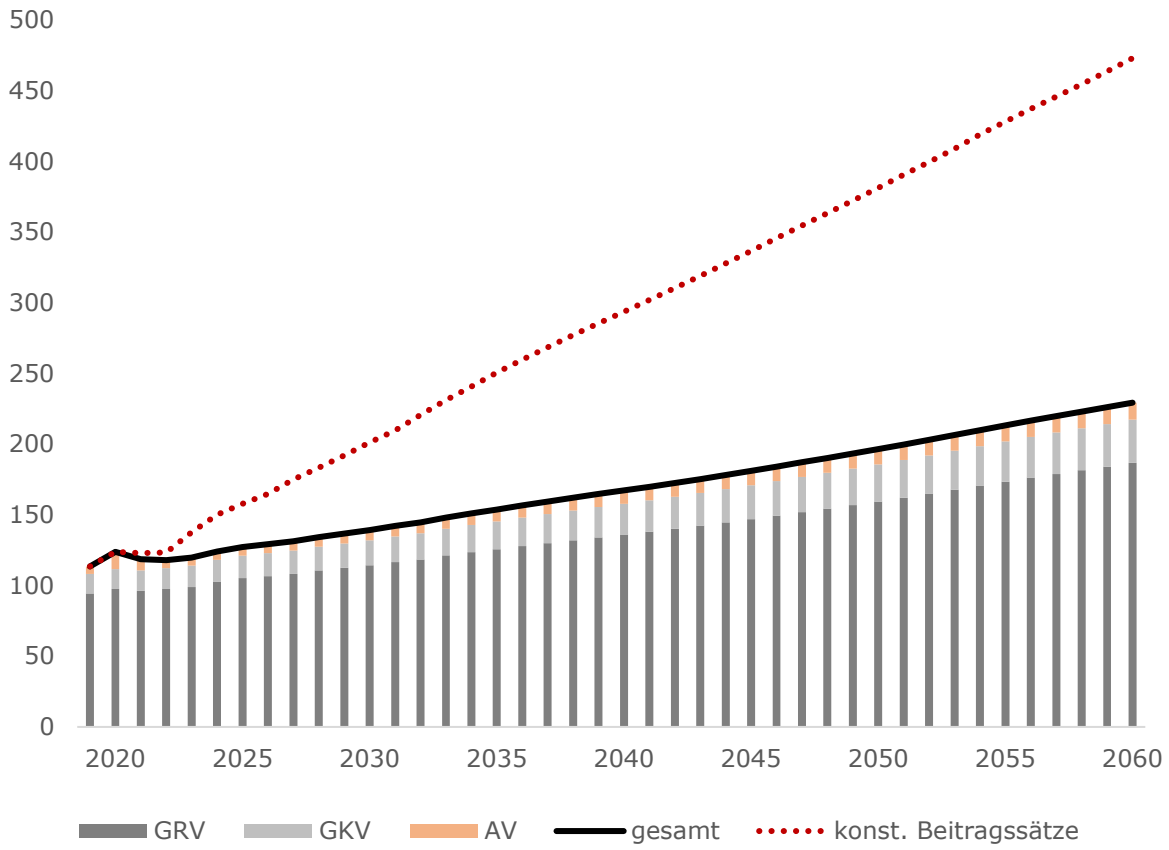
Alle betrachteten Sozialversicherungszweige erhalten Zuschüsse und Beiträge aus dem Bundeshaushalt. Zum einen sollen dadurch pauschal Leistungen ausgeglichen werden, die nicht in die Finanzierungsverantwortung der Beitragszahlenden, sondern der gesamten Gesellschaft fallen (versicherungsfremde Leistungen). Darüber hinaus dienen sie der Begrenzung der finanziellen Lasten der Beitragszahlenden und zur finanziellen Stabilisierung des Systems.

Zusammen beliefen sich die Mittel aus dem Bundeshaushalt für die Sozialversicherungen im Jahr 2019 auf 113,5 Mrd. Euro (in Preisen von 2015). Aufgrund der zunehmenden Alterung der Bevölkerung in den nächsten Jahrzehnten ist mit einem erhöhten Mittelbedarf zu rechnen. Unter Beibehaltung der geltenden Rechtslage verdoppeln sich im Basisszenario die Bundesmittel für die Sozialversicherungen bis zum Jahr 2060 auf 229,4 Mrd. Euro (vgl. Abbildung 19). Das entspricht einem Anstieg von 3,5 % des BIP im Jahr 2019 auf 4,9 % bis zum Jahr 2060.

Abbildung 19

Die Bundesmittel für die Sozialversicherungen steigen im Basisszenario deutlich

Die Bundesmittel steigen in Preisen des Jahres 2015 von 113,5 Mrd. Euro im Jahr 2019 auf 229,4 Mrd. Euro im Jahr 2060. Soweit die Sozialversicherungsbeitragssätze auf dem derzeitigen Niveau gehalten werden sollen, müssten die Ausgabensteigerungen vollständig über den Bundeshaushalt finanziert werden. In Preisen des Jahres 2015 würden dann die Bundesmittel für die Sozialversicherungen von 113,5 Mrd. Euro auf 473,1 Mrd. Euro im Jahr 2060 steigen. Die Abbildung zeigt die Bundesmittel in Mrd. Euro in den Jahren 2019 bis 2060.



Grafik: Bundesrechnungshof.

Quellen: DRV, BMG, BA, Statistisches Bundesamt, Werding, Läpple (2021).

Soweit die Sozialversicherungsbeitragssätze auf dem derzeitigen Niveau gehalten werden sollen, müssten die Ausgabensteigerungen vollständig über den Bundeshaushalt finanziert werden. Dies entspräche einem Szenario, in dem die Sozialgarantie über das Jahr 2021 hinaus verlängert würde. In Preisen des Jahres 2015 würden dann die Bundesmittel für die Sozialversicherungen von 113,5 Mrd. Euro auf 473,1 Mrd. Euro im Jahr 2060 steigen (vgl. Abbildung 16). Sie würden dann nicht 4,9 % des BIP, sondern 10,1 % ausmachen. Ein weiterer Vergleich verdeutlicht das Ausmaß dieses Ausgabenanstieges: Im Jahr 2019 hatte der gesamte Bundeshaushalt einen Anteil von 13,0 % am BIP.

6.5 Gesamte öffentliche Finanzen und Tragfähigkeitskennziffern

Die langfristige Tragfähigkeit der Sozialversicherungen und des gesamtstaatlichen Haushalts wird von der demografischen Entwicklung stark beeinflusst. Tragfähigkeitsprobleme werden nachfolgend durch die folgenden Kennziffern identifiziert:

- Gesamtstaatlicher Primärsaldo und gesamtstaatlicher Finanzierungssaldo sowie
- der gesamtstaatliche Schuldenstand relativ zum BIP.

Um diese Kennziffern zu berechnen, wird unterstellt, dass alle öffentlichen Einnahmen in Prozent des BIP konstant bleiben. Die Ausgaben der genannten öffentlichen Haushalte entwickeln sich dagegen wie oben für das Basisszenario projiziert. Differenzen zwischen Einnahmen und Ausgaben werden annahmegemäß durch zusätzliche öffentliche Schulden finanziert. Diese Berechnungsweise der Kennziffern ist eine Konvention, durch die zunächst die Größe etwaiger Tragfähigkeitsprobleme offengelegt werden soll, bevor über mögliche Maßnahmen zu ihrer Lösung nachgedacht wird.

Unter diesen Annahmen wird sich der gesamtstaatliche Primärsaldo[88] – nachdem er sich im Jahr 2020 durch die Covid-19 bedingten Ausgaben kurzfristig enorm verschlechtert hat – mittelfristig wieder erholen und positive Werte (Primärüberschuss) aufweisen. Ab dem Jahr 2028 ist mit einem Primärdefizit zu rechnen, das bis auf -2,7 % des BIP im Jahr 2040 und -5,2 % des BIP im Jahr 2060 zunimmt (vgl. Abbildung 20 a). Der Finanzierungssaldo ist nach dem Auslaufen der krisenbedingten Sondereffekte ab dem Jahr 2027 defizitär und bleibt dauerhaft im Defizit. Im Jahr 2040 wird das gesamtstaatliche Finanzierungsdefizit rechnerisch bei -3,4 % liegen und -8,8 % des BIP im Jahr 2060 erreichen.

Der Unterschied zwischen Primärsaldo und Finanzierungssaldo wird im Laufe der Zeit größer. Dies ist auf die wachsende Zinslast für die Staatsschulden zurückzuführen, die sich nur im Finanzierungssaldo bemerkbar macht und stärker zunimmt als das BIP.

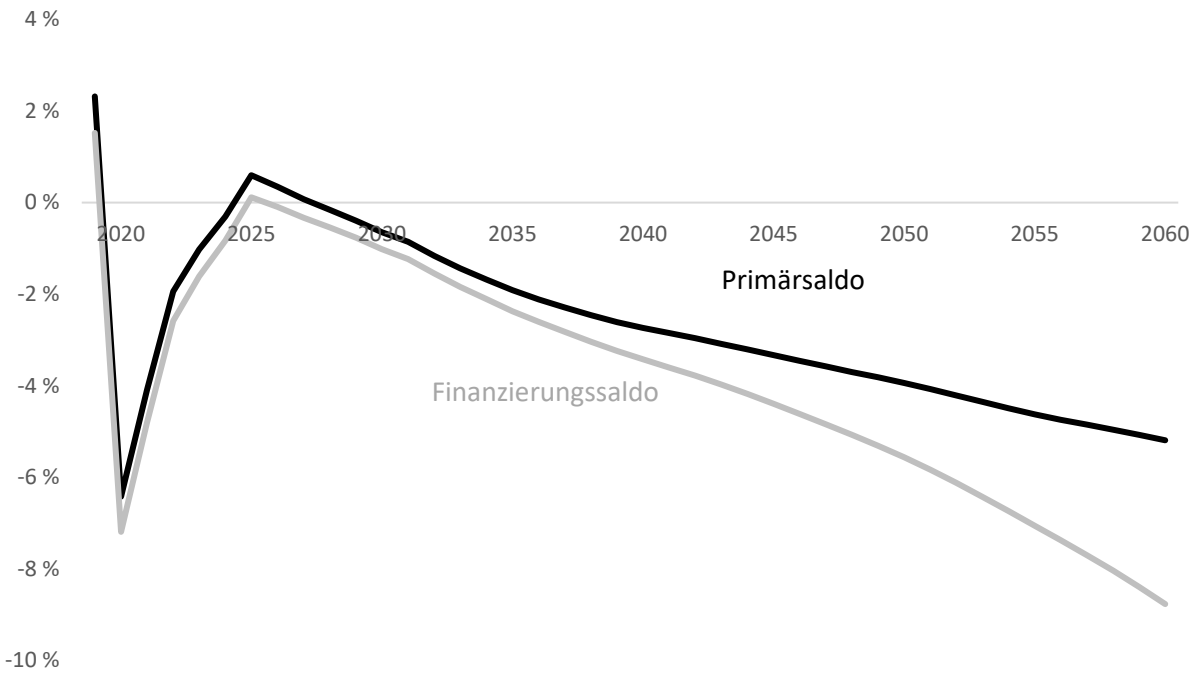
[88] Der Finanzierungssaldo der öffentlichen Haushalte entspricht der Differenz zwischen Staatseinnahmen und Staatsausgaben. Kreditaufnahmen zählen dabei nicht zu den Einnahmen und Tilgungen der Staatsschulden nicht zu den Ausgaben. Der Primärsaldo ist die Differenz zwischen den Staatseinnahmen (ohne Nettokreditaufnahme) und Staatsausgaben abzüglich der Zinszahlungen auf die Staatsschulden.

Abbildung 20

Gesamtstaatliche öffentliche Finanzen in den Jahren 2019 bis 2060 verschlechtern sich ohne Reformen deutlich

a) Gesamtstaatliche Finanzierungssalden

Die Abbildung zeigt den gesamtstaatlichen Primärsaldo, der sich im Jahr 2020 durch die Covid-19 bedingten Ausgaben kurzfristig verschlechtert hat. Er wird sich mittelfristig wieder erholen. Ab dem Jahr 2028 ist mit einem Primärdefizit zu rechnen, das bis auf -5,2 % des BIP im Jahr 2060 zunimmt. Der Finanzierungssaldo ist nach dem Auslaufen der krisenbedingten Sondereffekte ab dem Jahr 2027 dauerhaft defizitär und wird -8,8 % des BIP im Jahr 2060 erreichen. Die Abbildung zeigt den Primärsaldo und den Finanzierungssalden in Prozent des BIP bis zum Jahr 2060 unter der Annahme, dass die Einnahmequoten konstant bleiben und Defizite durch zusätzliche Schulden gedeckt werden.

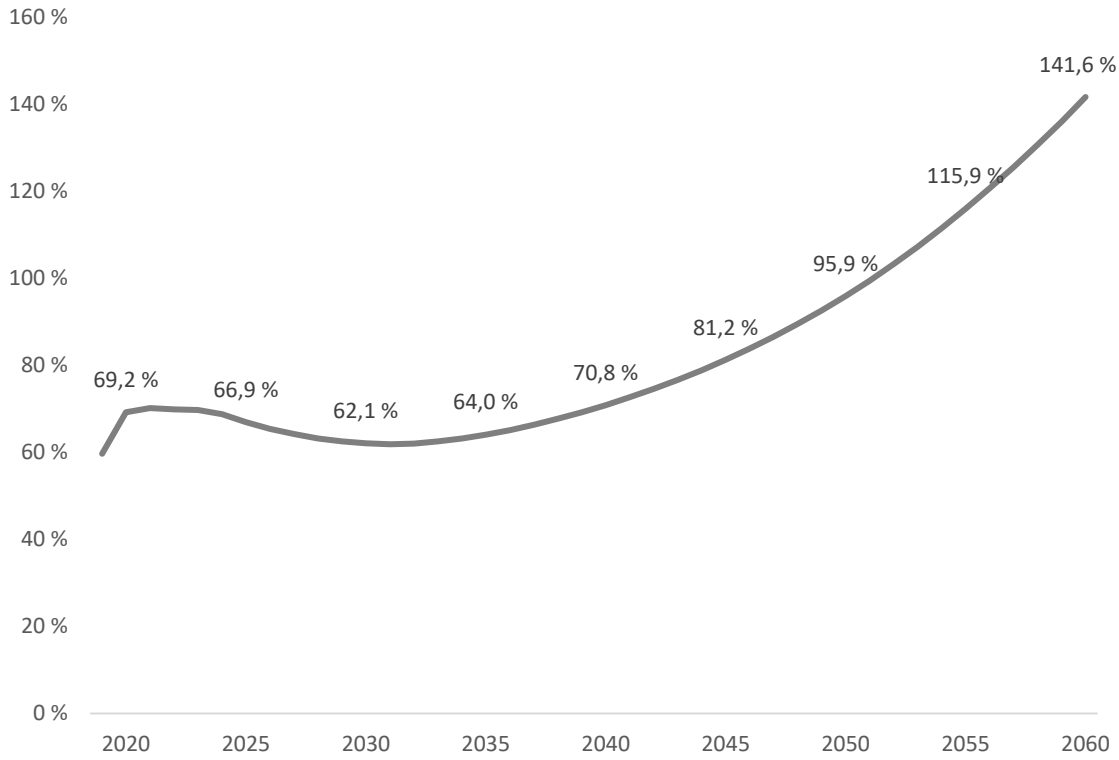


Grafik: Bundesrechnungshof.

Quellen: Statistisches Bundesamt (VGR), Deutsche Bundesbank, Werding, Läpple (2021).

b) Schuldenstand

Die gesamtstaatliche Schuldenstandsquote fiel im Jahr 2019 unter den „Maastricht-Referenzwert“ von 60 % des BIP. Sie wird sich langsam erhöhen und bis zum Jahr 2060 auf 141,6 % des BIP ansteigen. Die Abbildung zeigt die Schuldenstandsquote in Prozent des BIP bis zum Jahr 2060 unter der Annahme, dass die Einnahmequoten konstant bleiben und Defizite durch zusätzliche Schulden gedeckt werden.



Grafik: Bundesrechnungshof.

Quellen: Statistisches Bundesamt (VGR), Deutsche Bundesbank, Werding, Läpple (2021).

Die Entwicklung der Finanzierungssalden schlägt sich in einer entsprechenden Entwicklung des Schuldenstandes nieder. Bedingt durch die gute Konjunktur, die niedrigen Zinsen und eine nicht allzu expansive Haushaltsführung fiel die Schuldenstandsquote im Jahr 2019 unter den „Maastricht-Referenzwert“ von 60 %. Sie wird weiter leicht sinken, solange die gesamtstaatlichen Finanzierungssalden noch positive Werte aufweisen. Im Vergleich zu den Finanzierungssalden etwas verzögert folgt jedoch auch hier die Trendumkehr: Die gesamtstaatliche Schuldenstandsquote wird sich langsam erhöhen. Im Jahr 2031 wird sie 61,8 % des BIP und 70,8 % im Jahr 2040 erreichen. Der Projektion zufolge wird sie bis zum Jahr 2060 auf rechnerisch 141,6 % des BIP ansteigen (vgl. Abbildung 20 b).

6.6 Fiskalische Tragfähigkeit der öffentlichen Haushalte

Die fiskalische Tragfähigkeit öffentlicher Haushalte wird anhand von sogenannten Tragfähigkeitsindikatoren gemessen. Um sie zu berechnen[89], wird für die Ausgabenseite die vorher für das Basisszenario simulierte zukünftige Entwicklung der gesamtstaatlichen Ausgaben[90] unterstellt. Für die Einnahmeseite wird wieder angenommen, dass die gesamtstaatliche Einnahmenquote (in Prozent des BIP) in Zukunft konstant bleibt. Vereinfacht ausgedrückt geben die Differenzen zwischen dem Ausgabenpfad und dem Einnahmenpfad die Tragfähigkeitslücke an. Dabei kann die Tragfähigkeitslücke auf unterschiedliche Arten gemessen werden. Eine Möglichkeit besteht darin, die Tragfähigkeitslücke durch den Korrekturbedarf der zukünftigen Primärsalden zu beschreiben. Die Indikatorwerte zeigen Korrekturen der zukünftigen Primärsalden (in Prozent des laufenden BIP) an, die ab dem Jahr 2021 dauerhaft erforderlich sind, damit die zukünftige Entwicklung der gesamtstaatlichen öffentlichen Finanzen als „tragfähig" bezeichnet und die Tragfähigkeitslücke also geschlossen werden kann.

Aus den Simulationen für das Basisszenario ergibt sich hier eine Tragfähigkeitslücke in Höhe von 5,1 % des BIP. Das heißt, dass ab sofort und dauerhaft die öffentlichen Haushalte ihren Primärsaldo um 5,1 % des BIP verbessern müssen, um die Tragfähigkeitslücke zu schließen. Dies entspricht aktuell (auf Basis von Werten für das Jahr 2019) einem jährlichen Konsolidierungsbedarf von 11,4 % der gesamtstaatlichen Primärausgaben oder von 175,3 Mrd. Euro.

Die gesamte Tragfähigkeitslücke lässt sich auch aufgliedern nach den hier betrachteten Funktionen und erfassten Institutionen (vgl. Abbildung 21). Den größten Einzelbeitrag leistet dabei die GKV mit 2,2 Prozentpunkten, gefolgt von der GRV mit 1,6 Prozentpunkten. Gemessen an der Größe des Sicherungssystems erscheinen aber auch die Anteile der Beamtenversorgung und der SPV mit 0,3 bzw. 0,2 Prozentpunkten als nicht gering.

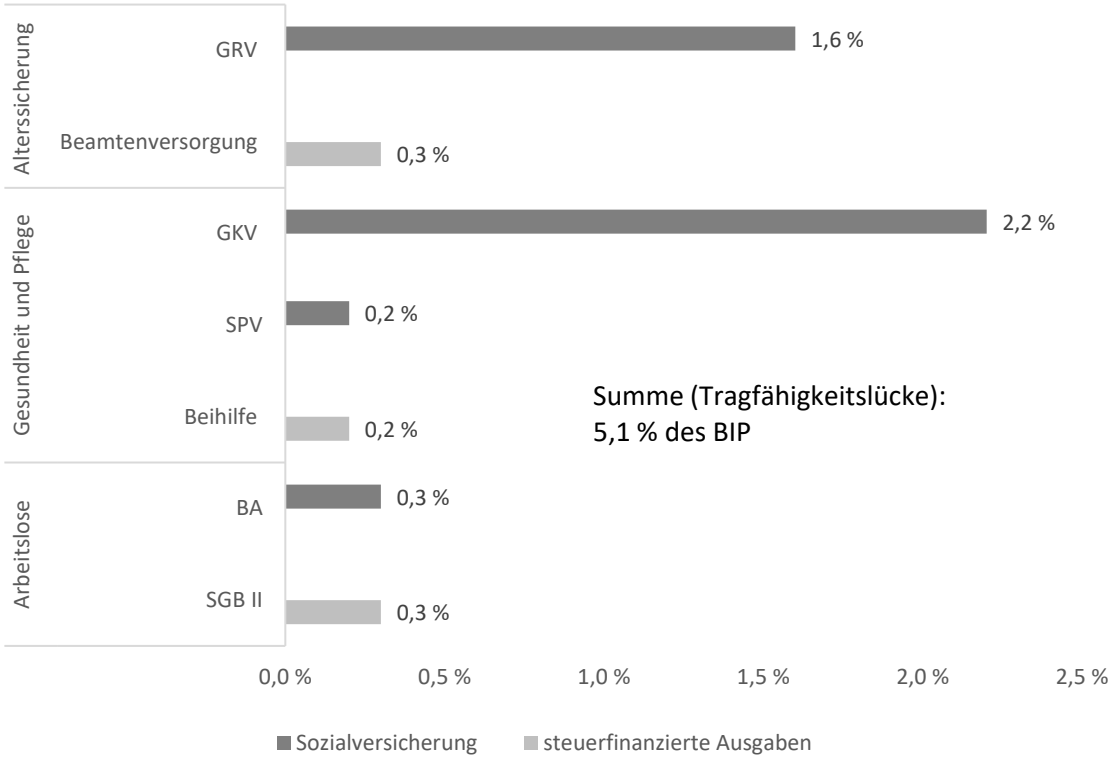
[89] Grundlage ist die sogenannte intertemporale Budgetbeschränkung. Diese besagt, dass der Staatshaushalt kurzfristig jederzeit Defizite aufweisen kann, aber über einen virtuell unendlichen Zeithorizont mit allen seinen zukünftigen Einnahmen in der Lage sein muss, alle zukünftigen Ausgaben und den in der Vergangenheit aufgelaufenen Schuldenstand zu decken. Anders ausgedrückt: Der Barwert aller zukünftigen Ausgaben plus aufgelaufene Schulden muss dem Barwert aller zukünftigen Einnahmen entsprechen.

[90] Dabei werden die nur für einen begrenzten Zeitraum (bis zum Jahr 2060) angestellten Ausgabenprojektionen auf einen unendlichen Zeithorizont ausgedehnt, indem unterstellt wird, dass die für das Jahr 2060 projizierte Ausgabenquote anschließend konstant bleibt. Die Erfassung zukünftiger Ausgaben wäre dabei vollständig, wenn die Situation im Jahr 2060 als „Steady State" angesehen werden könnte. Längerfristig angelegte Projektionen zeigen an, dass dies nicht der Fall sein dürfte, sondern dass aus heutiger Sicht auch bis zum Jahr 2080 mit einem weiteren Anstieg der gesamtstaatlichen Ausgabenquote zu rechnen ist (Werding, 2018). In diesem Fall unterschätzt der Tragfähigkeitsindikator die bestehende Tragfähigkeitslücke, stellt aber gleichwohl das angesichts der Ausgangsmaterialien für die vorliegenden Berechnungen (u. a. Bevölkerungsvorausberechnungen, die nur bis zum Jahr 2060 reichen) umfassendste Maß der Tragfähigkeitsprobleme dar.

Abbildung 21

Tragfähigkeitslücke in der GKV und der GRV am größten

Die Abbildung zeigt die Tragfähigkeitslücke ab dem Jahr 2021 in Höhe von 5,1 % des BIP und ihre Zusammensetzung nach den hier betrachteten Funktionen und Institutionen. Sie zeigt, dass ab sofort und dauerhaft die öffentlichen Haushalte den Primärsaldo um 5,1 % des BIP verbessern müssen, um die Tragfähigkeitslücke zu schließen.



Grafik: Bundesrechnungshof.

Quellen: Werding, Läpple (2021).

Der durch die Änderung des Primärsaldos angezeigte Konsolidierungsbedarf kann prinzipiell sowohl durch eine entsprechende Erhöhung der Einnahmenquoten als auch durch eine Senkung der zukünftigen Ausgabenquoten – oder durch eine Kombination aus beidem – realisiert werden. Die erforderlichen Einnahmeerhöhungen werden mit Anhebungen der Beitragssätze der Sozialversicherungen (sowie von Steuern zur Deckung der erforderlichen Bundesmittel sowie der rein steuerfinanzierten Ausgaben) erreicht. Alternativ dazu kann der Tragfähigkeitsindikator auch in eine Anpassung der Steuerquote umgerechnet werden, die – dann ohne eine Anpassung der Beitragssätze in der Sozialversicherung – zu den notwendigen Konsolidierungen führen würde. Um diese zu erreichen, müsste die Steuerquote sofort und dauerhaft (auf Basis von Werten für das Jahr 2019) von 23,2 % auf 28,3 % des BIP angehoben werden.

Die erforderlichen Ausgabensenkungen müssen keineswegs nur bei den hier explizit betrachteten Ausgaben ansetzen, die von der demografischen Alterung besonders betroffen sind. Denn dies könne den Sicherungszwecken der Sozialversicherungen und der anderen Ausgaben zuwiderlaufen. In Frage kommen daneben auch alle sonstigen staatlichen Ausgaben.

Die betrachteten Indikatoren für die Entwicklung der öffentlichen Finanzen lassen den Schluss zu, dass die öffentlichen Finanzen aufgrund der Effekte der demografischen Alterung langfristig Tragfähigkeitsprobleme haben werden. Die Tragfähigkeitslücken zeigen somit einen Handlungsbedarf an.

6.7 Auswirkungen von Alternativvarianten

Mit dem Basisszenario wurde eine moderate Variante für die zukünftige Entwicklung der Bevölkerung, der Wirtschaft und der Sozialversicherungen betrachtet. Um die Auswirkungen von Abweichungen von diesem Szenario abschätzen zu können, wurden folgende abweichende Annahmen geprüft:

- niedrigere Geburtenziffer: die zusammengefasste Geburtenziffer geht bis zum Jahr 2040 wieder auf 1,43 zurück, statt sich – wie im Basisszenario – bei 1,55 zu stabilisieren, und bleibt anschließend konstant,
- höhere Lebenserwartung: die Lebenserwartung bei Geburt steigt für Frauen bis zum Jahr 2060 auf 89,6 Jahre, für Männer auf 86,2 Jahre, statt auf 88,1 Jahre bzw. 84,4 Jahre wie im Basisszenario,
- höheres Produktivitätswachstum: die Wachstumsrate der totalen Faktorproduktivität beträgt im gesamten Simulationszeitraum 1,0 % p.a. statt 0,7 % und erhöht damit auch die Wachstumsraten der Arbeitsproduktivität und der Löhne entsprechend,
- lohnorientierte Anpassungen der Pflegeleistungen (die Leistungen der SPV werden nicht mit einem Mischindex aus Preis- und Lohnsteigerungen angepasst, sondern allein gemäß der Wachstumsrate der Löhne).

Die Simulation zeigt, dass sich die veränderten Annahmen auf die Ausgabenquote nur wenig auswirken. Die Ausgabenquote steigt von 21,8 % im Jahr 2019 auf 28,8 % des BIP im Jahr 2060 anstatt auf 29,3 % im Basisszenario. Die Variante mit lohnorientierten Anpassungen der Pflegeleistungen entfaltet hier die stärksten Effekte. So steigt in dieser Variante die für das Jahr 2060 projizierte Ausgabenquote aufgrund erhöhter Ausgaben der SPV um 0,6 Prozentpunkte.

Ähnliche Effekte ergeben sich in der Variante mit einem stärkeren Anstieg der Lebenserwartung: Dort ist die für das Jahr 2060 projizierte Ausgabenquote um 0,5 Prozentpunkte höher als im Basisszenario. In der Variante mit niedrigerer Geburtenziffer, aber auch in der Variante mit einem stärkeren Wachstum von Arbeitsproduktivität und Löhnen, variiert die Ausgabenquote im Jahr 2060 dagegen nur um ±0,2 Prozentpunkte. Ein höheres Produktivitätswachstum hat demnach zwar große Effekte für die Entwicklung des BIP, aber – unter den derzeitigen rechtlichen Rahmenbedingungen – jeweils auch nahezu gleich starke Effekte für die hier erfassten Ausgaben.

Auch bezogen auf den Gesamtsozialversicherungsbeitragssatz, für den im Basisszenario ein Anstieg von 40,0 % im Jahr 2020 auf 53,3 % im Jahr 2060 berechnet wird, sind die Abwei-

chungen bei anderen Annahmen eher gering. Die Abweichungen bei den Gesamtsozialversicherungsbeitragssätzen liegen bei +/- 0,3 bis 1,2 Prozentpunkten.

Auf die gesamtstaatliche Schuldenstandsquote haben die abweichenden Annahmen einen etwas größeren Einfluss, wenn annahmegemäß die Sozialversicherungsbeitragssätze konstant blieben[91] und der Bund die Ausgabensteigerungen bei den Sozialversicherungen über zusätzliche Schulden finanziert. Statt dem Anstieg der gesamtstaatlichen Schuldenstandsquote auf 142 % des BIP im Jahr 2060 variiert die Verschuldung in den Alternativszenarien. Bei geringerer Geburtenziffer und stärkerem Anstieg der Lebenserwartung erhöht sich der Wert leicht auf 144 % bzw. 149 % des BIP. Der Anstieg verstärkt sich in der Variante mit lohnorientierten Anpassungen der Pflegeleistungen auf 153 % des BIP. Hingegen reduziert sich die gesamtstaatliche Schuldenstandsquote auf 128 % im Falle eines höheren Produktivitätswachstums.

Auf die Tragfähigkeitslücke und die erforderliche Änderung der Steuerquote wirkt sich die geänderte Entwicklung der Leistungspauschalen in der SPV am stärksten aus. Die Tragfähigkeitslücke ist um +0,6 Prozentpunkte höher als im Basisszenario. Eine höhere Lebenserwartung hat ebenfalls einen deutlichen Effekt auf die Tragfähigkeitslücke von +0,5 Prozentpunkten (vgl. Tabelle 2).

[91] Diese Schuldenstandsquote bei konstanten Beitragssätzen ist eine Näherung für die in Tnr. 6.5 dargestellte Schuldenstandsquote bei konstanten Einnahmequoten. Zusätzlich zu den konstanten Beitragssätzen wird unterstellt, dass auch ein überproportionaler Anstieg der steuerfinanzierten Sozialausgaben für die Beamtenversorgung, die Beihilfe und für einen Teil der Arbeitsmarktmaßnahmen (v.a. Grundsicherungsleistungen) durch zusätzliche Verschuldung finanziert wird. Diese zusätzliche Annahme gilt im Folgenden immer, wenn die Entwicklung der Schuldenstandsquote bei konstanten Beitragssätzen beschrieben wird.

Tabelle 2

Tragfähigkeitsindikatoren ausgewählter Alternativvarianten

Variante	Tragfähigkeitslücke[a)] *in %*	Erforderliche Steuerquote[b)] *in %*
Basisszenario	***5,1***	***28,3***
Alternativvarianten:		
niedrigere Geburtenziffer	*5,3*	*28,4*
höhere Lebenserwartung	*5,6*	*28,7*
höheres Produktivitätswachstum	*5,1*	*28,2*
lohnor. Anpassg. Pflegeleistungen	*5,7*	*28,8*

Anmerkungen:
a) Ab sofort und dauerhaft erforderliche Korrektur der Primärsalden des gesamtstaatlichen Haushalts zur Einhaltung der intertemporalen Budgetbeschränkung des Staates (in Prozent des BIP).
b) Zur Schließung der Tragfähigkeitslücke allein durch höhere Steuereinnahmen (ausgehend vom Wert der Steuerquote für das Jahr 2019 von 23,2 % des BIP).

Tabelle: Bundesrechnungshof.

Quelle: Werding, Läpple (2021).

6.8 Bewertung der Ergebnisse des Basisszenarios und mögliche Lösungsansätze

Die demografische Entwicklung würde unter Beibehaltung der derzeitigen Rechtslage dazu führen, dass der Gesamtsozialversicherungsbeitragssatz bis zum Jahr 2060 auf 53,3 % steigt. Somit besteht die Gefahr einer finanziellen Überlastung sowohl für die Beschäftigten als auch für deren Arbeitgeber. Gleichzeitig würden sich die Bundesmittel von 113,5 Mrd. Euro im Jahr 2019 (preisbereinigt auf das Jahr 2015) bis zum Jahr 2060 verdoppeln, wodurch sich deren Anteil am BIP um rund 40 % auf 4,9 % des BIP erhöhen würde. Dies könnte zu einer finanziellen Überlastung des Bundes, mithin der Steuerzahlenden, führen. Die Tragfähigkeitslücke beträgt insgesamt 4 % des BIP für die GRV, die GKV und die SPV. Diese drei Sozialversicherungszweige wären damit für 80 % der Tragfähigkeitslücke des Gesamtstaates verantwortlich.

Trotz dieser dramatischen Finanzentwicklung würde sich das Rentenniveau bis zum Ende des Betrachtungszeitraumes auf 43,4 % vermindern, was die Akzeptanz des Rentenversicherungssystems gefährden könnte. Auch veränderte Annahmen zur Lebenserwartung, zur wirtschaftlichen Entwicklung, zur Geburtenrate oder zur Erwerbstätigkeit führen nicht dazu, dass sich die aufgezeigte finanzielle Situation grundlegend ändert.

Eine derartige Steigerung des Gesamtsozialversicherungsbeitragssatzes und der Bundeszuschüsse bei gleichzeitig sinkendem Rentenniveau ist nach Auffassung des Bundesrechnungshofes nicht tragbar, was Reformen unausweichlich werden lässt.

Angesichts der ohnehin starken finanziellen Belastung des Bundes durch die steigenden Bundeszuschüsse erscheint eine Begrenzung des Gesamtsozialversicherungsbeitragssatzes durch zusätzliche Bundesmittel kaum möglich. Soweit zusätzliche Bundesmittel nicht durch Umschichtungen im Bundeshaushalt oder zusätzliche Steuereinnahmen gedeckt werden, würde dies zu einer deutlichen Erhöhung der Verschuldung auf 141,6 % des BIP führen. Die finanziellen Belastungen aufgrund der demografischen Entwicklung würden nur in die Zukunft, zu Lasten der künftigen Generationen, verschoben. Die Verschuldung des Bundes würde, insbesondere bei steigenden Zinsen, die Grenze der Tragfähigkeit erreichen und den finanziellen Handlungsspielraum künftiger Regierungen und Generationen stark beschränken.

Reformen der GRV, GKV und SPV müssen daher so angelegt sein, dass die absehbaren finanziellen Belastungen der demografischen Entwicklung gleichmäßiger zwischen den Generationen aufgeteilt werden.

Aus Sicht des Bundesrechnungshofes müssen diese Reformen daher an verschiedenen Stellen und bei allen Sozialversicherungszweigen ansetzen. Daher sollte ein Reformbündel geschnürt werden, mit dem die folgenden Ziele erreicht werden können:

- Verminderung der Finanzierungslasten für die erwerbstätige Bevölkerung, z. B. durch eine Verringerung des Beitragssatzanstieges.
- Intergenerativer Lastentransfer: Indem Beitragsbelastungen vorgezogen werden, können heutige Generationen belastet und zukünftige Generationen entlastet werden.
- Erhöhung des Eigenfinanzierungsanteiles der Personen im Seniorenalter und
- Minderung der Beitragsbelastung in der Phase, in der die demografische Entwicklung zu einem starken Anstieg der Beitragssätze führt (Beitragssatzglättung).
- Dabei gilt es, Überbelastungen zu vermeiden. So sollte die Reduktion des Rentenniveaus begrenzt sein. Es sollte aber auch eine Überforderung des Bundes vermieden werden.

Fest steht, dass die Auswirkungen der demografischen Entwicklung nicht wegreformiert werden können. Vielmehr gilt es, Lasten zu reduzieren und gleichmäßiger auf Senioren, Beitragszahlende und den Bund und damit die derzeitigen und zukünftigen Steuerzahlenden zu verteilen, als dies derzeit in den Sozialversicherungszweigen angelegt ist.

7 Reformen in der Alterssicherung

7.1 Stellschrauben und Instrumente in der gesetzlichen Rentenversicherung

Die untersuchten Reformmaßnahmen in der GRV sollen eines oder mehrere der vorgenannten Ziele erfüllen. Die Reformmaßnahmen können dabei nur an den vier genannten Stellschrauben der GRV ansetzen (Tnr. 3.4):

- Beitragssatz,
- Rentenniveau,
- Rentnerquotient und
- Bundeszuschuss.

Um diese Stellschrauben zu beeinflussen, gibt es verschiedene Instrumente bzw. Reformmöglichkeiten. Nachfolgend werden einige dieser Instrumente bzw. Reformen beschrieben und anhand von Beispielen ihre Auswirkungen auf den GRV-Beitragssatz und das Rentenniveau simuliert.

7.2 Stellschraube Rentnerquotient

Die demografische Entwicklung schlägt sich durch die zunehmende Anzahl der Rentenbeziehenden und die abnehmende Anzahl der Beitragszahlenden in einer ungünstigen Entwicklung des Rentnerquotienten (zahlenmäßige Relation von Rentenbeziehenden zu Beitragszahlenden) nieder. Deshalb ist es naheliegend, mit Reformmaßnahmen an dieser Stellschraube anzusetzen. Die dafür geeigneten Instrumente sind:

- die Anhebung des Renteneintrittsalters und
- die Ausweitung des Versichertenkreises.

7.2.1 Instrument - Anhebung des gesetzlichen Renteneintrittsalters

Gehen Versicherte später in Rente, dämpft dies den Anstieg des Rentnerquotienten, weil die Zahl der Rentenbeziehenden nicht so schnell steigt und die Zahl der Beitragszahlenden nicht so schnell sinkt. Die GRV wird dadurch entlastet, weil die Rentenbezugsdauer kürzer und die Dauer der Beitragszahlung länger ist.

Schon im Jahr 2007 hatte der Gesetzgeber mit dem RV-Altersgrenzenanpassungsgesetz die Stellschraube Rentnerquotient beeinflusst. Ab dem Geburtsjahrgang 1947 erhöht sich die Regelaltersgrenze schrittweise bis zum Jahr 2031 auf 67 Jahre. Die neue Regelaltersgrenze

von 67 Jahren gilt erstmals für den Geburtsjahrgang 1964. Da die Lebenserwartung und damit die Rentenbezugszeit jedoch voraussichtlich weiter steigt, ist eine weitere Anhebung der Regelaltersgrenze eine Möglichkeit, die finanzielle Stabilität der GRV zu verbessern.

Denkbar wäre, die Regelaltersgrenze ausgehend von 67 Jahren nach Maßgabe der zunehmenden Lebenserwartung weiter zu erhöhen (dynamische Regelaltersgrenze). Als Standardvorschlag hat sich mittlerweile etabliert, dass sich die Regelaltersgrenze regelgebunden jeweils um zwei Drittel des prognostizierten durchschnittlichen Zuwachses der Lebenserwartung erhöht. Das verbleibende Drittel verlängert die erwartete Rentenlaufzeit. Steigt also die durchschnittliche Lebenserwartung um drei Jahre, erhöht sich die Regelaltersgrenze um zwei Jahre und die Rentenbezugszeit um ein Jahr.

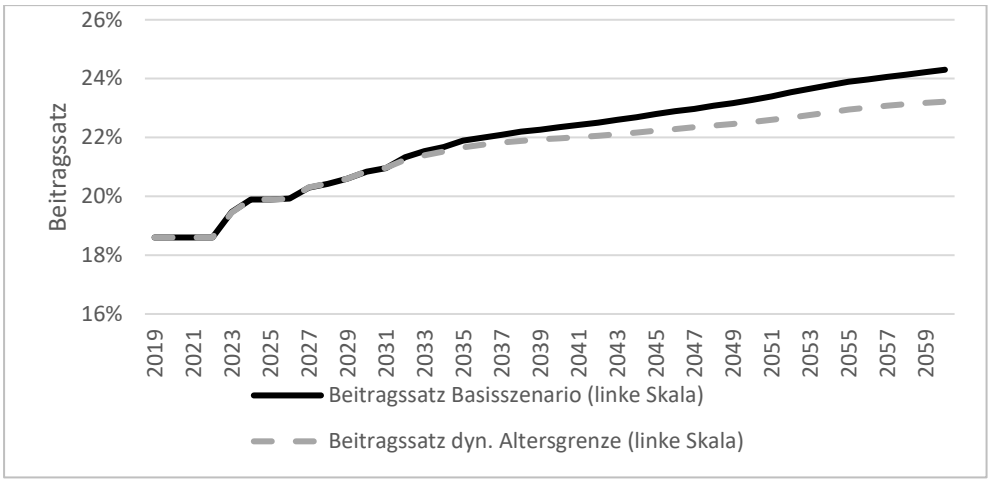
Wendet man diese Regel ab dem Jahr 2030 bzw. beginnend mit dem Geburtsjahrgang 1965 an, müsste die Regelaltersgrenze bis zum Jahr 2060 schrittweise auf 69,5 Jahre ansteigen. Diese Anhebung der Regelaltersgrenze würde den Beitragssatzanstieg bis zum Jahr 2060 dämpfen: Der Beitragssatz stiege auf 23,2 % anstatt auf 24,3 % im Basisszenario (vgl. Abbildung 22). Auch auf das Rentenniveau wirkt sich die Anhebung des gesetzlichen Renteneintrittsalters positiv aus: Rentnerquotient und Beitragssatz entwickeln sich günstiger. Damit wirken der Nachhaltigkeitsfaktor und der Beitragssatzfaktor in der Rentenanpassungsformel weniger rentendämpfend, sodass das Rentenniveau nicht so stark sinkt wie im Basisszenario. Im Jahr 2060 läge es bei 44,7 % und nicht bei 43,4 % (vgl. Abbildung 23).

Insgesamt ist die Anhebung der Regelaltersgrenze dazu geeignet, zukünftige Beitragszahlende zu entlasten und die Menschen im Seniorenalter stärker an den demografischen Lasten zu beteiligen, indem sie aufgrund ihrer gestiegenen Lebenserwartung erst später in Rente gehen oder höhere Abschläge in Kauf nehmen müssen.

Abbildung 22

Durch die Anhebung der Regelaltersgrenze ist der Beitragssatzanstieg geringer

Durch die schrittweise Anhebung des gesetzlichen Renteneintrittsalters ist der Beitragssatz im Jahr 2060 um 1,1 Prozentpunkte geringer als im Basisszenario. Die Abbildung zeigt die Entwicklung des Beitragssatzes in Prozent bis zum Jahr 2060.



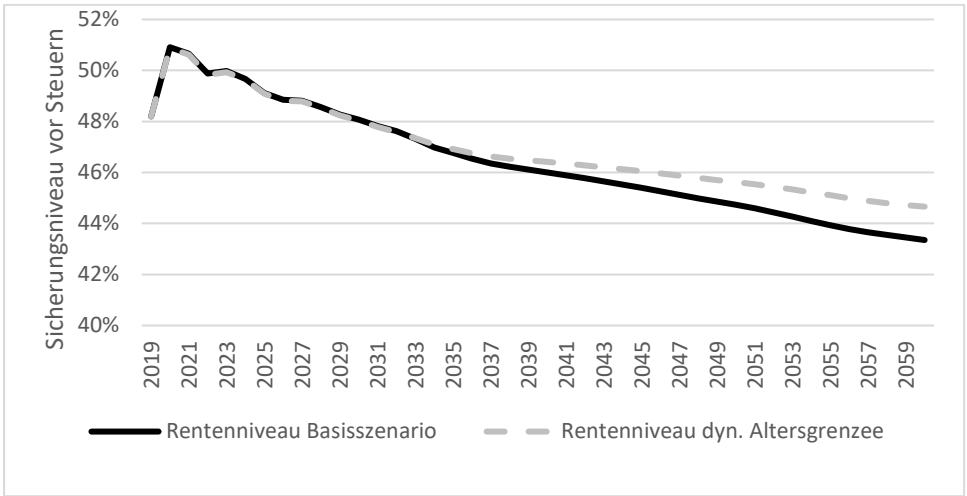
Grafik: Bundesrechnungshof.

Quelle für Grundzahlen: Werding, Läpple (2021).

Abbildung 23

Durch die Anhebung der Regelaltersgrenze ist das Rentenniveau höher

Durch die schrittweise Anhebung des gesetzlichen Renteneintrittsalters ist das Rentenniveau um 1,3 Prozentpunkte höher als im Basisszenario. Die Abbildung zeigt die Entwicklung des Rentenniveaus in Prozent bis zum Jahr 2060.



Grafik: Bundesrechnungshof.

Quelle für Grundzahlen: Werding, Läpple (2021).

7.2.2 Instrument - Ausweitung des Versichertenkreises

Wird der Versichertenkreis um zusätzliche Beitragszahlende erweitert, die sofort Beiträge zahlen, aber erst später Rente beziehen, wird der Rentnerquotient (vorübergehend) reduziert. Dies entlastet die GRV, weil sie von den zusätzlichen Versicherten Beiträge einnimmt und zunächst keine oder nur geringe Ausgaben für die zusätzlichen Versicherten hat.

Derzeit sind nicht alle Erwerbstätigen in der GRV versicherungspflichtig. Es wird daher regelmäßig gefordert, den Pflichtversichertenkreis der GRV z. B. um die Selbstständigen oder die Beamtinnen und Beamten auszuweiten.

7.2.2.1 Einbeziehung der Selbstständigen in die gesetzliche Rentenversicherung

Erweitert man beispielsweise die Versicherungspflicht auf die Selbstständigen, die bisher noch nicht in einem obligatorischen Alterssicherungssystem abgesichert sind, kann der Rentnerquotient kurzfristig reduziert werden. Zunächst sind die neuen Versicherten nur Beitragszahlende. Erst später werden sie zu Rentenbeziehenden. Dies dämpft den Beitragssatzanstieg in der GRV. Da dieses Instrument jedoch mittel- und langfristig stets zu einer Ausweitung des Kreises der Rentenempfänger führt, ist der Entlastungseffekt für das Rentensystem in der Regel nur von vorübergehender Natur. Der langfristige Effekt hängt von der Struktur des neu einbezogenen Versichertenkreises im Vergleich zu den „Altversicherten“ ab. Hat der neu einbezogene Personenkreis z. B. im Vergleich zu den „Altversicherten“ eine im Durchschnitt höhere Lebenserwartung, ist der langfristige Effekt sogar negativ. Das heißt, der Beitragssatz ist höher als ohne Reform. Ist die Lebenserwartung im Durchschnitt geringer, gibt es langfristig einen positiven Effekt auf den Beitragssatz.

Für das Reformszenario „Selbstständige in die GRV“ wird Folgendes unterstellt:

- Alle Selbstständigen, die bisher nicht über eine obligatorische Altersvorsorge verfügen, werden ab dem Jahr 2023 grundsätzlich in die GRV einbezogen. Die Größe des Adressatenkreises dieser Reform lässt sich nicht leicht beziffern. Er wird häufig auf etwa 80 % aller Selbstständigen, d. h. derzeit rund 3,2 Mio. Personen, geschätzt.[92]
- Bei Einführung der Versicherungspflicht wird eine Altersgrenze definiert: Selbstständig Tätige unterhalb dieser Altersgrenze werden in die GRV einbezogen. Ältere Selbstständige sind aufgrund eines Vertrauensschutzes nicht betroffen.
- Personen, die nach Einführung der Versicherungspflicht eine selbstständige Tätigkeit neu aufnehmen, werden dagegen unabhängig von ihrem Alter in die GRV einbezogen.
- Für die Höhe der Altersgrenze werden drei Varianten von 30 Jahren, 40 Jahren und 50 Jahren unterstellt.

[92] Vgl. z. B. Gasche, M. und Rausch J., Auswirkungen einer Einbeziehung der Selbständigen in die Gesetzliche Rentenversicherung, Perspektiven der Wirtschaftspolitik 14(3-4), 2013, S. 305-345. sowie Preis, U. und Temming, F., Für ein modernes Rentenrecht: Die Einbeziehung von Selbstständigen in die gesetzliche Rentenversicherung (GRV), Forschungsbericht Nr. 487, Bundesministerium für Arbeit und Soziales, Berlin 2017.

- Als beitragspflichtige Einnahmen gelten jeweils die Einkünfte aus selbstständiger Tätigkeit. Vereinfachend wird hier unterstellt, dass sie den durchschnittlichen beitragspflichtigen Entgelten der sozialversicherungspflichtig Beschäftigten nach geltendem Recht entsprechen.
- Mangels empirischer Daten zur Lebenserwartung dieses Kreises der Selbstständigen wird unterstellt, dass sich die Lebenserwartung Selbstständiger nicht von der der sonstigen Versicherten der GRV unterscheidet.[93]

In diesem Reformszenario ergibt sich für die mittlere Altersgrenze von 40 Jahren im Jahr 2023 ein um 0,7 Prozentpunkte geringerer Beitragssatz. Das Rentenniveau steigt gleichzeitig um 0,2 Prozentpunkte. Ein höheres Rentenniveau ergibt sich vor allem deshalb, weil durch die Reform der Rentnerquotient und damit der Nachhaltigkeitsfaktor in der Rentenanpassungsformel günstig beeinflusst werden. Die Rentenanpassungen fallen höher aus als im Basisszenario, und das Rentenniveau sinkt nicht so schnell (vgl. Abbildung 24). Im Jahr 2060 ist der Beitragssatz um 0,4 Prozentpunkte geringer und das Rentenniveau um 0,9 Prozentpunkte höher. Die maximalen Effekte werden in den Jahren 2047 bzw. 2049 erreicht und belaufen sich auf 0,8 Prozentpunkte für den Beitragssatz bzw. 1,3 Prozentpunkte für das Rentenniveau (vgl. Abbildung 25).

Je höher die Altersgrenze für eine Einbeziehung Selbstständiger in die GRV gesetzt wird, desto mehr wird der Anstieg des Beitragssatzes ab dem Jahr 2023 zunächst verlangsamt und aufgrund der Auswirkungen auf den Nachhaltigkeitsfaktor gleichzeitig auch der Rückgang des Rentenniveaus gedämpft. Eine höhere Altersgrenze bewirkt aber auch, dass Selbstständige, die in der GRV für eine gewisse Übergangszeit nur als zusätzliche Beitragszahlende auftreten, bereits rascher zu zusätzlichen Rentenbeziehenden werden. Dies beschleunigt den Anstieg des Beitragssatzes und auch den Rückgang des Rentenniveaus nach dem Jahr 2035 wieder.

Auf Dauer nähern sich die Werte für den Beitragssatz und für das Rentenniveau wieder ihren für das Basisszenario projizierten Werten an. Damit sind die Effekte für den Beitragssatz und das Rentenniveau nur vorübergehend. Im Fall einer recht hohen Altersgrenze von 50 Jahren ist dieser Prozess bis zum Jahr 2060 bereits weitgehend abgeschlossen (vgl. Abbildung 25), bei einer niedrigen Altersgrenze von 30 Jahren dauert er insgesamt zwei Jahrzehnte länger.

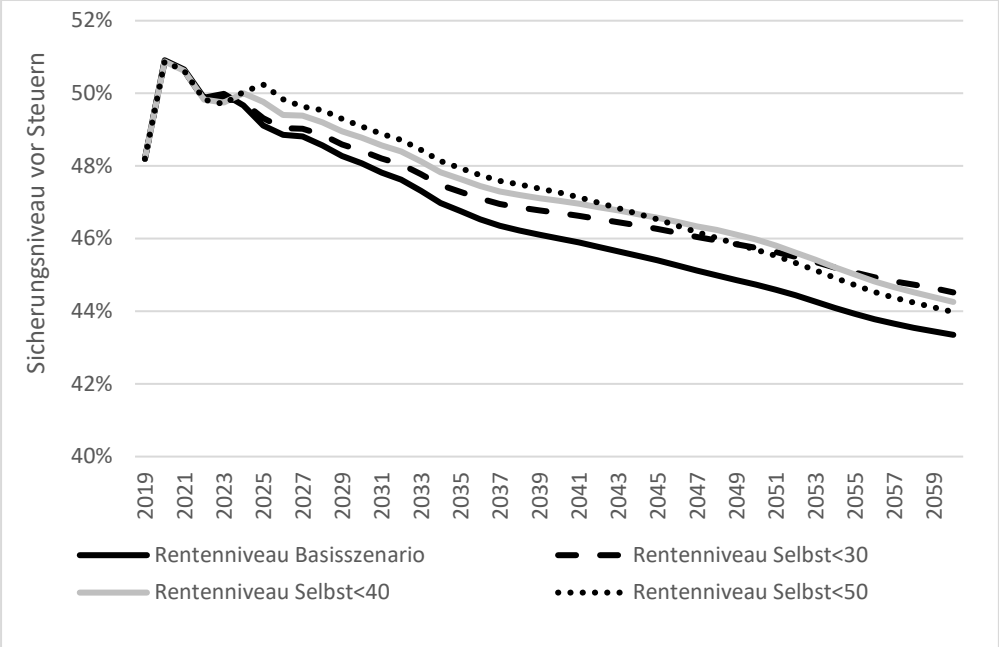
Mit der Einbeziehung der nicht obligatorisch abgesicherten Selbstständigen kann ein niedrigerer Beitragssatz in der Phase erreicht werden, in der der Beitragssatz besonders stark steigt (Beitragssatzglättung). Die Altersgrenze für die Einbeziehung der Selbstständigen kann dabei als „Timinginstrument“ dienen. Je höher die Altersgrenze, desto früher der Beitragssatzeffekt, desto früher ist die Beitragsentlastung aber auch wieder verschwunden.

[93] Diese Annahme ist plausibel, wenn man davon ausgeht, dass die beiden Personengruppen ähnlich hohe Einkommen haben und ähnliche Tätigkeiten ausüben.

Abbildung 24

Einbeziehung der Selbstständigen bewirkt (vorübergehend) höheres Rentenniveau

Durch die Einbeziehung der Selbstständigen ist das Rentenniveau höher als im Basisszenario. Die Abbildung zeigt die Entwicklung des Rentenniveaus in Prozent bis zum Jahr 2060.



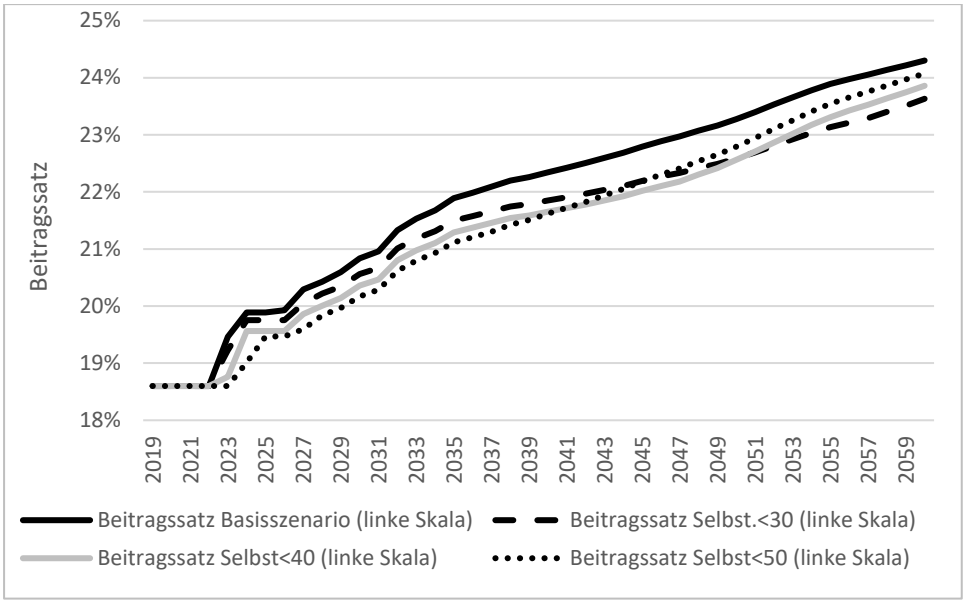
Grafik: Bundesrechnungshof.

Quelle für Grundzahlen: Werding, Läpple (2021).

Abbildung 25

Einbeziehung der Selbstständigen bewirkt (vorübergehend) niedrigeren Beitragssatz

Durch die Einbeziehung der Selbstständigen verläuft der Beitragssatz bis zum Jahr 2060 auf einem niedrigeren Niveau als im Basisszenario. Die Abbildung zeigt die Entwicklung des Beitragssatzes in Prozent bis zum Jahr 2060.



Grafik: Bundesrechnungshof.

Quelle für Grundzahlen: Werding, Läpple (2021).

7.2.2.2 Einbeziehung der Beamtinnen und Beamten in die gesetzliche Rentenversicherung

Auch die Einbeziehung der Beamtinnen und Beamten in die GRV wird regelmäßig gefordert. Auf den Beitragssatz und das Rentenniveau hätte diese Erweiterung des Versichertenkreises kurz- und mittelfristig ähnliche positive Effekte wie die Einbeziehung der Selbstständigen. Im Gegensatz zur Einbeziehung der Selbstständigen ist jedoch davon auszugehen, dass der Beitragssatz langfristig höher und das Rentenniveau niedriger sein würden als im Basisszenario. Dies liegt daran, dass die Beamtinnen und Beamten – empirisch gesichert – eine höhere Lebenserwartung haben als die in der GRV versicherten Personen. Die Maßnahme könnte somit eine Beitragssatzglättung erreichen. Langfristig würde sie jedoch zu einer stärkeren Belastung junger und zukünftiger Generationen von Beitragszahlenden führen.

Zudem hat die Einbeziehung von Beamtinnen und Beamten negative Effekte auf die öffentlichen Haushalte insbesondere von Bund und Ländern. Die Pensionen der vorhandenen Personen im Ruhestand müssen weiterhin aus dem Haushalt finanziert werden. Gleichzeitig müssen die Dienstherren für neu in die Rentenversicherung einbezogene Beamtinnen und Beamte Rentenversicherungsbeiträge entrichten. Außerdem müsste eine Zusatzversorgung neben der Rente aufgebaut und finanziert werden, um das derzeit durch das Grundgesetz abgesicherte Versorgungsniveau (Alimentationsprinzip) zu erhalten.

7.2.2.3 Erhöhung der Erwerbstätigenquote

Der Versichertenkreis kann auch durch eine Erhöhung der Erwerbstätigenquote (z. B. durch eine höhere Erwerbsbeteiligung von Frauen und Älteren) oder durch eine Erhöhung des Erwerbspersonenpotenzials (z. B. durch Zuwanderung) erweitert werden. Diese Instrumente sind jedoch schwieriger zu steuern als die Ausweitung des Kreises der Pflichtversicherten. Zur Steigerung der Erwerbsneigung von Frauen ist z. B. ein Bündel von arbeitsmarktpolitischen sowie gesellschafts- und familienpolitischen Maßnahmen (z. B. flexible Arbeitszeiten und Sicherstellung der Kinderbetreuung) erforderlich. Für eine höhere Erwerbsbeteiligung der älteren Menschen ist neben einer Anhebung der Altersgrenze z. B. eine alters- und gesundheitsgerechte Arbeitsplatzausstattung ein begünstigender Faktor. Die höhere Erwerbsneigung von Frauen und Älteren hat in den letzten Jahren erheblich zur verbesserten finanziellen Situation der GRV beigetragen.

7.3 Stellschraube Rentenniveau

7.3.1 Instrument - Rentenanpassungsformel

Das Instrument, mit dem das Rentenniveau hauptsächlich gesteuert werden kann, ist die Rentenanpassungsformel (Tnr. 3.2.2). Durch den erwarteten Beitragssatzanstieg und den aufgrund der demografischen Entwicklung zu erwartenden Anstieg des Rentnerquotienten werden sowohl der Beitragssatzfaktor als auch der Nachhaltigkeitsfaktor in den nächsten Jahren dämpfend auf die Rentensteigerungen wirken und damit das Rentenniveau reduzieren. Somit sind in der Rentenanpassungsformel bereits wirksame Elemente zur Steuerung des Rentenniveaus und damit zur finanziellen Stabilisierung der Rentenversicherung etabliert.

Sie sind dazu geeignet, die Beitragsbelastung zukünftiger Generationen zu mildern und den Beitragssatzanstieg vor allem in den 2020er und 2030er Jahren zu dämpfen (Beitragssatzglättung).

7.3.2 Instrument - Rentenniveau-Haltelinie

Reformvorschläge in der jüngsten Vergangenheit zielen darauf ab, die Wirkungen der Rentenanpassungsformel zu begrenzen. So wird ein dauerhaftes Mindestrentenniveau (Haltelinie) angestrebt.[94] Droht die Haltelinie überschritten zu werden, wird die Rentenanpassungsformel ausgesetzt und werden die Renten in dem Maße erhöht, dass das Rentenniveau die Haltelinie nicht unterschreitet.

[94] Zuletzt von der Kommission „Verlässlicher Generationenvertrag“ (2020).

Begründet wird eine Rentenniveau-Haltelinie damit, dass

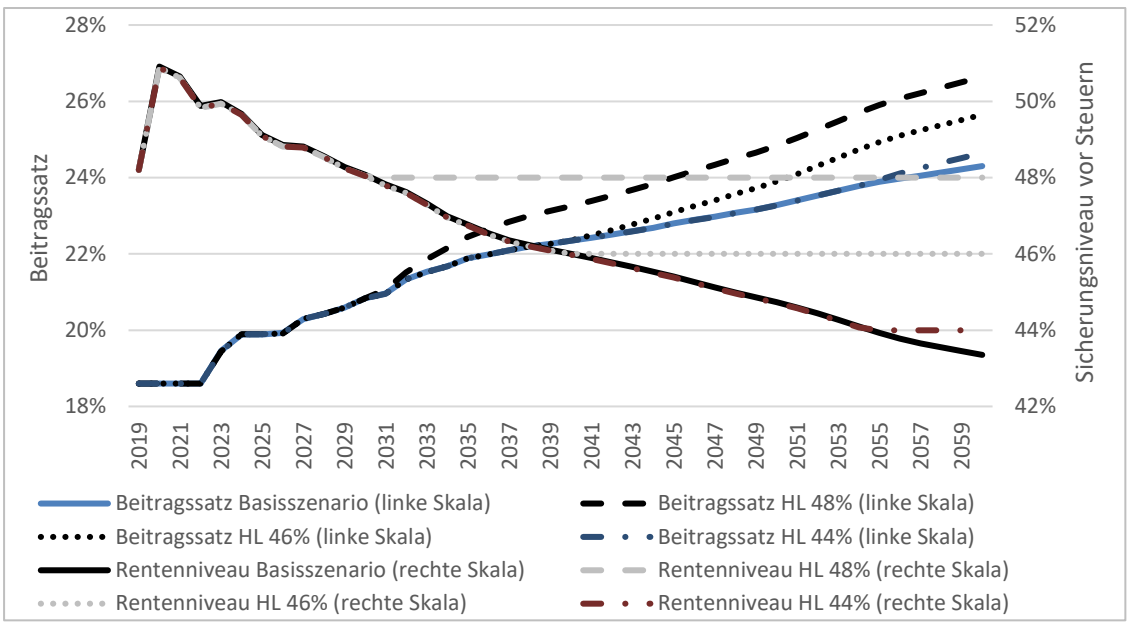
- bei einem zu niedrigen Rentenniveau der Abstand zur Grundsicherung im Alter verloren geht und somit eine beitragsfinanzierte Rentenversicherung ihre Legitimation verliert. Denn das Grundsicherungsniveau können Berechtigte auch ohne Beitragszahlung erreichen.
- bei einem zu niedrigen Rentenniveau die Lohnersatzfunktion der Rente verloren geht. Im Hintergrund steht die Frage, ob die Rente ein ausreichendes Versorgungsniveau bietet, sodass Rentenbeziehende mit dem Renteneinkommen ihren Lebensstandard halten können.

Bis zum Jahr 2025 ist derzeit die Rentenniveau-Haltelinie von 48 % gesetzlich vorgesehen. Wäre sie über das Jahr 2025 gültig, würde sie ab dem Jahr 2031 bindend wirken. Der Beitragssatz würde bis zum Jahr 2060 nicht auf 24,3 % steigen, sondern auf 26,6 %. Eine Haltelinie von 46 % würde im Jahr 2040 greifen und zu einem Beitragssatz von 25,7 % im Jahr 2060 führen. Eine Haltelinie von 44 % griffe im Jahr 2055, und der Beitragssatz im Jahr 2060 läge bei 24,6 % (vgl. Abbildung 26).

Abbildung 26

Je höher die Rentenniveau-Haltelinie, desto höher der Beitragssatz

Eine Rentenniveau-Haltelinie von 48 % greift im Jahr 2031 und würde zu einem Beitragssatz von 26,6 % im Jahr 2060 führen - 2,3 Prozentpunkte über dem Basisszenario. Die Haltelinie von 46 % hat einen Beitragssatz von 25,7 % und eine Halteline von 44 % einen Beitragssatz von 24,6 % im Jahr 2060 zur Folge. Die Abbildung zeigt die Entwicklung des Beitragssatzes und des Rentenniveaus in Prozent.



Grafik: Bundesrechnungshof.

Quelle für Grundzahlen: Werding (2021). HL: Haltelinie.

Neben den stärker steigenden Beitragssätzen werden in den hier betrachteten Szenarien außerdem höhere Zuschüsse an die GRV aus Haushaltsmitteln des Bundes erforderlich. Von 2,9 % des BIP im Jahr 2019 müssen diese bis zum Jahr 2060 – statt auf 4,0 % wie im Basisszenario – mit einer Haltelinie von 46 % auf 4,1 % und mit einer Haltelinie von 48 % auf 4,2 % des BIP steigen. Bei einer Haltelinie von 44 % sind die Bundesmittel im Vergleich zum Basisszenario dagegen kaum höher.

7.3.3 Instrument - Reaktivierung des Nachholfaktors

Wegen der Corona-Pandemie sind im Jahr 2020 die Bruttolöhne und -gehälter je Arbeitnehmer gesunken. Dies führte dazu, dass der für die Rentenanpassung des Jahres 2021 relevante Lohnfaktor ebenfalls gesunken ist und zu einer Rentensenkung geführt hätte. Dies verhinderte jedoch die sogenannte Rentengarantie. Sie sorgt dafür, dass es keine Rentensenkungen gibt. Die Renten blieben somit konstant (Nullanpassung), während die relevante Lohngröße gesunken ist. Dies führte im Jahr 2021 zu einem Anstieg des Rentenniveaus.

Bis zum Jahr 2018 sorgte der sogenannte Nachholfaktor dafür, dass die unterbliebenen Rentensenkungen derart nachgeholt wurden, dass zukünftige Rentensteigerungen geringer ausfielen, bis der Nachholbedarf abgebaut war. Den Nachholfaktor setzte der Gesetzgeber aber bis zum Jahr 2026 aus. Somit führt die Rentengarantie zusammen mit der Aussetzung des Nachholfaktors zu Mehrausgaben der GRV und tendenziell zu höheren Beitragssätzen. Eine Reformmöglichkeit in der GRV besteht deshalb darin, den Nachholfaktor im Jahr 2022 zu reaktivieren und die im Jahr 2021 unterbliebenen Rentensenkungen nachzuholen.[95]

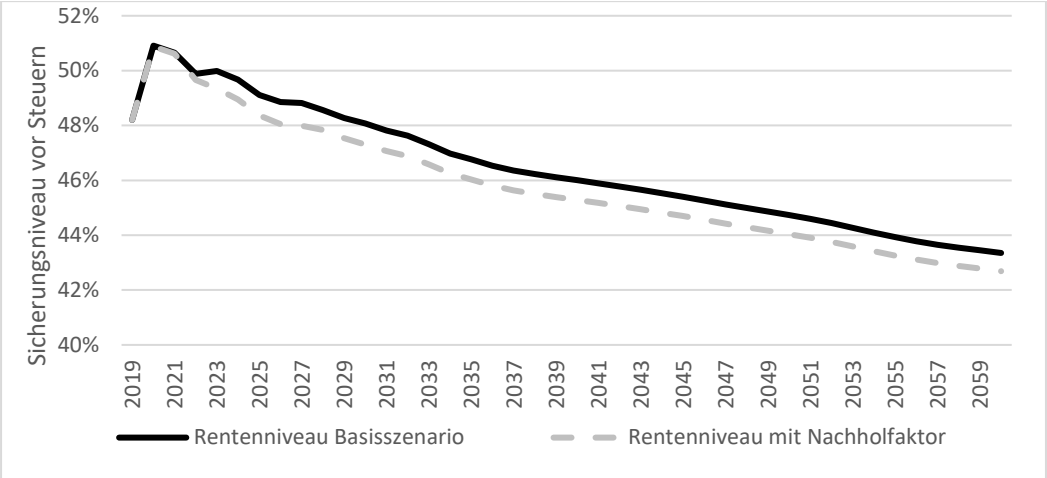
Wird der Nachholfaktor ab dem Jahr 2022 wieder angewendet, würde der Anstieg des Rentenniveaus aufgrund der Rentengarantie rückgängig gemacht. Das Rentenniveau verliefe auf einem niedrigeren Niveau als im Basisszenario und würde im Jahr 2060 42,7 % betragen (vgl. Abbildung 27). Im Gegenzug stiege der Beitragssatz der GRV ab dem Jahr 2023 langsamer an und würde im Jahr 2060 23,9 % erreichen (vgl. Abbildung 28). Er wäre damit um 0,4 Prozentpunkte niedriger.

[95] Vgl. zur Rentengarantie und zur Reaktivierung des Nachholfaktors: Bericht des Bundesrechnungshofes nach § 88 Absatz 2 BHO an das BMAS: Die finanzielle Entwicklung der gesetzlichen Rentenversicherung – Rentengarantie und Notwendigkeit des Nachholfaktors, Gz.: IX 2 - 2021 - 0053. https://www.bundesrechnungshof.de/de/veroeffentlichungen/produkte/beratungsberichte/2021/finanzielle-entwicklung-der-gesetzlichen-rentenversicherung-rentengarantie-und-notwendigkeit-eines-nachholfaktors.

Abbildung 27

Reaktivierung des Nachholfaktors im Jahr 2022 führt zu niedrigerem Rentenniveau

Holte man die im Jahr 2021 unterbliebene Rentensenkung nach, wäre im Jahr 2060 das Rentenniveau um 0,7 Prozentpunkte geringer. Die Abbildung zeigt das Rentenniveau in Prozent bis zum Jahr 2060.



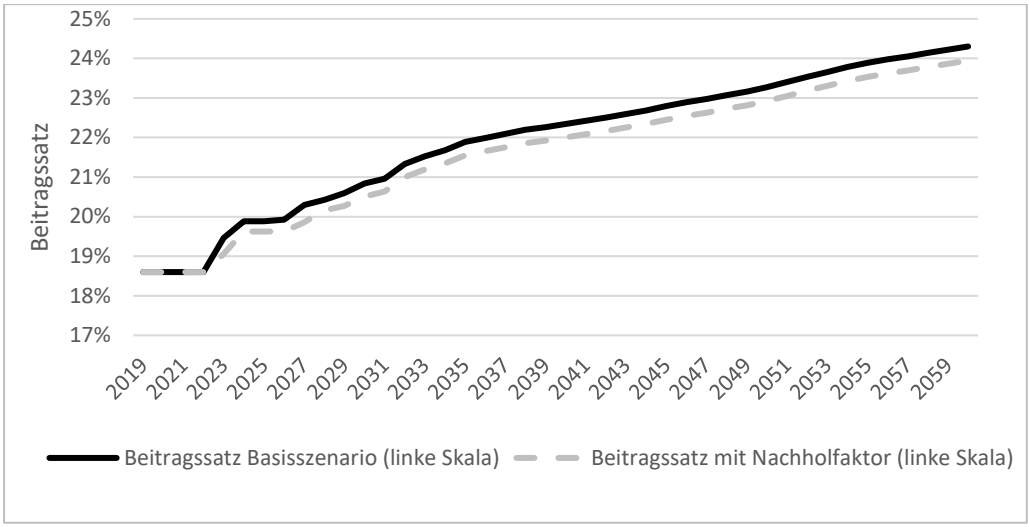
Grafik: Bundesrechnungshof.

Quelle für Grundzahlen: Werding, Läpple (2021).

Abbildung 28

Reaktivierung des Nachholfaktors im Jahr 2022 führt zu niedrigerem Beitragssatz

Holte man die im Jahr 2021 unterbliebene Rentensenkung nach, wäre im Jahr 2060 der Beitragssatz um 0,4 Prozentpunkte geringer. Die Abbildung zeigt den Beitragssatz in Prozent bis zum Jahr 2060.



Grafik: Bundesrechnungshof.

Quelle für Grundzahlen: Werding, Läpple (2021).

7.3.4 Instrument - Anhebung der Beitragsbemessungsgrenze

Auch die Anhebung der BBG wirkt über die Stellschraube Rentenniveau. Das durchschnittliche Rentenniveau sinkt, weil durch eine höhere BBG die durchschnittlichen beitragspflichtigen Löhne je Beschäftigten steigen. Zur Finanzierung der Ausgaben steht nun mehr beitragspflichtiges Einkommen zur Verfügung. Dies wirkt tendenziell senkend auf den Beitragssatz. Allerdings erwerben die Versicherten mit einem Einkommen über der BBG mit den höheren Beiträgen auch höhere Rentenansprüche (Äquivalenzprinzip), was mittel- und langfristig die Ausgaben erhöht. Der Beitragssatzeffekt verschwindet somit langfristig wieder.[96]

Nur wenn mit den höheren Beiträgen keine zusätzlichen Rentenansprüche verbunden wären, wäre ein langfristiger Effekt auf den Beitragssatz möglich.[97] In diesem Fall würde jedoch das Äquivalenzprinzip außer Kraft gesetzt.

7.4 Stellschraube Beitragssatz

Die gesetzlichen Regelungen sind derzeit so gestaltet, dass der Beitragssatz gleichsam automatisch angepasst wird, wenn Ausgaben und Einnahmen der GRV auseinanderlaufen. Da aus hohen Beitragssätzen eine Überbelastung der Beitragszahlenden und deren Arbeitgeber entstehen kann, ist es sinnvoll, den Beitragssatzanstieg zu begrenzen. Der Anstieg des Beitragssatzes im Basisszenario auf 24,3 % im Jahr 2060 könnte schon zu einer solchen Überbelastung beitragen, zumal die Beitragssätze in den anderen Sozialversicherungszweigen auch aufgrund der demografischen Entwicklung ansteigen werden. Deshalb hält es der Bundesrechnungshof für erforderlich, den Beitragssatzanstieg (auch) in der GRV zu dämpfen. Dies kann durch Maßnahmen geschehen, die an den anderen Stellschrauben ansetzen (Tnrn. 7.2 und 7.3).

7.4.1 Instrument - Beitragssatzhaltelinie

Eine Maßnahme, die direkt an der Stellschraube Beitragssatz ansetzt, wäre eine gesetzliche Beitragssatzobergrenze (Haltelinie), verknüpft mit einer Beitragssatzgarantie des Bundes. Sie würde zwar die Belastung der Beitragszahlenden begrenzen, die Last aber auf den Bund verlagern, der dann höhere Zuschüsse leisten müsste. Der Bund trüge letztlich das konjunkturelle und demografische Risiko der GRV alleine. Dabei besteht die Gefahr, dass der Bund überbelastet wird. Wenn der Bund diese Ausgabensteigerung bei der GRV nicht aus dem Haushalt gegenfinanzieren kann, stiege dadurch die Verschuldung des Bundes.

---

[96] Bei der derzeitigen Ausgestaltung der Rentenanpassungsformel würde sich der mit einer höheren BBG verbundene Anstieg der durchschnittlichen beitragspflichtigen Löhne schnell (im übernächsten Jahr) in einer höheren Rentenanpassung niederschlagen, so dass der kurzfristige Beitragssatzeffekt kaum vorhanden sein würde.

[97] Zusätzlich müsste der „Einmaleffekt“ auf die beitragspflichtigen Löhne im Lohnfaktor der Rentenanpassungsformel eliminiert werden, damit es nicht umgehend zu einem Anstieg der Renten kommt.

7.5 Ausbau der zusätzlichen Altersvorsorge

Mit der Rentenreform des Jahres 2001 führte der Gesetzgeber die sogenannte Riester-Rente ein. Sie sollte die durch diese Rentenreform in der GRV entstehende Rentenlücke durch eine zusätzliche individuelle und kapitalgedeckte Altersvorsorge schließen. Die Riester-Rente ist nicht obligatorisch. Sie wird aber steuerlich und mit Zulagen staatlich gefördert. Die festen, jährlichen Zulagen führen grundsätzlich zu einer mit dem Einkommen der Versicherten abnehmenden Förderquote.[98] Die steuerliche Förderung ergibt sich durch die nachgelagerte Besteuerung, bei der in der Erwerbsphase Altersvorsorgebeiträge die steuerliche Bemessungsgrundlage mindern. Aufgrund des progressiven Tarifs der Einkommensteuer nimmt der Steuervorteil mit dem Einkommen der Betroffenen zu.[99] Die volle Förderung erhält, wer – ausgehend von einem Mindesteigenbeitrag von 60 Euro im Jahr – einschließlich der erhaltenen Zulagen jährlich mindestens 4 % des im Vorjahr erzielten beitragspflichtigen Entgeltes bis zur Beitragsbemessungsgrenze der GRV in ein zertifiziertes Altersvorsorgeprodukt investiert.

Die Zahl abgeschlossener Riester-Verträge stieg in den Jahren 2004 bis 2010 stark an. Seit etwa dem Jahr 2015 stagniert sie annähernd bei insgesamt rund 16,5 Mio. Verträgen.[100] Es wird geschätzt, dass damit rund 44 % der Förderberechtigten Riester-Verträge abgeschlossen haben.[101] Die Zahl aktiv besparter und geförderter Verträge liegt mit zuletzt rund 10,5 Mio. deutlich darunter.[102] Als Gründe für diese Entwicklung werden oftmals eine große Intransparenz des Marktes sowie die niedrigen Effektivrenditen der Riester-Produkte gesehen.[103] Als intransparent gelten sowohl die Vorschriften zur Förderberechtigung als auch die Vielzahl angebotener Produkte, die für die Verbraucher nur schwer vergleichbar sind. Geringe Effektivrenditen resultieren zum einen aus hohen Vertriebs- und Verwaltungskosten für individuelle Verträge, zum anderen aus der Anlage der eingezahlten Mittel in vergleichsweise wenig rentierlicher Form, u. a. wegen gesetzlich geforderter Garantien, die die Anlagerisiken vermindern sollen. Im aktuellen Umfeld niedriger Zinsen hat sich das Renditeproblem weiter verschärft.

Daher wird hier ein Reformszenario mit reformierter, zusätzlicher Altersvorsorge betrachtet. Es unterscheidet sich vom derzeit geltenden Recht in folgenden Punkten:

---

98 Hinzu kommen besondere Zulagen, die von Anzahl und Alter etwaiger Kinder abhängen, aber ebenfalls als feste Beträge konzipiert sind und bei Beziehern niedriger Entgelte einen höheren Förderanteil erzeugen.

99 Umgekehrt werden entsprechend höhere Alterseinkommen in der Rentenphase steuerlich stärker belastet. Die Nettoeffekte in Abhängigkeit vom Einkommen in Erwerbs- und Rentenphase lassen sich dabei nicht leicht typisierend bestimmen. In der Erwerbsphase wird die Zulagenförderung mit den Effekten der steuerlichen Absetzung von Vorsorgeaufwendungen verrechnet, so dass effektiv die jeweils günstigere Form der Förderung zum Tragen kommt. Die Zulagen stellen auch für Geringverdiener eine – gemessen an den selbst aufzubringenden Vorsorgeersparnissen – relativ hohe Förderung sicher.

100 Vgl. Bundesministerium für Arbeit und Soziales 2021.

101 Vgl. Börsch-Supan, A., Bucher-Koenen, T., Goll, N. und Maier, C., 15 Jahre Riester – eine Bilanz, SVR-Arbeitspapier Nr. 12/2016.

102 Vgl. Bundesministerium der Finanzen 2020.

103 Vgl. Börsch-Supan et al. 2016, Seiten 50-64.

- Es wird unterstellt, dass die ergänzende Altersvorsorge mit Wirkung ab dem Jahr 2022 für alle sozialversicherungspflichtig Beschäftigten obligatorisch wird.
- Es wird ein Standardprodukt etabliert, das – z. B. nach dem Vorbild Schwedens[104] – sowohl von einem unabhängigen (quasi-)staatlichen Fonds angeboten werden kann als auch von konkurrierenden privaten Finanzdienstleistern.[105]
- Das staatliche Standardprodukt wird einer Person zugewiesen, wenn sie sich nicht für ein anderes Produkt entscheidet.
- Die Anlage konzentriert sich auf höher rentierliche Anlageformen, insbesondere in branchenmäßig und international breit gestreuten Aktienportfolien. Durch diese starke Streuung und die lange Anlagedauer zum Zweck der Altersvorsorge sind die Ertragsrisiken einer solchen Anlage relativ gering, trotz möglicherweise großer Kursschwankungen der einzelnen Aktien. Verluste eingezahlter Mittel sind daher auch ohne entsprechende Garantien kaum zu erwarten.[106]
- Die im langfristigen Mittel zu erwartenden Erträge solcher Aktienportfolios werden hier (nach Abzug von Kosten) mit nominal 5 % p. a. bzw. real 3,4 % p. a. angesetzt. Gemessen an Erfahrungswerten mit vergleichbaren Portfolien, die auch die teils heftigen Kurschwankungen an den Aktienmärkten im Rahmen der Finanz- und Wirtschaftskrise der Jahre 2008 und 2009 oder bei Ausbruch der Covid-19-Pandemie im Frühjahr 2020 berücksichtigen, erscheint diese Annahme eher moderat.[107]

Die staatliche Förderung wird in der bisherigen Form beibehalten.

Betrachtet werden hier drei Varianten, in denen die obligatorischen, jährlichen Einzahlungen (Sparquote) 4 %, 3 % oder 2 % der beitragspflichtigen Bruttoentgelte der sozialversicherungspflichtig Beschäftigten betragen.

[104] Vgl. Börsch-Supan et al. 2017.

[105] Unterscheiden könnten sich die verschiedenen Anbieter eines solchen, standardisierten Produktes immer noch bezüglich der effektiv erzielten Anlagerenditen und der anfallenden Verwaltungskosten.

[106] Vgl. Bucher-Koenen *et al*. 2019. Um temporäre Kursverluste auszugleichen, die auf Aktienmärkten immer wieder auf breiter Basis anfallen können, empfiehlt es sich außerdem, die für einzelne Anleger angesammelten Mittel gemäß einem „Lebenszyklusmodell" in den Jahren vor Erreichen des Rentenalters sukzessive stärker in festverzinsliche, allerdings weniger rentierliche Anlageformen umzuschichten (vgl. Börsch-Supan et al. 2017, S. 31-33, 71; Bucher-Koenen et al. 2019, insbes. Seiten 17-20).

[107] Zum Vergleich: Der Aktienfonds „MSCI World", der die Wertentwicklung von derzeit über 1 600 börsennotierten Unternehmen aus 23 Industrieländern abbildet, erzielte im Zeitraum der Jahre 1969 bis 2020 eine durchschnittliche Rendite von 7,9 % p.a. (MSCI 2021). Der im Rahmen des schwedischen Prämienrenten-Systems meist genutzte Mischfonds „AP7 Såfa", der von einer unabhängigen, staatlichen Agentur als „Default-Option" angeboten wird, weist in den Jahren 2003 bis 2020 eine Durchschnittsrendite von 9,9 % p.a. auf. Beträge, die dort bereits bei der Gründung des Fonds im Jahr 2000 angelegt wurden, erreichen wegen großer Kursverluste beim Abflauen des vorhergehenden „eCommerce-Booms" in den beiden Folgejahren insgesamt nur eine durchschnittliche Rendite von 5,5 % p.a. Bei Altersvorsorgeprodukten mit jährlichen Einzahlungen sind solche frühen Einbrüche für den Gesamtverlauf jedoch weniger bedeutsam, weil spätere Einlagen den angegebenen höheren Renditen unterlagen. Die Verwaltungskosten betragen in Schweden bezogen auf die angelegten Mittel derzeit rund 0,2 % p.a. (Swedish Pensions Agency, div. Jg.).

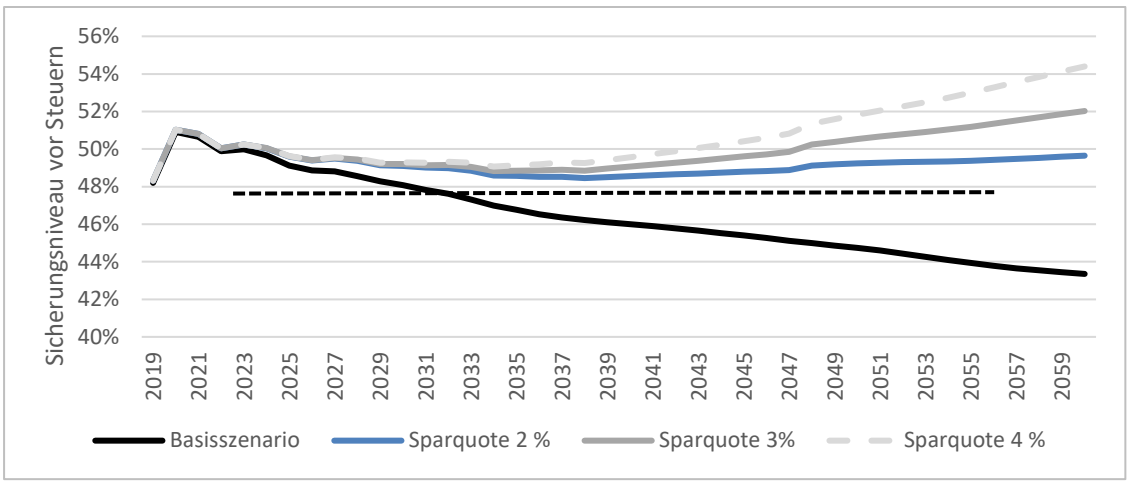
Es zeigt sich, dass das kombinierte Rentenniveau aus gesetzlichen Renten und zusätzlicher Altersvorsorge[108] bereits bei einer obligatorischen, jährlichen Sparquote von 2 % der jeweiligen Bruttoentgelte im Simulationszeitraum durchgängig über 48 % gehalten werden kann (vgl. Abbildung 29). Längerfristig entfalten sich die Effekte der kapitalgedeckten Vorsorge immer stärker, sodass das Rentenniveau bei einer Sparquote von 2 % ab dem Jahr 2040 wieder ansteigt und sich einem Rentenniveau von 50 % bis zum Jahr 2060 annähert. Bei Sparquoten von 3 % und 4 % liegt das kombinierte Rentenniveau langfristig sogar bei fast 52 % bzw. 55 %.

Dem langfristig deutlich erhöhten Rentenniveau stehen in den hier betrachteten Varianten ab dem Jahr 2022 verpflichtend anfallende Einzahlungen in die ergänzende Altersvorsorge in Höhe von 2 %, 3 % oder 4 % der beitragspflichtigen Bruttoentgelte gegenüber. Sie sind jedoch Ausdruck des (gewünschten) intertemporalen und intergenerativen Lastentransfers: stärkerer Belastung heutiger und geringerer Belastung zukünftiger Generationen.

Abbildung 29

Kombiniertes Rentenniveau aus gesetzlicher Rente und zusätzlicher obligatorischer Vorsorge liegt stets über 48 %

Die Abbildung zeigt das kombinierte Rentenniveau aus gesetzlicher Rente und zusätzlicher obligatorischer Altersvorsorge für unterschiedliche Sparquoten in Prozent.
Eine obligatorische Sparquote von 4 % des Bruttoentgelts hätte im Jahr 2060 ein kombiniertes Rentenniveau von 55 % zur Folge. Bei einer Sparquote von 3 % sind es 52 % und bei der Sparquote von 2 % fast 50 %.



Grafik: Bundesrechnungshof.

Quelle für Grundzahlen: Werding, Läpple (2021).

[108] Die Effekte der ergänzenden Vorsorge werden dabei so umgerechnet, dass sie konzeptionell mit den Angaben zum Sicherungsniveau der GRV vergleichbar sind.

8 Reformen in der gesetzlichen Krankenversicherung

8.1 Stellschrauben und Instrumente in der gesetzlichen Krankenversicherung

Um die Entwicklung des steigenden Beitragssatzes abzumildern, sind verschiedene Maßnahmen vorstellbar. Diese Maßnahmen können grundsätzlich an drei Stellschrauben ansetzen:

- der Versichertenstruktur,
- den durchschnittlichen beitragspflichtigen Einnahmen und
- den durchschnittlichen Ausgaben je Versicherten.

Im Folgenden werden einige dieser Maßnahmen betrachtet.

Systemwechsel z. B. zur Bürgerversicherung oder zu Pauschalbeiträgen (sog. Kopfpauschale) werden häufig im Zusammenhang mit Reformen der Sozialversicherungen diskutiert. Da jedoch ein grundsätzlicher Systemwechsel nicht Gegenstand dieses Berichts ist, werden nur einzelne Elemente wie die Erhöhung der Beitragsbemessungsgrenze oder die Einbeziehung anderer Einkunftsarten als Beitragsgrundlage sowie eine Ausweitung des Versichertenkreises dargestellt.

8.2 Stellschraube - Versichertenstruktur

Durch die Einbeziehung von Beamtinnen und Beamten, Selbstständigen und freiberuflich Tätigen wird die Versichertenstruktur beeinflusst. Da die neu einbezogenen Personen vermutlich im Durchschnitt mehr Beiträge zahlen als sie Ausgaben verursachen, könnte der GKV-Beitragssatz dadurch zunächst sinken. Durch die möglicherweise geringeren Risiken von Personen dieser Berufsgruppen könnten zudem die durchschnittlichen Leistungsausgaben sinken. Die durchschnittlichen beitragspflichtigen Einnahmen könnten steigen, da diese Personenkreise eher über ein überdurchschnittliches Einkommen verfügen.[109]

[109] Bertelsmann-Stiftung, online unter https://www.bertelsmann-stiftung.de/de/themen/aktuelle-meldungen/2020/februar/duales-system-kostet-mitglieder-der-gesetzlichen-krankenversicherung-bis-zu-145-euro-pro-jahr, Zugriff vom 31. August 2021.

8.2.1 Instrument - Erweiterung des Versichertenkreises um die Selbstständigen

Durch die Einführung einer Versicherungspflicht für Selbständige und freiberuflich Tätige in der GKV könnte der Beitragssatz zunächst sinken. Denn es ist davon auszugehen, dass dieser Personenkreis die Einnahmen erhöht, während die Ausgaben durch die günstige Altersstruktur zunächst nicht im gleichen Ausmaß steigen. Ob sich auch langfristig ein positiver Effekt auf den Beitragssatz ergibt, hängt z. B. von der Morbidität oder der Lebenserwartung dieses Personenkreises ab sowie von sozioökonomischen Aspekten wie dem Einkommen oder der Anzahl der mitversicherten Familienangehörigen.

8.2.2 Instrument – Erweiterung des Versichertenkreises um Beamtinnen und Beamte

Eine weitere Reformmöglichkeit bietet die Versicherungspflicht von Beamtinnen und Beamten und ihrer Familienangehörigen in der GKV. Bislang werden die Gesundheits- und Pflegekosten dieses Personenkreises grundsätzlich über die PKV und das System der Beihilfe gedeckt.

Erweitert man den Versichertenkreis um die Beamtinnen und Beamten, ergeben sich ähnliche Auswirkungen wie oben für die Selbstständigen beschrieben.

Für das Reformszenario wird unterstellt, dass alle Personen, die ab dem Jahr 2022 neu verbeamtet werden, in die GKV einbezogen werden. Personen, deren Beamtenverhältnis zu diesem Stichtag bereits besteht, verbleiben, gegebenenfalls mit ihren Familienangehörigen, im PKV-System und dem System der Beihilfe. Durch diese Reform stiege in der GKV zunächst die Anzahl der Personen, die mehr Beiträge zahlen als sie Leistungen verursachen. Das wirkt sich positiv auf die Beitragssatzentwicklung aus. Bis zum Jahr 2060 ließe sich so der Anstieg der GKV-Beitragssätze im Vergleich zum Basisszenario um einen Prozentpunkt auf 20,8 % verringern (vgl. Abbildung 30).

Allerdings ist damit zu rechnen, dass sich dieser Effekt längerfristig, d. h. über das Jahr 2060 hinaus, wieder zurückbildet. Bis zum Ende des Simulationszeitraumes im Jahr 2060 haben jedoch die ab dem Jahr 2022 neu verbeamteten Personen überwiegend noch nicht die späte Lebensphase erreicht, in der sie die Leistungen der GKV und SPV verstärkt beanspruchen. Würden sich Beamtinnen und Beamte und ihre Angehörigen nicht hinsichtlich risikorelevanter Merkmale von der bisherigen Versichertenpopulation der GKV unterscheiden, sollten die Effekte der Einbeziehung in diese Systeme nach dem Jahr 2060 wieder vollständig verschwinden. Untersuchungen zeigen allerdings, dass Beamte und ihre Angehörigen eine höhere Lebenserwartung aufweisen als der Bevölkerungsdurchschnitt.[110] Dies könnte die

[110] Statistisches Bundesamt; WISTA; 2017; Lebenserwartung von Beamtinnen und Beamten; Befunde und Auswirkungen auf künftige Versorgungsausgaben.

Leistungsausgaben und den Beitragssatz der GKV auf Dauer sogar noch weiter erhöhen, soweit dies nicht durch eine generell geringere Morbidität dieser Personengruppe aufgewogen wird.

Ein weiterer Aspekt ist die Entwicklung der Beihilfeausgaben. Da die Beihilfe im Reformszenario nur noch Beamtinnen und Beamten gewährt wird, die vor dem Jahr 2022 ernannt wurden, nehmen die dafür anfallenden Ausgaben nach dem Jahr 2030 deutlich ab. Im Gegenzug müssen die Dienstherren allerdings in wachsendem Maße Arbeitgeberanteile der Beiträge zur GKV entrichten. Gemessen am BIP wachsen die dafür anfallenden Ausgaben bis zum Jahr 2060 um 0,4 Prozentpunkte an, während sich die Ausgaben für die Beihilfezahlungen um rund 0,25 Prozentpunkte verringern. Der Unterschied ist darauf zurückzuführen, dass die einkommensbezogenen Beiträge zur GKV für neu verbeamtete Personen im Simulationszeitraum zunächst höher ausfallen als ihre tatsächlichen Gesundheits- und Pflegekosten, die für die jeweiligen Beihilfeausgaben maßgeblich wären.

Beim Übergang von einem Kostenerstattungssystem wie der Beihilfe, das jeweils einen Teil der tatsächlich anfallenden Kosten übernimmt, zu Beitragssystemen mit verschiedenen Elementen der Umverteilung sind solche Effekte unvermeidlich. Längerfristig sollten sich diese Effekte allerdings zurückbilden, da die Beihilfeausgaben nach der Reform auf Dauer auf null sinken. Die Dienstherren müssen dann nur noch die hälftigen Beiträge auf die geringeren Alterseinkünfte an die GKV entrichten. Bis zum Jahr 2060 wird eine solche Situation allerdings bei weitem noch nicht erreicht sein.

8.3 Stellschraube durchschnittliche beitragspflichtige Einnahmen

Durch eine Erhöhung der durchschnittlichen beitragspflichtigen Einnahmen je Mitglied könnte der Beitragssatz gesenkt werden, da die Beitragsgrundlage zur Finanzierung der Ausgaben größer ist. Dazu werden im Folgenden verschiedene Instrumente vorgestellt.

8.3.1 Instrument – Ausweitung der beitragspflichtigen Einnahmearten

Während die Ausgaben der GKV ähnlich gestiegen sind wie das BIP, sind die beitragspflichtigen Einnahmen der GKV-Mitglieder hingegen deutlich geringer gewachsen. Dies kann z. B. an einem Rückgang des Anteils des Lohneinkommens am Gesamteinkommen oder an den geringeren Erhöhungen der Renten liegen.[111]

Die Einbeziehung von weiteren Einkunftsarten wie zum Beispiel Einkünfte aus Vermietung und Verpachtung oder Kapitalerträge führt zu höheren beitragspflichtigen Einnahmen je

[111] Sozialpolitik aktuell, online unter http://www.sozialpolitik-aktuell.de/files/sozialpolitik-aktuell/_Politikfelder/Gesundheitswesen/Datensammlung/PDF-Dateien/abbVI52.pdf , Zugriff vom 5. Mai 2021.

Mitglied und somit zu niedrigeren Beitragssätzen. Dabei könnten für unterschiedliche Einkommensarten verschiedene oder einheitliche Beitragsbemessungsgrenzen festgelegt werden.[112]

Bisher gilt die Einbeziehung von weiteren Einkunftsarten als verwaltungsaufwendig, da Krankenkassen derzeit nicht digital auf Steuerdaten zugreifen können. Eine manuelle Erfassung, wie heute schon bei den Selbstständigen, wäre notwendig. Alternativ dazu könnte eine digitale Lösung der Beitragsermittlung mit Hilfe der Steuerbehörden angestrebt werden.

Des Weiteren besteht die Gefahr, dass bei einer Einbeziehung weiterer Einkunftsarten gesetzlich Versicherte vermehrt in die PKV abwandern.

8.3.2 Instrument – Erhöhung der Beitragsbemessungsgrenze

Ein weiterer Vorschlag, die Beitragsgrundlage der GKV zu verbreitern, ist die Anhebung der Beitragsbemessungsgrenze. Bei einer höheren Beitragsbemessungsgrenze steigen die durchschnittlichen beitragspflichtigen Einnahmen. Dem gegenüber stehen nur geringere Mehrausgaben, z. B. aufgrund höherer Krankengeldzahlungen. Der Beitragssatz könnte gesenkt werden.

8.3.3 Instrument – Beitragszuschläge bei vorzeitigem Eintritt in die Rente

Höhere beitragspflichtige Einnahmen je Mitglied können auch erreicht werden, wenn Versicherte Beitragszuschläge bei einem vorzeitigen Eintritt in die Rente leisten müssten. Gehen Versicherte vorzeitig aus einer Beschäftigung in Rente, muss beim Renteneintritt nicht nur die GRV auf Beitragseinnahmen verzichten. Auch an die GKV und SPV zahlen diese Versicherten als Frührentner geringere Beiträge auf die geringeren Alterseinkünfte. In der GRV werden bei vorzeitigem Renteneintritt grundsätzlich Rentenabschläge erhoben. Für die GRV entsteht damit – sofern die Rentenabschläge ausreichend hoch sind – langfristig durch einen vorgezogenen Renteneintritt kein finanzieller Nachteil. Dies gilt jedoch nicht für die GKV. Im Gegenteil: Zum einen werden vorzeitig nur Beiträge auf die im Vergleich zum Arbeitsentgelt niedrigeren Alterseinkünfte entrichtet. Darüber hinaus sind die Alterseinkünfte durch den vorzeitigen Renteneintritt wegen der Abschläge etwas niedriger als bei „normalem“ Renteneintritt. Entsprechend sind die Beiträge an die GKV geringer, da die GKV-Beiträge nur auf die niedrigeren Alterseinkünfte entrichten werden. Um diese beiden negativen Effekte für die Beitragseinnahmen der GKV zu vermeiden, könnten vorzeitige Altersrentner verpflichtet werden, den Nachteil für die GKV zu kompensieren. Dies kann z. B. durch eine Einmalzahlung an die GKV bei Renteneintritt oder einen Beitragszuschlag geschehen. In der GRV besteht heute schon die Möglichkeit, durch eine Zahlung die Abschläge zu vermeiden. Hier besteht

[112] Wissenschaftliches Institut der AOK, 2013, Die Krankenversicherung der Zukunft - Anforderungen an ein leistungsfähiges System, Seite 142.

die Möglichkeit, dass die Arbeitgeber für ihre Beschäftigten die Zahlung leisten, wenn sie vorzeitig das Unternehmen verlassen.

Folgendes Reformszenario wäre z. B. vorstellbar und praktikabel: GKV-Versicherte, die eine gesetzliche Altersrente mit Abschlägen beziehen, müssen die GKV-Beiträge nicht auf die Rente mit Abschlägen zahlen, sondern auf die Rente ohne Abschläge. Damit fallen für diese Versicherten die beitragspflichtigen Einnahmen und damit die Beiträge höher aus. Durch diesen Beitragszuschlag für vorzeitige Altersrentner könnte der GKV-Beitragssatz bis zum Jahr 2060 um 0,3 Prozentpunkte geringer ausfallen (vgl. Abbildung 30).[113]

8.4 Stellschraube Ausgaben je Versicherten

8.4.1 Instrument – Begrenzung der Ausgaben

Der Beitragssatzanstieg könnte begrenzt werden, wenn die durchschnittlichen Ausgaben je Versicherten reduziert werden könnten. Eine Möglichkeit, dies zu erreichen, besteht darin, den Leistungskatalog der GKV einzuschränken. Solche Maßnahmen sind jedoch schwierig durchzusetzen und werfen neben Fragen nach der sozialen Gerechtigkeit weitreichende ethische Fragen auf. Effizienzsteigerungen bei den Leistungserbringern z. B. durch Digitalisierung und Kostensenkung bei der Versorgungsstruktur können dagegen die durchschnittlichen Ausgaben je Versicherten positiv beeinflussen, ohne dass dadurch das Leistungsniveau und die Qualität sinken müssen.

8.4.2 Instrument – Zuzahlungen

Durch Zuzahlung beteiligen sich die Versicherten an der Finanzierung der für sie erbrachten Leistungen. Die Ausgaben der GKV fallen dann geringer aus (vgl. Tnr. 4.2.5). Dies reduziert in der GKV die durchschnittlichen Ausgaben je Versicherten und wirkt sich positiv auf die Beitragssatzentwicklung aus.

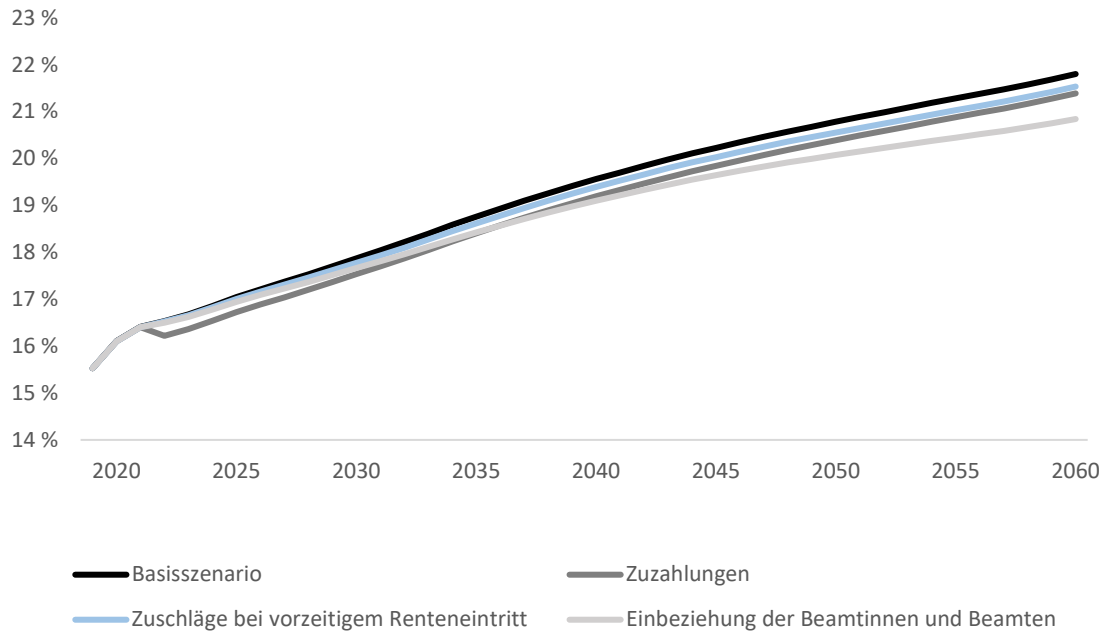
Als Reformszenario wird hier beispielhaft eine Verdopplung der bisher zu leistenden Zuzahlungen von 4,2 Mrd. Euro auf 8,4 Mrd. Euro betrachtet. Dies würde umgehend zu einer Verringerung des Beitragssatzes um 0,4 Prozentpunkte führen. Dieser Effekt bleibt in dieser Stärke bis zum Jahr 2060 erhalten. Die Umsetzung dieser Maßnahme könnte den Beitragssatzanstieg auf 21,4 % der beitragspflichtigen Einnahmen der Mitglieder begrenzen, gegenüber einem Beitragssatz von 21,8 % im Basisszenario ohne eine Erhöhung der Zuzahlungen (vgl. Abbildung 30).

[113] Nicht berücksichtigt sind dabei Verhaltensänderungen, dass die Versicherten etwa den Renteneintritt wegen des Zuschlags zeitlich verschieben.

Abbildung 30

Reformen können Beitragssatzanstieg in der gesetzlichen Krankenversicherung bremsen

Durch die verschiedenen Reformansätze könnte der Beitragssatz im Vergleich zum Basisszenario gesenkt werden. Die Abbildung zeigt den Beitragssatz in Prozent bis zum Jahr 2060.



Grafik: Bundesrechnungshof.

Quelle für Grundzahlen: Werding, Läpple (2021).

9 Reformen in der Pflegeversicherung

9.1 Stellschrauben und Instrumente in der sozialen Pflegeversicherung

Die Reformziele einer intergenerativen Lastenumverteilung zugunsten der jüngeren und zukünftigen Generationen, die stärkere Beteiligung der Menschen im Seniorenalter an den demografischen Lasten und die Beitragssatzglättung lassen sich auch in der SPV erreichen. Möglich ist dies über eine Beeinflussung der genannten Stellschrauben Leistungsniveau und Leistungsbeziehendenquotient.

9.2 Stellschraube Leistungsbeziehendenquotient

Der Leistungsbeziehendenquotient gibt die zahlenmäßige Relation der Leistungsbeziehenden (Pflegebedürftigen) zu den Beitragszahlenden in der SPV an (vgl. Tnr. 5.6.2). Durch die starke Zunahme der Anzahl der pflegebedürftigen Menschen und die sinkende Zahl der Beitragszahlenden wird dieser Quotient stark steigen und zu einem Beitragssatzanstieg in der SPV führen. Dem können Instrumente bzw. Reformen entgegenwirken, die den Leistungsbeziehendenquotient günstig beeinflussen. Dies sind im Wesentlichen:

- Ausweitung des Versichertenkreises und
- Reduktion der Wahrscheinlichkeit, pflegebedürftig zu werden.

9.2.1 Instrument - Ausweitung des Versichertenkreises

Durch die Ausweitung des Versichertenkreises steigt die Anzahl der Beitragszahlenden, während die Anzahl der Pflegebedürftigen zunächst weitgehend unbeeinflusst bleibt. Dies eröffnet die Möglichkeit, zumindest vorübergehend, den Beitragssatzanstieg zu bremsen. Eine positive langfristige Wirkung auf den Beitragssatz ist davon abhängig, wie die sozioökonomischen Eigenschaften und die Morbidität des neu einbezogenen Personenkreises im Vergleich zu den bisher in der SPV versicherten Personen sind. Vor allem spielen die (altersspezifischen) Pflegewahrscheinlichkeiten des neuen Versichertenkreises im Vergleich zu denen der „Altversicherten“ eine Rolle.

Eine Möglichkeit, den Versichertenkreis zu erweitern, besteht darin, die neu eingestellten Beamtinnen und Beamten ab dem Jahr 2023 in die SPV einzubeziehen. Genauso wie in der GKV (Tnr. 8.2.2) führt dies zu einem niedrigeren Beitragssatzpfad zumindest bis zum Jahr 2060. Der SPV-Beitragssatz ist im Jahr 2060 um 0,2 Prozentpunkte geringer als im Basisszenario. Allerdings ist genauso wie in der GKV damit zu rechnen, dass sich diese Effekte längerfristig wieder zurückbilden oder sogar ins Gegenteil verkehren.

9.2.2 Instrument - Reduktion der Wahrscheinlichkeit, pflegebedürftig zu werden

Der Beitragssatzanstieg kann auch durch Maßnahmen gebremst werden, die die altersspezifischen Pflegewahrscheinlichkeiten senken bzw. den Grad der Pflegebedürftigkeit positiv beeinflussen. Ist der Grad der Pflegebedürftigkeit nicht so hoch oder wird die Pflegebedürftigkeit zeitlich verschoben bzw. gänzlich vermieden, entstehen Ausgaben für die SPV erst gar nicht. Dies kann z. B. durch Prävention oder durch den medizinisch-technischen Fortschritt geschehen.

9.3 Stellschraube Leistungsniveau

Das Leistungsniveau der SPV wird hier an der Relation der durchschnittlichen Ausgaben für einen Pflegebedürftigen zum durchschnittlichen beitragspflichtigen Einkommen der Beitragszahlenden gemessen (Tnr. 5.5). Steigen die Durchschnittseinkommen stärker als die durchschnittlichen Pflegeausgaben, wirkt dies positiv auf die Beitragssatzentwicklung.

Dies kann z. B. erreicht werden durch:

- Einbeziehung weiterer Einkommensarten (z. B. Vermögenseinkommen) in die Beitragsbemessungsgrundlage,
- Erhöhung der beitragspflichtigen Einnahmen bestimmter Personengruppen oder
- Erhöhung oder Aufhebung der Beitragsbemessungsgrenze.

Bei allen drei Maßnahmen steigt das durchschnittliche beitragspflichtige Einkommen an. Zur Finanzierung der Leistungsausgaben steht ein größeres Finanzvolumen zur Verfügung. Dies macht es möglich, den Beitragssatz zu senken. Die Ausführungen zu diesen Maßnahmen im Bereich der GKV gelten auch für die SPV.

9.3.1 Instrument - Erhöhung des beitragspflichtigen Einkommens bestimmter Personengruppen

Als Beispiel für die Erhöhung des beitragspflichtigen Einkommens bestimmter Personengruppen wird das schon für die GKV behandelte Reformszenario „höhere Beiträge für Frührentner“ betrachtet (siehe Tnr. 8.3.3). Rentenbeziehende, die eine mit Abschlägen versehene Rente erhalten, zahlen dabei auf die Bruttorente ohne Abschläge Beiträge an die SPV. Dies bedeutet, dass die betroffenen Personen zusätzliche Beiträge auf den jeweiligen Gegenwert ihrer Abschläge zahlen.

Eine solche Reform führt bis zum Jahr 2060 zu einem um 0,05 Prozentpunkte geringeren Beitragssatzanstieg in der SPV.

9.3.2 Instrument – Änderung der Leistungspauschalen

Das Leistungsniveau kann politisch gut durch die Festlegung der Leistungspauschalen beeinflusst und damit der Beitragssatz gesteuert werden. Blieben z. B. die Leistungspauschalen nominal konstant (Aufhebung der Dynamisierung), sänke das Leistungsniveau. Eine Beitragssatzsenkung wäre möglich. Dieser Reformvorschlag wird als „kalte Abschaffung der Pflegeversicherung“ bezeichnet. Tatsächlich würde dadurch der Versicherungscharakter der SPV allmählich verloren gehen.

Aktuelle Reformschritte gehen in eine andere Richtung. So sieht das Gesetz zur Weiterentwicklung der Gesundheitsversorgung (GVWG)[114] vor, dass Pflegebedürftige in der stationären Pflege einen prozentualen Leistungszuschlag auf ihre Eigenbeteiligung erhalten.[115] Dies führt letztlich zu einem Anstieg des Leistungsniveaus und beschleunigt den Beitragssatzanstieg. Entsprechend sieht das GVWG eine Erhöhung des Beitragssatzes für Kinderlose um 0,1 Prozentpunkte auf 0,35 % vor.[116]

9.4 Ausbau des Pflegevorsorgefonds

Durch Kapitalbildung können Lasten zukünftiger Generationen auf die heutigen Generationen verlagert werden. Dies leistet in der SPV der Pflegevorsorgefonds. Indem die heutigen Beitragszahler mit höheren Beiträgen in den Fonds einzahlen, können die Beitragszahler ab dem Jahr 2035 mit der Auflösung des Fonds von Beiträgen entlastet werden. Derzeit kann durch den Fonds nur eine geringe Entlastung zukünftiger Beitragszahler erreicht werden (vgl. Tnr. 5.4). Durch eine Reform könnte das Fondsvolumen vergrößert werden. Betrachtet werden folgende Reformszenarien:

Der Pflegevorsorgefonds wird so ausgestattet und dafür der Beitragssatz ab dem Jahr 2022 so erhöht, dass ab dem Jahr 2035 bis zum Jahr 2060 ein allgemeiner Beitragssatz von 3 % gehalten wird.

Der Pflegevorsorgefonds wird so ausgestattet und dafür der Beitragssatz im Jahr 2022 so angehoben, dass er bis zum Jahr 2060 konstant ist.

Die Ausgabenentwicklung in der SPV bleibt von den Reformszenarien unberührt. Beide Reformszenarien zielen darauf ab, heutige Generationen, vor allem die derzeit noch in der Erwerbsphase befindliche „Babyboomer-Generation“, stärker zu belasten und die jüngeren Generationen zu entlasten.

Um den Beitragssatz ab dem Jahr 2022 zu stabilisieren, muss der rechnerische Beitragssatz (Durchschnittswert aus allgemeinem Beitragssatz und Beitragszuschlag für Kinderlose) im Jahr 2022 auf 3,6 % steigen. Er bleibt dann bis zum Ende des Simulationszeitraums im Jahr 2060 auf diesem Niveau konstant (vgl. Abbildung 31). Um den allgemeinen Beitragssatz (ohne Berücksichtigung des Zusatzbeitragssatzes für Kinderlose) im gesamten Zeitraum der Jahre 2035 bis 2060 bei 3 % zu stabilisieren, muss er in den Jahren von 2022 bis 2034

[114] Bundesgesetzblatt Jahrgang 2021 Teil I, Nr. 44, 19. Juli 2021, S. 2754-2804.

[115] § 43c SGB XI sieht vor, dass Pflegebedürftige der Pflegegrade 2 bis 5, die bis zu zwölf Monaten vollstationäre Leistungen beziehen, einen Leistungszuschlag in Höhe von 5 % ihres zu zahlenden pflegebedingten Eigenanteils erhalten. Pflegebedürftige, die seit mehr als zwölf (24) Monaten vollstationäre Leistungen beziehen, erhalten einen Leistungszuschlag in Höhe von 25 (45) % ihres zu zahlenden pflegebedingten Eigenanteils. Ab 36 Monaten vollstationärer Leistungen beträgt der Zuschlag 70 %.

[116] Der geänderte § 55 SGB XI sieht vor, dass der Beitragszuschlag von 0,25 % auf 0,35 % steigt. Die Rechtsänderungen des GVWG sind in den Simulationsrechnungen nicht abgebildet, da diese den Rechtstand bis März 2021 berücksichtigen.

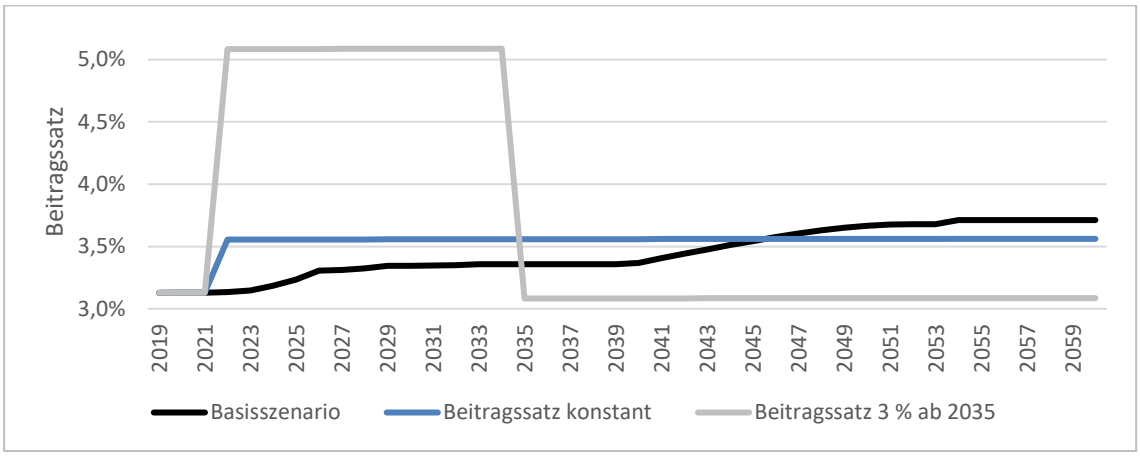
deutlich auf 5,0 % angehoben werden. Der rechnerische Beitragssatz unter Berücksichtigung des Zusatzbeitragssatzes für Kinderlose steigt dann auf 5,1 %, ab dem Jahr 2035 liegt er konstant bei 3,1 % (vgl. Abbildung 31). Das erste Reformszenario bewirkt eine einfache Glättung des für das Basisszenario projizierten Verlaufs des Beitragssatzes der SPV. Das zweite Reformszenario erfordert dagegen temporär eine deutliche Beitragssatzerhöhung.[117]

Während sich der Mittelbestand des Pflegevorsorgefonds im Basisszenario bis zum Jahr 2034 auf einen Höchstwert von fast 35 Mrd. Euro bzw. 0,7 % des laufenden BIP erhöht, muss er zur Stabilisierung des allgemeinen SPV-Beitragssatzes auf 3 % ab dem Jahr 2035 auf 487 Mrd. Euro bzw. 10,7 % des BIP ausgebaut werden. Für einen ab dem Jahr 2022 durchgängig konstanten Beitragssatz ergibt sich im Jahr 2034 ein Höchstbestand der Mittel im Pflegevorsorgefonds von 132 Mrd. Euro bzw. 2,8 % des BIP.

Abbildung 31

Pflegevorsorgefonds ermöglicht Beitragssatzsenkung für zukünftige Beitragszahlende

Mit einem Anstieg des **Beitragssatzes** ab dem Jahr 2022 auf 5,1 % wird das erforderliche Fondsvolumen aufgebaut, das eine Beitragssatzsenkung ab dem Jahr 2035 auf 3,1 % erlaubt. Ein durchweg konstanter Beitragssatz wird erreicht, wenn er ab dem Jahr 2022 auf 3,6 % steigt. Die Abbildung zeigt die Entwicklung des rechnerischen SPV-Beitragssatzes (inklusive Zusatzbeitrag für Kinderlose) für die beiden Reformszenarien und das Basisszenario.



Grafik: Bundesrechnungshof.

Quelle für Grundzahlen: Werding, Läpple (2021).

Das **Fondsvolumen** muss auf 487 Mrd. Euro bis zum Jahr 2034 aufgebaut werden, damit eine Beitragssatzsenkung auf 3,1 % ab dem Jahr 2035 möglich ist. Das Fondsvolumen für einen ab

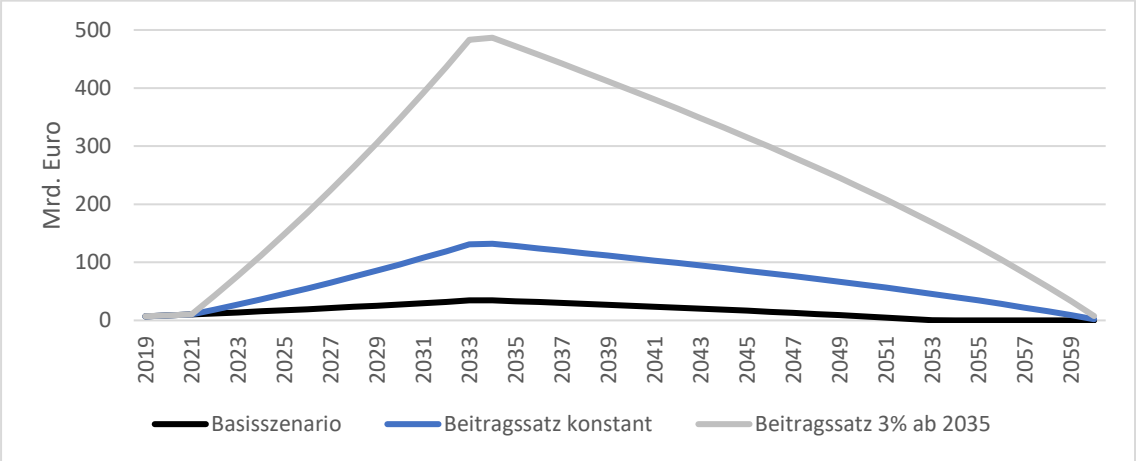
[117] Wegen der plangemäßen Erschöpfung der Mittel im Pflegevorsorgefonds im Jahr 2060 muss der Beitragssatz der SPV unmittelbar anschließend prinzipiell wieder auf das für das Basisszenario zum Ende des Simulationszeitraums ermittelte Niveau ansteigen. Längerfristig bewirkt jedoch das sinkende Leistungsniveau wegen der hier getroffenen Annahmen zur jährlichen Fortschreibung der Pflegeleistungen (mit Hilfe eines Mischindex aus Lohnwachstum und Inflationsrate) eine leichte Beitragssatzsenkung.

dem Jahr 2022 konstanten Beitragssatz muss auf 132 Mrd. Euro im Jahr 2034 anwachsen (vgl. Abbildung 32). Die Abbildung zeigt die Entwicklung des Volumens des Pflegevorsorgefonds in Mrd. Euro für die beiden Reformszenarien und das Basisszenario.

Abbildung 32

Volumen des Pflegevorsorgefonds ist umso größer, je stärker die Beitragssatzsenkung für zukünftige Beitragszahlende sein soll

Das Fondsvolumen muss auf 487 Mrd. Euro steigen, um eine Beitragssatzsenkung ab dem Jahr 2035 auf 3,1 % zu ermöglichen. Für einen durchweg konstanten Beitragssatz muss das Fondsvolumen bis auf 132 Mrd. Euro im Jahr 2034 wachsen. Die Abbildung zeigt das Fondsvolumen in Mrd. Euro bis zum Jahr 2060.



Grafik: Bundesrechnungshof.

Quelle für Grundzahlen: Werding, Läpple (2021).

10 Kombinierte Effekte der einzelnen Reformelemente

Auf Grundlage der prognostizierten Entwicklung der finanziellen Situation der Sozialversicherungszweige und den zuvor im Einzelnen betrachteten Reformelementen lassen sich die Auswirkungen auf die finanzielle Situation der Sozialversicherungszweige und den Bund in einem kombinierten Reformszenario abbilden. Dabei wurden folgende Reformelemente kombiniert:

- Einführung einer dynamischen Regelaltersgrenze in der GRV,
- Einbeziehung selbstständig werdender Personen und aller Selbstständigen (unter 40 Jahren) ohne obligatorische Altersvorsorge in die GRV,
- Mindestrentenniveau in der GRV von 44 % (Rentenniveau-Haltelinie) ab dem Jahr 2026,
- Einführung einer obligatorischen, ergänzenden Altersvorsorge für alle sozialversicherungspflichtig Beschäftigten (mit einer jährlichen Sparquote von 3 % der Bruttoentgelte),

- Wiedereinsetzung des „Nachholfaktors“ ab dem Jahr 2022,
- Erhöhung der Zuzahlungen Versicherter zu Leistungen der GKV,
- Erhebung höherer Beiträge in GKV und SPV von Beziehenden einer vorzeitigen Altersrente zur Kompensation von Beitragsausfällen durch vorzeitige Renteneintritte,
- Einbeziehung neu ernannter Beamtinnen und Beamter in GKV und SPV,
- Ausbau des Pflegevorsorgefonds zur Stabilisierung des SPV-Beitragssatzes (ab dem Jahr 2022).
- Keine Beitragssatz-Haltelinie und keine Beitragssatzgarantie in der GRV über das Jahr 2025 hinaus.

10.1 Auswirkungen der Reformelemente auf die gesetzliche Rentenversicherung

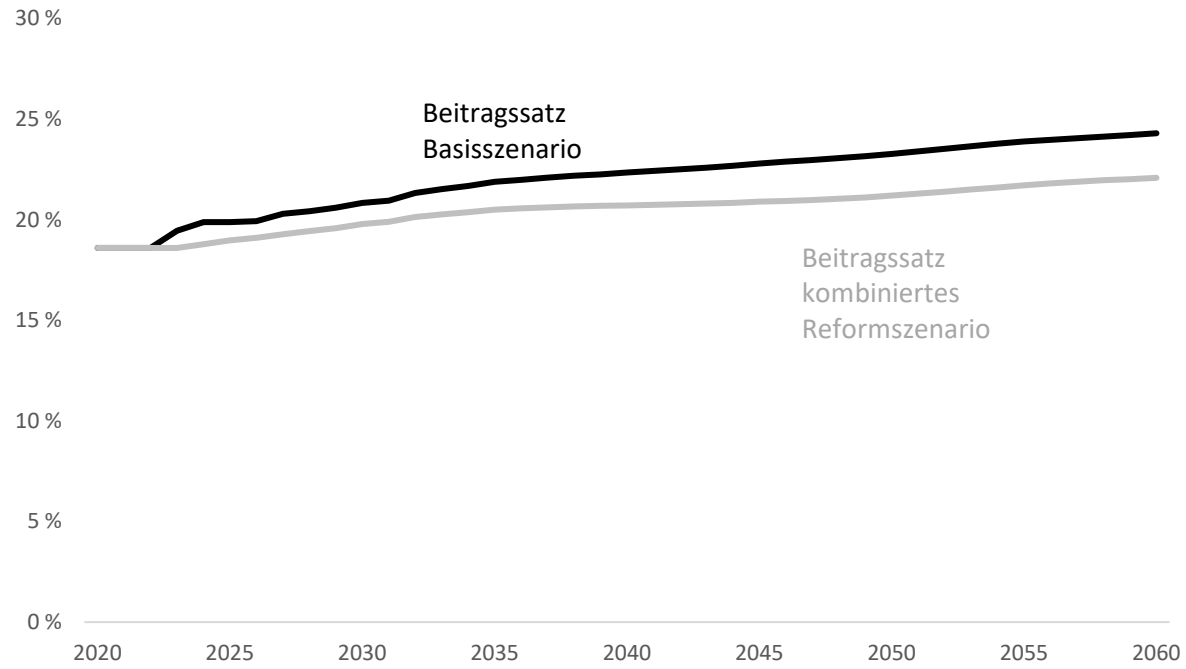
Die dargestellten kombinierten Reformelemente wirken sich insgesamt günstig auf das Sicherungsniveau und den Beitragssatz der GRV aus (vgl. Abbildung 33 und 34). Bis zum Jahr 2037 sinkt das Sicherungsniveau der GRV etwas stärker als im Basisszenario. Ab dem Jahr 2047 fällt das Sicherungsniveau unter 46 % und erreicht bis zum Jahr 2060 44,6 % – und nicht 43,4 % wie im Basisszenario. Damit kommt die Rentenniveau-Haltelinie von 44 % nicht zur Anwendung.

Von besonderer Bedeutung für das Rentenniveau ist die Einführung einer obligatorischen ergänzenden Altersvorsorge auf kapitalgedeckter Basis für alle sozialversicherungspflichtig Beschäftigten. Dadurch entwickelt sich das kombinierte Rentenniveau aus gesetzlicher Rente und privater Zusatzvorsorge bereits ab dem Jahr 2027 günstiger als im Basisszenario: Bis zum Jahr 2034 bleibt das Sicherungsniveau von GRV und ergänzender Vorsorge bei rund 48,5 % stabil und beginnt dann sogar wieder anzusteigen. Im Jahr 2060 ergibt sich so ein kombiniertes Sicherungsniveau von 53,4 %.

Abbildung 33

Beitragssatz der gesetzlichen Rentenversicherung bis zum Jahr 2060 im kombinierten Reformszenario deutlich niedriger

Die Abbildung zeigt, dass sich die kombinierten Reformelemente in den beiden Reformszenarien insgesamt günstig auf den Beitragssatz der GRV auswirken. Die Abbildung zeigt den Beitragssatz in der GRV bis zum Jahr 2060 in Prozent.



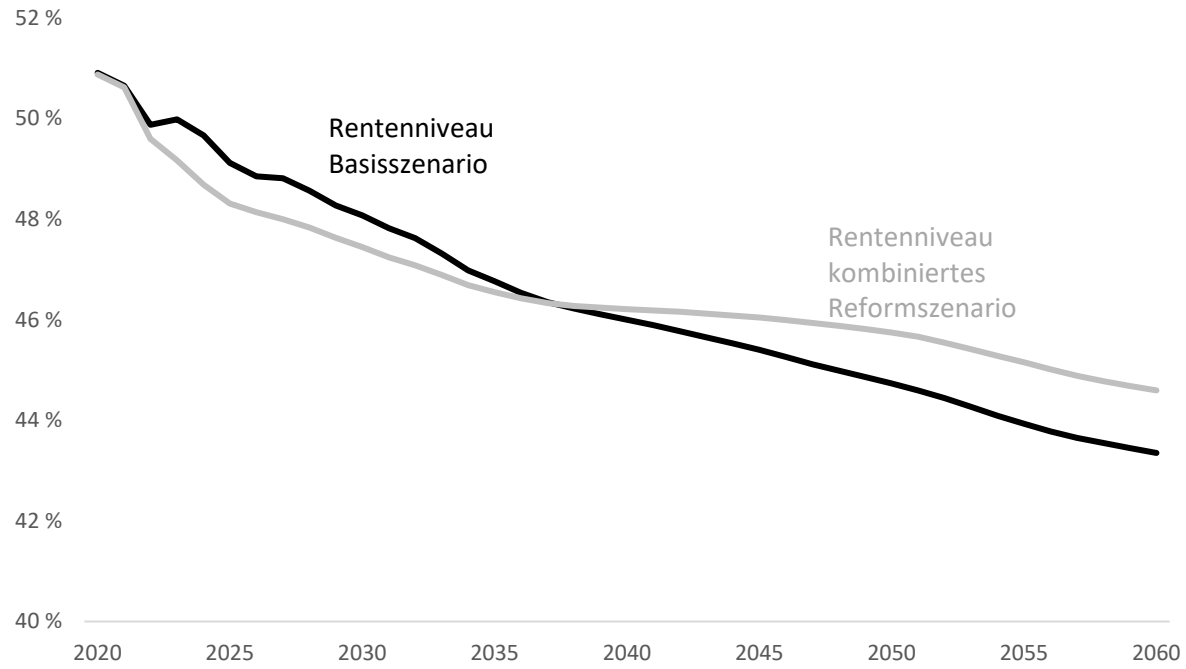
Grafik: Bundesrechnungshof.

Quelle für Grundzahlen: Werding, Läpple (2021).

Abbildung 34

Rentenniveau der gesetzlichen Rentenversicherung ist im kombinierten Reformszenario langfristig höher

Die kombinierten Reformelemente führen ab dem Jahr 2037 zu einem höheren Rentenniveau als im Basisszenario. Die Abbildung zeigt das Sicherungsniveau vor Steuern (Rentenniveau) in Prozent bis zum Jahr 2060 im Basisszenario und im kombinierten Reformszenario.



Grafik: Bundesrechnungshof.

Quelle für Grundzahlen: Werding, Läpple (2021).

Durch die Kombination der Reformelemente steigt der Beitragssatz der GRV bis zum Jahr 2060 auf 22,1 % statt auf 24,2 % im Basisszenario. Die Reformelemente wirken sich dabei folgendermaßen aus: Die Wiedereinsetzung des Nachholfaktors hat auf Sicherungsniveau und GRV-Beitragssatz rasch einsetzende, dauerhafte, aber relativ geringe Effekte. Die Einbeziehung Selbstständiger ohne obligatorische Altersvorsorge in die GRV entfaltet ebenfalls rasch Wirkungen, die sich bis in die 2040er Jahre verstärken, anschließend allerdings wieder zurückbilden. Die Einführung einer dynamischen Regelaltersgrenze wirkt ab dem Jahr 2030 und verstärkt sich zunehmend, sodass sie bis zum Jahr 2060 die größten Effekte auf den Beitragssatz und das Rentenniveau hat. Auch die Einführung einer obligatorischen, ergänzenden Altersvorsorge benötigt einige Zeit, bevor ihre Effekte für das kombinierte Sicherungsniveau ausreichend hervortreten.

10.2 Auswirkungen der Reformelemente auf die gesetzliche Krankenversicherung und die soziale Pflegeversicherung

In der GKV bewirken die Reformelemente zusammen genommen einen deutlich niedrigeren GKV-Beitragssatz: Im Basisszenario steigt er bis zum Jahr 2060 auf 21,8 %, mit den Reformelementen hingegen lediglich auf 19,5 % und ist damit um 2,3 Prozentpunkte niedriger (vgl. Abbildung 35).

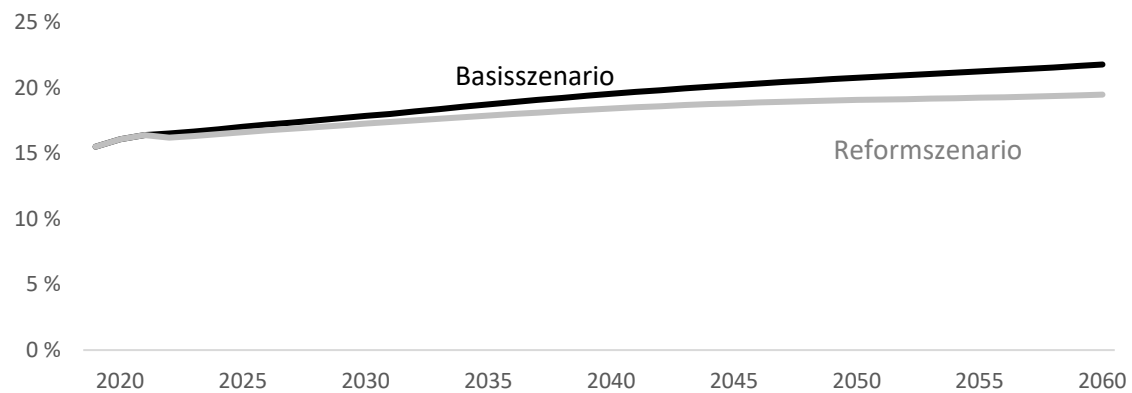
Die erhöhten Zuzahlungen für GKV-Leistungen tragen dazu 0,4 Prozentpunkte bei. Die Effekte höherer Beitragszahlungen von Personen, die vorzeitig in Rente eingetreten sind, betragen 0,3 Prozentpunkte. Die Einbeziehung neu ernannter Beamter in die GKV hat bis zum Jahr 2060 mit einem Prozentpunkt die größten Auswirkungen. Sobald ein wachsender Anteil dieses Personenkreises die Altersphase erreicht, in der die Inanspruchnahme von Leistungen der GKV typischerweise stark ansteigt, dürften sich die Effekte jedoch wieder zurückbilden. Die Beitragssatzeffekte der beiden zuvor genannten Maßnahmen sind dagegen dauerhaft.

Ferner wirken sich die Reformen im Bereich der GRV auch auf die GKV aus. Das höhere Renteneintrittsalter und das höhere Sicherungsniveau beeinflussen die Beitragszahlungen positiv. Dies dämpft den Anstieg des GKV-Beitragssatzes zusätzlich.

Abbildung 35

Der Beitragssatzanstieg in der gesetzlichen Krankenversicherung kann durch Reformen gebremst werden

Der Beitragssatz (in Prozent der beitragspflichtigen Einnahmen) würde im Reformszenario bis zum Jahr 2060 auf 19,5 % steigen. Im Basisszenario betrug er im Jahr 2060 21,8 %. Die Abbildung zeigt den Beitragssatz in der GKV in Prozent bis zum Jahr 2060.



Grafik: Bundesrechnungshof.

Quelle für Grundzahlen Werding, Läpple (2021).

In der SPV ergeben sich ähnliche Effekte wie in der GKV. Im Basisszenario bremsen die höheren Beitragszahlungen von Frührentnern zusammen mit der Einbeziehung von Beamtinnen

und Beamten in die SPV den Anstieg des Beitragssatzes leicht. Dieser Effekt wird nach dem Jahr 2060 allerdings wieder zurückgehen. Der Ausbau des Pflegevorsorgefonds, durch den der SPV-Beitragssatz bis zum Jahr 2060 konstant gehalten werden kann, wirkt sich dabei zusätzlich günstig aus. Die Kombination der drei Reformelemente in der SPV ergibt, dass der ab dem Jahr 2022 erforderliche Beitragssatz bei 3,4 % liegt.

Als Nebeneffekt steigen durch diese Maßnahmen die Zuschüsse des Bundes für die einzelnen Sozialversicherungszweige bis zum Jahr 2060 im kombinierten Reformszenario weniger stark als im Basisszenario – auf 4,5 % des BIP anstatt auf 4,9 % des BIP.

10.3 Auswirkungen auf den Gesamtsozialversicherungsbeitragssatz und den Bundeshaushalt

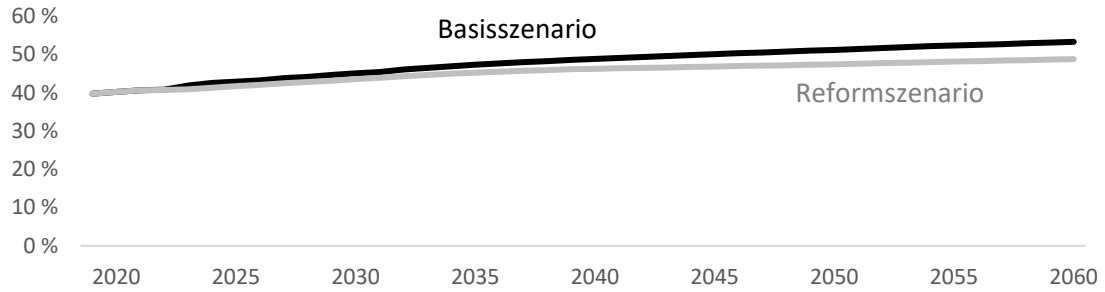
Werden alle dargestellten Reformelemente kombiniert, wird der Anstieg des Gesamtsozialversicherungsbeitragssatzes deutlich gedämpft (vgl. Abbildung 36).[118] Im Basisszenario steigt er bis zum Jahr 2060 auf 53,3 %. Die hier dargestellten Reformelemente verringern den Anstieg deutlich um 4,5 Prozentpunkte, sodass er nur auf 48,8 % steigt. Auf die Ausgabenquote wirken sich die kombinierten Reformelemente ebenfalls aus: Sie erhöht sich bis zum Jahr 2060 nicht auf 29,3 %, sondern auf 28,8 % des BIP. Die eher moderate Verringerung der Ausgabenquote um 0,5 Prozentpunkte ist insbesondere darauf zurückzuführen, dass durch die Einbeziehung der Selbstständigen und der Beamtinnen und Beamten in GKV und SPV die Ausgaben in diesen Umlagesystemen langfristig steigen.

[118] Die Ausgaben zur Arbeitslosenversicherung, für die Beamtenversorgung und die Beihilfen zu Gesundheits- und Pflegeausgaben für Beamtinnen und Beamte sind nicht Gegenstand dieses Berichtes. Sie werden zur Vervollständigung der Gesamtentwicklung in den sozialen Sicherungssystemen dargestellt.

Abbildung 36

Reformen bremsen den Anstieg des Gesamtsozialversicherungsbeitragssatzes in den Jahren 2019 bis 2060 deutlich

Die Beitragssätze zur GRV, GKV und SPV sowie zur Arbeitslosenversicherung steigen insgesamt durch die Reformen nicht auf 53,3 %, sondern auf 48,8 %. Die Abbildung zeigt die Summe der Beitragssätze der Sozialversicherungszweige in Prozent.



Grafik: Bundesrechnungshof.

Quelle für Grundzahlen: Werding, Läpple (2021).

Die kombinierten Reformelemente wirken sich auch auf den gesamtstaatlichen Finanzierungssaldo und den Schuldenstand günstig aus. Sie bremsen den Anstieg des Finanzierungssaldos:[119] Er steigt auf 7,9 % des BIP, anstatt 8,8 % im Basisszenario. Ähnlich entwickelt sich die gesamtstaatliche Schuldenstandsquote. Im Basisszenario nimmt sie bis zum Jahr 2060 auf 141,6 % des BIP zu. Die Reformelemente verringern den Anstieg um über 13 Prozentpunkte auf 128,5 % des BIP.

Insgesamt kann das skizzierte Reformbündel die finanzielle Situation der Sozialversicherungszweige und auch des Bundes deutlich verbessern. Klar wird aber auch, dass diese Reformen die finanziellen Lasten aufgrund der demografischen Entwicklung nicht beseitigen können. Sie sorgen jedoch für eine Milderung und eine andere – gleichmäßigere – Verteilung zwischen den Generationen.

Auch mit den Reformelementen kann der Gesamtsozialversicherungsbeitragssatz nicht unter 48 % gedrückt werden. Auf den Beitragssatz der GKV wirken sich die Reformelemente am deutlichsten aus. Der Wert des Tragfähigkeitsindikators S2, der im Basisszenario bei 5,1 % des BIP liegt, beträgt mit den kombinierten Reformelementen immer noch 4,6 % des BIP. Ein Konsolidierungsbedarf bleibt bestehen. Der gesamtstaatliche Primärsaldo[120] müsste also sofort und dauerhaft um 4,6 % des BIP verbessert werden. Rechnerisch müsste dazu die Steuerquote – ausgehend vom Wert für das Jahr 2019 in Höhe von 23,2 % – auf 27,7 % des BIP steigen. Ohne Maßnahmen zur Konsolidierung würde die Verschuldung auf 128,5 % des BIP steigen (vgl. Abbildung 37).

[119] Finanzierungssaldo und Schuldenstandsquote berechnet unter der Annahme, dass alle Einnahmequoten konstant bleiben und Defizite durch zusätzliche Schulden ausgeglichen werden.

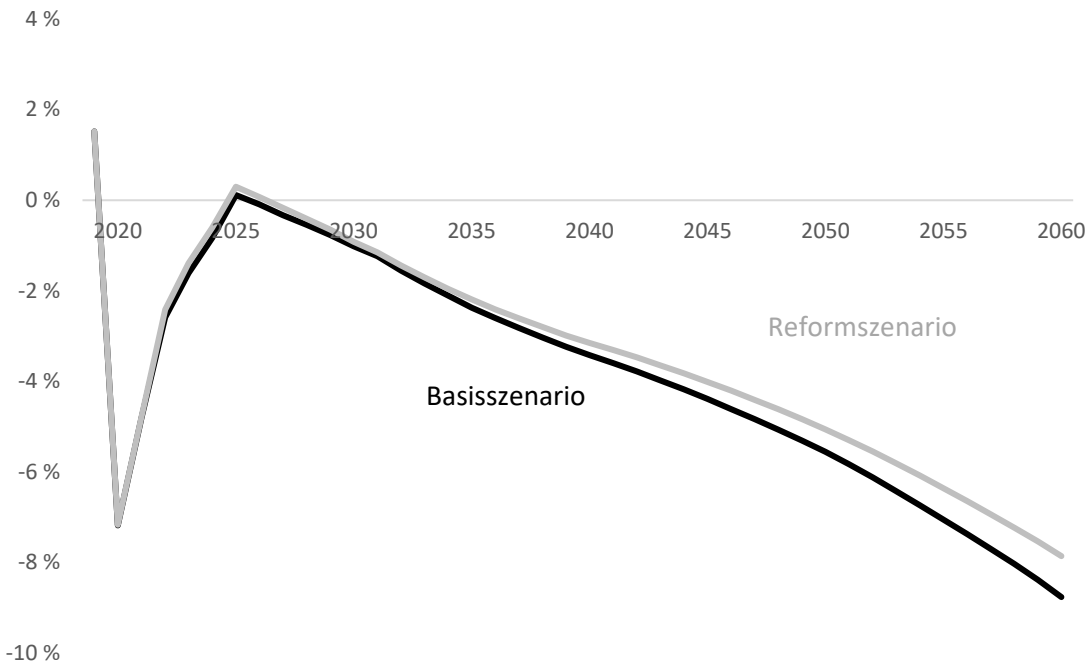
[120] Zum Begriff Primärsaldo siehe Tz. 6.5.

Abbildung 37

Reformelemente wirken sich positiv auf die gesamtstaatlichen öffentlichen Finanzen aus

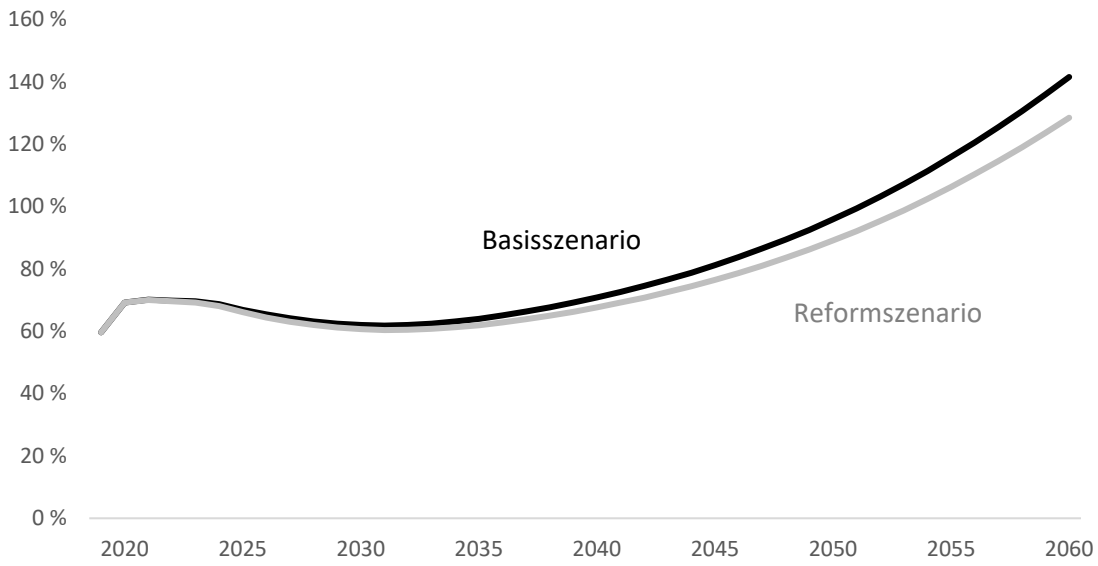
a) Finanzierungssaldo

Die kombinierten Reformelemente wirken sich günstig auf den gesamtstaatlichen Finanzierungssaldo aus. Sie bremsen den Anstieg des Finanzierungssaldos. Er steigt bis zum Jahr 2060 auf 7,9 % des BIP, anstatt auf 8,8 % im Basisszenario. Die Abbildung zeigt den Finanzierungssaldo in Prozent des BIP bis zum Jahr 2060 unter der Annahme, dass die Einnahmequoten konstant bleiben und Defizite durch zusätzliche Schulden gedeckt werden.



b) Schuldenstand

Wenn der Gesamtsozialversicherungsbeitragssatz bis zum Jahr 2060 bei 40 % konstant bliebe, würde sich die gesamtstaatliche Schuldenstandsquote im Reformszenario auf 128,5 % des BIP erhöhen. Im Basisszenario läge sie bei knapp 142 % des BIP. Die Abbildung zeigt die Schuldenstandsquote (Maastricht-Definition) in Prozent des BIP bis zum Jahr 2060 unter der Annahme, dass die Einnahmequoten konstant bleiben und Defizite durch zusätzliche Schulden gedeckt werden.



Grafik: Bundesrechnungshof.

Quellen: Prof. Werding, Läpple (2021).

11 Würdigung und Empfehlung

11.1 Reformen wegen demografischer Situation unausweichlich

Die Projektion der finanziellen Situation der Sozialversicherungszweige aufgrund der demografischen Entwicklung in Deutschland bis zum Jahr 2060 zeigt deutlich, dass die Ausgaben in der GRV, der GKV und der SPV stark steigen werden. Diese Ausgaben muss im Umlageverfahren im Wesentlichen eine schrumpfende Erwerbsbevölkerung finanzieren. Standen im Basisjahr 2019 noch einer Person im Seniorenalter mehr als drei Personen im Erwerbsalter gegenüber, sind es im Jahr 2060 nur noch eine zu zwei. Bereits Mitte der 2020er-Jahre wird sich die Auswirkung des Alterungsprozesses verstärken, wenn die Generation der „Babyboomer“ nach und nach das Seniorenalter erreicht.

Die vom Bundesrechnungshof betrachtete „mittlere“ Variante mit moderaten Annahmen zur Entwicklung von Demografie, Arbeitsmarkt und Wirtschaft führt voraussichtlich zu einem Anstieg des Gesamtsozialversicherungsbeitragssatzes von derzeit rund 40,0 % auf 53,3 % im

Jahr 2060. Die Bundeszuschüsse werden, gerechnet in Preisen von 2015, voraussichtlich von 113,5 Mrd. Euro im Jahr 2020 (3,5 % des BIP) auf 229,4 Mrd. Euro im Jahr 2060 steigen (4,9 % des BIP). Würde die politische Maßgabe umgesetzt werden, nach der der Gesamtsozialversicherungsbeitragssatz bei 40,0 % bleiben soll („Sozialgarantie“) und der Bund die Ausgabensteigerungen finanziert, müssten sich die Bundesmittel bis zum Jahr 2060 in Preisen von 2015 auf 473,1 Mrd. Euro erhöhen (10,1 % des BIP). Soweit diese Ausgaben nicht über Umschichtungen im Bundeshaushalt oder zusätzliche Einnahmen finanziert werden, würde die Verschuldung des Bundes im Jahr 2060 auf 141,6 % des BIP steigen.[121] Die derzeitigen gesetzlichen Vorgaben in der GRV führen zudem dazu, dass das Rentenniveau bis zum Jahr 2060 auf 43,4 % sinken wird.

Veränderungen der Annahmen wie niedrigere Geburtenrate, höhere Lebenserwartung, höheres Produktivitätswachstum oder lohnorientierte Anpassungen der Pflegeleistungen haben nur begrenzte Auswirkungen auf die Höhe des Gesamtsozialversicherungsbeitragssatzes, des Bundeszuschusses oder auf die Höhe der Verschuldung des Bundes. Besonders hervorzuheben ist, dass ein höheres Produktivitätswachstum, das häufig als Lösung der demografisch bedingten Finanzprobleme angesehen wird, unter den derzeitigen rechtlichen Rahmenbedingungen (einkommensorientierte Rentenanpassungsformel, Leistungen der GKV mit hohem Arbeitskostenanteil und an der Lohnentwicklung orientierte Anhebung der Leistungspauschalen in der SPV) höhere Ausgaben bei den Sozialversicherungszweigen nach sich zieht. Ein Produktivitätswachstum kann nicht dazu beitragen, die finanziellen Herausforderungen der demografischen Entwicklung zu lösen. Es mildert die Probleme aber ab, da sich eine wohlhabendere Gesellschaft z. B. höhere Sozialversicherungsbeitragssätze eher leisten kann. Erheblich geringere Beitragssätze in den Sozialversicherungszweigen können jedoch dadurch nicht erreicht werden.

11.2 Exemplarische Reformszenarien verbessern die finanzielle Lage

Die vom Bundesrechnungshof exemplarisch analysierten Reformszenarien enthalten im Hinblick auf die festgelegten Ziele folgende Maßnahmen:

- Das zeitliche Vorziehen von Beitragsbelastungen (obligatorische private Altersvorsorge in der GRV und Erhöhung der Zuführung zum Pflegevorsorgefonds) zugunsten einer geringeren Belastung in der Zukunft (intergenerativer Lastentransfer).

[121] Dabei ist die Annahme konstanter Beitragssätze als Näherung für die tatsächlich für die Berechnung der Schuldenstandsquoten unterliegenden Annahme konstanter Einnahmequoten zu verstehen. Zusätzlich wird unterstellt, dass auch ein überproportionaler Anstieg der steuerfinanzierten Sozialausgaben für die Beamtenversorgung, die Beihilfe und für einen Teil der Arbeitsmarktmaßnahmen (v. a. Grundsicherungsleistungen) durch zusätzliche Verschuldung finanziert wird. Diese zusätzliche Annahme gilt im Folgenden immer, wenn die Entwicklung der Schuldenstandsquote beschrieben wird.

- Eine Beitragssatzglättung durch die kurzfristige Erhöhung der Einnahmen[122] aufgrund der Einbeziehung von bisher nicht Versicherten in die Sozialversicherungssysteme (Einbeziehung Selbstständiger in die GRV und Einbeziehung von neu ernannten Beamtinnen und Beamten in die GKV und SPV), bevor mittel- und langfristig auch die Ausgaben steigen.
- Leistungskürzungen (höheres Renteneintrittsalter) und höhere Zuzahlungen in der GKV. Damit werden Menschen im Seniorenalter stärker an den demografischen Lasten beteiligt und die Beitragszahlenden entlastet.
- Einen Ausgleich von Nachteilen für die Sozialversicherungen („Nachholfaktor“ für Rentenanpassungen und Kompensationsleistungen für die GKV und die SPV für Beitragsausfälle durch vorzeitige Renteneintritte). Damit werden die Beitragszahlenden entlastet.
- Die Einführung einer „Haltelinie“ für das Sicherungsniveau in der GRV und
- eine direkte Belastungsbegrenzung für den Bund, da die Beitragssatz-Haltelinie und die Beitragssatzgarantie in der GRV nicht fortgeführt werden.

Die Umsetzung aller Maßnahmen könnte den Anstieg des Gesamtsozialversicherungsbeitragssatzes um 4,5 Prozentpunkte dämpfen (48,8 % anstatt 53,3 % im Basisszenario). Die Bundeszuschüsse an die Sozialversicherungen würden sich bis zum Jahr 2060 auf 4,5 % des BIP, statt auf 4,9 % des BIP im Basisszenario, erhöhen. Würde der Gesamtsozialversicherungsbeitragssatz konstant bei 40 % gehalten werden und der Bund die Ausgabensteigerung durch eine höhere Verschuldung finanzieren, stiege die Verschuldung bis zum Jahr 2060 auf 128 % des BIP statt auf 141,6 % des BIP im Basisszenario.

Die exemplarisch betrachteten Reformmaßnahmen führen auch dazu, dass sich die langfristige Tragfähigkeit der Finanzen der Sozialversicherungszweige und die gesamtstaatlichen öffentlichen Finanzen verbessern. Die Tragfähigkeitslücke liegt im Basisszenario bei 5,1 % und sinkt durch die analysierten Reformen auf 4,6 % des BIP. Um die Finanzen dauerhaft zu konsolidieren, wäre es erforderlich, den jährlichen Primärüberschuss um 4,6 % des BIP zu verbessern und nicht um 5,1 % des BIP wie im Basisszenario. Die Steuerquote müsste z. B. dafür um 19,7 % anstatt um 23,2 % steigen.

Das vom Bundesrechnungshof analysierte Maßnahmenpaket erscheint als ein ausgewogener Beitrag zur Stabilisierung der finanziellen Situation der Sozialversicherung, bei dem das Rentenniveau sogar höher liegen kann.

Die untersuchten Reformansätze machen deutlich, dass der Handlungsspielraum für eine Begrenzung des Gesamtsozialversicherungsbeitragssatzes oder der Verschuldung des Bundes eher gering ist. Leistungskürzungen können die finanzielle Situation der Sozialversicherungen stabilisieren. Sie sind aber politisch nur im begrenzten Umfang möglich, ohne die soziale Gerechtigkeit infrage zu stellen. So ist die umstrittene und unbeliebte dynamische Heraufsetzung der Regelaltersgrenze bei steigender Lebenserwartung beispielsweise ein wichtiges Instrument, um die Finanzen der Sozialversicherungszweige zu stabilisieren. Dabei soll nur ein Teil der längeren Rentenbezugszeiten aufgrund der höheren Lebenserwartung durch eine

[122] Ausweitung des Umlagesystems, wobei höhere Einnahmen und höhere Ausgaben zeitlich auseinanderfallen und somit eine Beitragssatzglättung erreicht werden kann.

längere Erwerbsphase kompensiert werden. Bereits dies dämpft deutlich den Anstieg des Beitragssatzes der GRV. Durch die höheren Beiträge während der (längeren) Erwerbsphase werden aber auch die Finanzen der GKV und der SPV in erheblichem Umfang stabilisiert. Wirtschaftlich leistungsfähige Versicherte haben die Möglichkeit, mit Abschlägen bei der Rente auch vorzeitig in die Rente zu gehen. Dann sollten aber auch die Beitragsausfälle bei der GKV und der SPV ausgeglichen werden.

Der Bundesrechnungshof hat bei der Analyse der Reformansätze auch im Blick, dass das Rentenniveau nach der derzeitigen Systematik sinken wird. Dies führt zu einem Rentenniveau, das eine adäquate Versorgung im Alter gefährdet und die Akzeptanz des Systems schwächt. Daher hat er die Effekte von „Haltelinien“ für das Rentenniveau berechnet. Diese „Haltelinien“ haben aber auch eine negative Seite: Der Beitragssatz und/oder der Bundeszuschuss müssen steigen. Letzteres erhöht ggf. die Verschuldung des Bundes. Schon die vorgeschlagenen Reformen innerhalb der umlagefinanzierten Systeme ohne private Zusatzversicherung erhöhen das Rentenniveau auf über 44 %, sodass eine Haltelinie von 44 % nicht greifen müsste. Mit Einführung einer obligatorischen ergänzenden Altersvorsorge kann ein Versorgungsniveau erreicht werden, dass durchgängig über 48 % liegt. Die kapitalgedeckte ergänzende Altersvorsorge ist dabei ein Instrument, durch das die betreffenden Personen ihre Alterssicherung selbst tragen und nicht jüngere oder zukünftige Generationen wie im Umlagesystem. Die Versicherten sparen in der Erwerbsphase einen Kapitalstock an, der in der Rentenphase abgebaut wird und das Versorgungsniveau stabilisiert. Die Erwerbsbevölkerung muss diese Rentenleistungen nicht durch Beiträge finanzieren, wodurch ihre Belastung durch den Gesamtsozialversicherungsbeitragssatz niedriger ist.

Den gleichen Ansatz verfolgt die Erhöhung der Zuführung zum Pflegevorsorgefonds. Ein besser ausgestatteter Pflegevorsorgefonds könnte als kapitalgedecktes Teilsystem so konzipiert werden, dass ein konstanter Beitragssatz in der SPV in den Jahren 2022 bis 2060 gewährleistet wird. Er kann aber auch so konzipiert werden, dass in der Phase von hohen Steigerungen des Gesamtsozialversicherungsbeitragssatzes der Beitragssatz der SPV in den Jahr 2035 bis 2060 auf einem niedrigeren Niveau gehalten werden kann.

Die Einbeziehung Selbstständiger in die GRV und die Einbeziehung neu ernannter Beamter in die GKV und die SPV erweitern die Umlagesysteme, wodurch sich die Tragfähigkeitslücke vergrößert. Sie sind aber geeignet, die Beitragssätze vorübergehend zu mindern (Beitragssatzglättung), indem sie in einer Phase von hohen Beitragssatzsteigerungen junge Beitragszahler in das Sozialversicherungssystem integrieren. Diese neuen Beitragszahler zahlen Beiträge und nehmen die Leistungen (Rente, Gesundheits- und Pflegeleistungen) weitestgehend erst (viel) später in Anspruch. Damit ergibt sich ein positiver – wenngleich nur vorübergehender – Finanzierungseffekt. Eine langfristige Beitragssatzsenkung tritt dadurch nicht ein, weil die neu einbezogenen Personen später höhere Ausgaben verursachen.

Das analysierte Reformpaket wird zudem noch durch die Wiedereinsetzung des „Nachholfaktors“ für Rentenanpassungen und eine Erhöhung der Zuzahlungen Versicherter zu Leistungen der GKV flankiert. Diese beiden Maßnahmen führen zu einer nicht unerheblichen Entlastung bei den Sozialversicherungsbeiträgen und des Bundes. Sie tragen zudem dazu bei, die Tragfähigkeitslücke zu vermindern.

Durch die Analyse der Reformansätze lassen sich Möglichkeiten und Grenzen von Reformen der Sozialversicherungszweige ableiten. Leistungskürzungen tragen sehr stark zur finanziellen Stabilisierung der Sozialversicherungszweige bzw. des Bundes bei. Sie sind jedoch nur in begrenztem Maße umsetzbar, da sie insbesondere für finanziell schlechter gestellte Personen unter dem Gesichtspunkt der sozialen Teilhabe nicht tragbar erscheinen.

Das vorhandene Umlagesystem in der Sozialversicherung sollte durch den Ausbau von kapitalgedeckten Elementen (obligatorische ergänzende Altersvorsorge, Erhöhung des Pflegevorsorgefonds oder andere kapitalgedeckte Modelle) ergänzt werden. Dazu könnten kurzfristig auch die Sozialversicherungsbeitragssätze erhöht werden. Es bleibt nur noch wenig Zeit, bis die „Babyboomer“ in Rente gehen und die finanziellen Belastungen in den Sozialversicherungszweigen stark steigen. Nur in der kurzen noch verbleibenden Zeit können die „Babyboomer“ in ihrer Erwerbsphase an der Vorsorge beteiligt werden. In der Rentenphase mit den geringeren Alterseinkommen ist eine Beteiligung an den steigenden Ausgaben nur begrenzt möglich.

Insgesamt werden durch das analysierte Reformpaket die mit der demografischen Veränderung einhergehenden finanziellen Belastungen gleichmäßiger auf die Generationen verteilt. Die Lasten werden nicht so stark – wie derzeit in den überwiegend umlagefinanzierten Systemen angelegt – in die Zukunft verschoben. Die Auswirkungen auf den Gesamtsozialversicherungsbeitragssatz und den Bund werden deutlich begrenzt. Gleichzeitig wird ein deutlich besseres Rentenniveau als ohne diese Reformen erreicht.

11.3 Sofortiges Handeln der politisch Verantwortlichen nötig

Aufgrund der demografischen Entwicklung besteht die Gefahr, dass zukünftige Generationen von Beitragszahlenden und Steuerzahlenden zu stark belastet, die wirtschaftliche Entwicklung gefährdet und der Bund an die Grenzen seiner finanziellen Leistungsfähigkeit gebracht wird. Deshalb muss der Gesetzgeber handeln. Und er muss es schnell tun.

Nach Auffassung des Bundesrechnungshofes ist es unabdingbar, das bisherige Sozialversicherungsystem an die sich verändernde Altersstruktur der Bevölkerung anzupassen. Die sinkende Anzahl der Personen im Erwerbsalter wird auch die Wirtschaft vor große Probleme stellen. Dies wird auch Auswirkungen auf die wirtschaftliche Leistungsfähigkeit der Volkswirtschaft und damit auf die Steuereinnahmen des Staates haben. Der Alterungsprozess der Bevölkerung lässt sich nicht aufhalten. Er ist auch eine Folge und Errungenschaft des Wohlstandes in gut funktionierenden sozial orientierten Industriestaaten. Verbesserte Lebensbedingungen und medizinische Versorgung haben dem überwiegenden Anteil der Bevölkerung ein immer längeres Leben in relativem Wohlstand und Gesundheit ermöglicht.

Der Bundesrechnungshof vertritt aufgrund seiner Analyse die Auffassung, dass die politisch Verantwortlichen nicht umhinkommen werden, die verlautbarten politischen Ziele Beitragssatzstabilität, Erhaltung des derzeitigen Versorgungsniveaus sowie keine Steuererhöhungen und keine Neuverschuldung aufzugeben. Denn die Einhaltung dieser politischen Ziele kommt

einer Quadratur des Kreises gleich. Nur durch eine moderate Erhöhung der Sozialversicherungsbeitragssätze und durch ausgewogene Leistungsverminderungen, flankiert durch Bundeszuschüsse, können nach Auffassung des Bundesrechnungshofes die Ausgaben der Sozialversicherungszweige mit den Einnahmen in Einklang gebracht werden. Welche Maßnahmen umgesetzt werden sollen, müssen die politisch Verantwortlichen entscheiden. Diese Entscheidungen sind unbequem und unpopulär, müssen aber getroffen werden.

Um Vorschläge bewerten zu können und sie in ein konkretes politisches Programm münden zu lassen, ist es nach Auffassung des Bundesrechnungshofes zunächst erforderlich, die Situation der Sozialversicherungszweige transparent darzulegen. Hierzu gehört neben dem Tragfähigkeitsbericht zu den öffentlichen Finanzen[123]

- der regelmäßige Ausweis bestehender Ansprüche (Rentenanwartschaften, niedriger Beitrag in der Krankenversicherung der Rentner, Versorgung im Pflegefall) in den Sozialversicherungssystemen (GRV, GKV und SPV) auf der Basis von versicherungsmathematischen Berechnungen und
- der Ausweis von versicherungsfremden Leistungen, die die Sozialversicherungen erbringen, die aber in die Finanzierungsverantwortung des Bundes fallen.[124]

In Gesetzesvorhaben sollte die Gesetzesfolgenabschätzung einen langfristigen Zeitraum umfassen. Die finanziellen Folgen der Gesetze sollten z. B. auf der Basis von versicherungsmathematischen Berechnungen für den Versichertenbestand oder auf der Grundlage von Tragfähigkeitsanalysen wie im Tragfähigkeitsbericht abgeschätzt werden. Diese Analysewerkzeuge sind derzeit in den Ressorts nicht oder nur rudimentär vorhanden. Sie sind aber unabdingbar für eine nachhaltige und generationengerechte Gesetzgebung im Bereich der Sozialversicherungszweige.

Die Analyse des Bundesrechnungshofes hat gezeigt, dass sich auf der Basis eines konsistenten Modells die mit konkreten Reformmaßnahmen einhergehenden Veränderungen bei den Sozialversicherungsbeitragssätzen, den Bundeszuschüssen und der Verschuldung des Bundes in Modellrechnungen darstellen lassen.

Es ist die Aufgabe der Politik, die Maßnahmen zu beschließen. Dabei sind stets auch die Aspekte der sozialen Gerechtigkeit zu beachten. Der Staat sollte durch eine entsprechende Ausgestaltung seines Steuer- und Transfersystems die finanziellen Lasten sozial ausgewogen verteilen, ohne dass zusätzliche bürokratische und kostenintensive neue Umverteilungssysteme in der Sozialversicherung geschaffen werden müssen.

---

[123] Bundesministerium der Finanzen, Vierter Bericht zur Tragfähigkeit der öffentlichen Finanzen, Jahr 2016, Seite 13.

[124] Vgl. Bericht an den Haushaltsausschuss des Deutschen Bundestages nach § 88 Absatz 2 BHO über die finanzielle Lage der gesetzlichen Krankenversicherung; Gegenstand und Auskömmlichkeit des Bundeszuschusses an die gesetzliche Krankenversicherung vom 1. Februar 2021.

Die politisch Verantwortlichen sollten auf der Basis der transparenten Bestandsanalyse in einem weiteren Schritt ein politisches Programm für alle Sozialversicherungszweige erarbeiten, in dem Maßnahmen und deren finanzielle Auswirkungen aufgeführt sind. Bei der Auswahl dieser Maßnahmen wird darüber entschieden, in welchem Umfang Leistungen gekürzt, Sozialversicherungsbeiträge erhöht und die finanziellen Auswirkungen durch eine höhere Verschuldung des Bundes auf künftige Generationen abgewälzt werden. Die Analyse des Bundesrechnungshofes hat aufgezeigt, wie ein Reformpaket aussehen könnte und welche finanziellen Wirkungen es erzeugen würde. Die Reformvorschläge und deren finanzielle Auswirkungen liegen auf dem Tisch. Es fehlen die politischen Entscheidungen. Je länger diese Entscheidungen aufgeschoben werden, umso größer werden die finanziellen Belastungen und die ökonomischen Verwerfungen, insbesondere für künftige Generationen.

12 Gemeinsame Stellungnahme des Bundesministeriums für Arbeit und Soziales, des Bundesministeriums für Gesundheit sowie des Bundesministeriums der Finanzen

12.1 Allgemeine Vorbemerkungen und Anmerkungen zu den Simulationsberechnungen

Die Bundesministerien führten in ihrer gemeinsamen Stellungnahme vom 18. Januar 2022 aus, dass Projektionen für sehr lange Zeiträume auf Basis des geltenden Rechts (hier bis zum Jahr 2060) grundsätzlich keine Prognosen darstellten, sondern lediglich Anhaltspunkte für im Zeitablauf auftretenden politischen Handlungsbedarf geben könnten. Denn annahmegemäß sei ausgeschlossen, dass der Gesetzgeber im Projektionszeitraum agiere. Konkrete Ergebnisse der Projektionen seien schon aufgrund des langen Betrachtungshorizonts in hohem Maße unsicher und hingen maßgeblich von den zugrundeliegenden Annahmen ab. Um die Unsicherheit über zukünftige Entwicklungen abzubilden, würden bei solchen Berechnungen in der Regel unterschiedliche Varianten präsentiert, um damit ein Spektrum möglicher demografischer und ökonomischer Entwicklungen im Sinne eines Korridors aufzuzeigen. Dieses Vorgehen sei bei wissenschaftlichen Gutachten üblich und werde auch für den Tragfähigkeitsbericht des BMF und im jährlichen Rentenversicherungsbericht der Bundesregierung genutzt (Kombination von drei Entgelt- mit drei Beschäftigungsvarianten). In dem vorliegenden Bericht würden Auswirkungen annahmebedingter Abweichungen vom Basisszenario nur anhand von vier alternativen Setzungen und lediglich für zwei Tragfähigkeitsindikatoren dargestellt. Die Auswahl der alternativen Annahmen sei nicht begründet. Insbesondere wiesen sie keinen Korridor zwischen günstigeren und ungünstigeren Varianten aus, mit dem die enorme Bandbreite möglicher Entwicklungen dargestellt werde, um der erheblichen Unsicherheit bei derart langfristigen Projektionen Rechnung zu tragen. Vor diesem Hintergrund sei der Aussagewert der unter anderem in der Zusammenfassung ohne Einschränkung

dargestellten zukünftigen Entwicklungen – und daraus abgeleitet die politischen Empfehlungen – kritisch zu sehen.

Ungeachtet dessen lieferten die im Bericht aufgezeigten Berechnungen grundsätzlich keine neuen Erkenntnisse, die über die bekannten Studien mit ebensolchen Berechnungen hinausgingen. Auf Basis des vorliegenden Berichtsentwurfs könnten jedenfalls die Aussagen, dass die Herausforderungen in Folge der demografischen Entwicklung „bislang nicht ausreichend betrachtet“ würden und die angeblich daraus folgenden „Fehlentwicklungen“, wie etwa die Sozialversicherungspolitik sei eine "Politik nach Kassenlage", nicht hergeleitet werden.

Zu den Simulationsberechnungen wiesen die Bundesministerien darauf hin, dass bei derart langen Projektionszeiträumen (bis zum Jahr 2060) selbst vermeintlich sichere Vorausberechnungen zur Demografie großen Unsicherheiten unterlägen. Um die Bandbreite möglicher Entwicklungen zu berücksichtigen, weise die aktuelle 14. koordinierte Bevölkerungsvorausberechnung des Statistischen Bundesamtes aus dem Jahr 2018 mehrere Varianten aus. Ein Vergleich der aktuellen mit früheren Bevölkerungsvorausberechnungen zeige, dass ein geringer als erwarteter Anstieg der Lebenserwartung sowie eine höher als erwartete Nettozuwanderung und Geburtenzahl dazu geführt haben, dass sich der für das Jahr 2050 projizierte Altenquotient deutlich reduziert habe.

Neben der Unsicherheit über die demografische und gesamtwirtschaftliche Entwicklung hinsichtlich der Einnahmeentwicklung sei in der GKV insbesondere auch die Ausgabenentwicklung von sehr hoher Unsicherheit geprägt, da diese von einer Vielzahl an angebots- und nachfrageseitigen Faktoren in vielen sehr unterschiedlich regulierten Leistungsbereichen abhängig sei. Neben unmittelbar ausgabenrelevanten gesetzgeberischen Maßnahmen (etwa Änderungen zur Definition des Leistungskatalogs der GKV) hätten Vergütungsvereinbarungen zwischen Krankenkassen, Leistungserbringern und deren Organisationen auf verschiedenen Verhandlungsebenen und in verschiedenen Leistungsbereichen sowie die Richtlinien und Beschlüsse des Gemeinsamen Bundesausschusses zur Konkretisierung des Leistungsumfanges einen großen und kaum vorhersagbaren Einfluss auf die Ausgabenentwicklung. All dem könne in der gängigen Projektionspraxis durch den annahmegemäßen Ausschluss gesetzgeberischer Tätigkeiten im Projektionszeitraum nicht hinreichend Rechnung getragen werden. Auch seien die Leistungen der GKV und der SPV (anders als bei der GRV) nicht an die Höhe der gezahlten Beiträge gekoppelt, sondern von Sach- und Dienstleistungen geprägt, deren Preis-, Mengen- und Strukturentwicklung langfristig kaum valide vorhersagbar sei. Vor diesem Hintergrund erstelle das BMG keine langfristigen Modellrechnungen zur GKV.

Auch für die SPV werde nur eine Grobabschätzung der langfristigen Entwicklung der Pflegebedürftigenzahl vorgenommen, welche jedoch für die Finanzentwicklung der SPV maßgeblich sei und somit zu einer weiteren Unsicherheit der Projektion führe.

12.2 Stellungnahme zu den Vorausberechnungen

Gesetzliche Rentenversicherung (Tnr. 3)

Zur Darstellung der zukünftigen Entwicklung der GRV werde die Entwicklung des Rentnerquotienten, des Rentenniveaus, des Beitragssatzes sowie der Bundeszuschüsse bis zum Jahr 2060 ausgewiesen. Grundlage des Basisszenarios bilde der rentenrechtliche Status quo zu Beginn des Jahres 2020. Die Ergebnisse könnten von Seiten des BMAS nicht im Detail nachvollzogen werden. Die dargestellten Tendenzen scheinen aber nicht grundsätzlich von anderen Berechnungen dieser Art abzuweichen.

Gesetzliche Krankenversicherung (Tnr. 4)

In den Szenarienrechnungen des Bundesrechnungshofes für die GKV stiegen die Leistungsausgaben in Relation zum BIP bis zum Jahr 2060 deutlich stärker als in vergleichbaren Schätzungen wie dem Tragfähigkeitsbericht des BMF aus dem Jahr 2020 oder dem aktuellen Ageing Report der EU-Kommission. Auch für die Berechnungen zur GKV gelte, dass weder Methodik und Annahmen noch die Ergebnisse der Langfristprojektion durch den Bundesrechnungshof im Detail erläutert würden und daher auch nicht nachvollzogen werden könnten. Ergänzend sei anzumerken, dass der Bericht nicht nur die Auswirkungen der demografischen Entwicklung, sondern auch die Finanzwirkungen des medizinisch-technischen Fortschritts thematisiere. In der Berechnung der Entwicklung der altersspezifischen Pro-Kopf-Ausgaben werde allerdings das BIP als Grundlage genutzt, ohne eine Verbindung zum medizinisch-technischen Fortschritt herzustellen. Hier wäre es sachgerecht und transparent gewesen, die vergleichsweise moderaten finanziellen Effekte der demografischen Entwicklung auf die GKV-Ausgabenentwicklung isoliert auszuweisen und die Finanzwirkungen des medizinisch-technischen Fortschritts in Alternativvarianten darzustellen.

Soziale Pflegeversicherung (Tnr. 5)

Die Annahmen zur Einnahmeentwicklung seien bei der SPV wie bei den anderen beiden Sozialversicherungszweigen langfristig mit hoher Unsicherheit verbunden. Auf der Ausgabenseite werde – wie in vielen wissenschaftlichen Modellen – eine konstante altersspezifische Pflegewahrscheinlichkeit unterstellt. Die für die Entwicklung der Leistungsbeträge unterstellte Anlehnung teilweise an die allgemeine Preisentwicklung und teilweise an die Lohnentwicklung stimme nicht mit der aktuellen Rechtslage überein, die eine inflationsorientierte Leistungsdynamisierung vorsehe, wobei die Leistungsdynamisierung die Bruttolohnentwicklung nicht übersteigen dürfe. Entsprechend müsste der Ausgabenanstieg in der Modellrechnung des Bundesrechnungshofes schwächer sein, da im langjährigen Durchschnitt die Löhne stärker stiegen als die Preise. Auch der prognostizierte langfristig notwendige Beitragssatzanstieg würde schwächer ausfallen.

12.3 Finanzielle Gesamtsituation der Sozialversicherung und deren Risiken für den Bund

Die Herausforderungen für die Tragfähigkeit der öffentlichen Finanzen, die sich im geltendem Rechtsrahmen durch die demografische Entwicklung für die Finanzierung der Sozialversicherung bis zum Jahr 2060 ergeben könnten, nehme auch der Tragfähigkeitsbericht des BMF in den Blick, der regelmäßig eine Vorausschau auf die Tragfähigkeit der öffentlichen Finanzen vorlege. Gegenüber den Simulationsrechnungen des Bundesrechnungshofes berücksichtige der Tragfähigkeitsbericht eine größere Bandbreite möglicher Entwicklungen und trage so auch der mit dem Projektionszeitraum steigenden Unsicherheit stärker Rechnung. Aus den Modellrechnungen könne auf Handlungsbedarf geschlossen werden, der unter den getroffenen unterschiedlichen Annahmen im Projektionszeitraum entstehen würde. Als Prognosen seien diese jedoch nicht zu verstehen.

Hinsichtlich der ausgewiesenen künftigen Entwicklung wichtiger Indikatoren wie der Schuldenstandsquote und der Tragfähigkeitslücke sei darauf hinzuweisen, dass diese auch auf einer Annahme zur künftigen Zinsentwicklung beruhe. Wie der Tragfähigkeitsbericht des BMF zeige, seien diese Indikatoren in hohem Maße von der getroffenen Zinsannahme abhängig. Auch hier würde die Berücksichtigung einer größeren Bandbreite möglicher Zinsentwicklungen der Unsicherheit zukünftiger Entwicklungen stärker Rechnung tragen.

12.4 Stellungnahme zu den Reformszenarien

Die Bundesministerien führen in ihrer Stellungnahme aus, dass zu vielen der vom Bundesrechnungshof vorgebrachten Reformmaßnahmen die Parteien der Regierungskoalition mit dem am 7. Dezember 2021 unterzeichneten Koalitionsvertrag bereits konkrete Pläne vorgelegt hätten.

Anhebung des gesetzlichen Renteneintrittsalters (Tnr. 7. 2. 1)

Um den negativen Folgen der demografischen Entwicklung auf die GRV entgegenzuwirken, werde die Regelaltersgrenze nach geltendem Recht vom vollendeten 65. auf das vollendete 67. Lebensjahr angehoben. Die Regierungskoalition hätte festgelegt, dass es in der 20. Legislaturperiode darüber hinaus keine Anhebung des gesetzlichen Renteneintrittsalters geben werde.

Die schrittweise Anhebung der Regelaltersgrenze schaffe Planungssicherheit für alle Beteiligten – für die Betriebe, die Beschäftigten, die Tarifpartner, die Sozialversicherungen und die Politik. Bei einer durch – Kopplung an die Lebenserwartung – dynamischen Regelaltersgrenze wie vom Bundesrechnungshof vorgeschlagen würde diese Planungssicherheit aufgegeben. Denn die Messung bzw. Schätzung des Anstiegs der Lebenserwartung sei schwierig und die Umsetzung des Vorschlags mit praktischen Problemen verbunden.

Ausweitung des Versichertenkreises (Tnr. 7.2.2)

Hinsichtlich der Einbeziehung von Selbstständigen in die GRV sähe der Koalitionsvertrag für die 20. Legislaturperiode vor, für alle neuen Selbstständigen, die keinem obligatorischen Alterssicherungssystem unterliegen, eine Pflicht zur Altersvorsorge mit Wahlfreiheit einzuführen. Selbstständige sollten in der GRV versichert sein, sofern sie nicht im Rahmen eines Opt-Outs ein privates Vorsorgeprodukt wählten, das bestimmte Voraussetzungen erfüllen müsse.

Das BMAS teilt die Darstellung des Bundesrechnungshofes, dass die Einbeziehung der Beamtinnen und Beamten in die GRV langfristig zu einer stärkeren Belastung junger und zukünftiger Generationen von Beitragszahlenden führen würde und je nach Ausgestaltung negative Effekte auf die öffentlichen Haushalte insbesondere von Bund und Ländern haben könnte.

Stellschraube Rentenniveau (Tnr. 7. 3)

Der Bundesrechnungshof stelle zutreffenderweise dar, dass die Rentenanpassungsformel geeignet sei, die Beitragssatzbelastung der kommenden Generationen zu dämpfen. Zur Haltelinie beim Sicherungsniveau vor Steuern (Rentenniveau) werde ebenso auf den Koalitionsvertrag verwiesen wie zur Reaktivierung des Ausgleichsbedarfes ("Nachholfaktor").

Der Bundesrechnungshof stelle zutreffend fest, dass eine Anhebung der Beitragsbemessungsgrenze zu einem geringeren Beitragssatz führen könne. Dies träfe allerdings nur zu, wenn den zusätzlichen Beiträgen keine zusätzlichen Anwartschaften gegenüberstünden, was jedoch dem die GRV prägenden Äquivalenzprinzip widersprechen würde.

Ausbau der zusätzlichen Altersvorsorge (Tnr. 7. 5)

Hinsichtlich des Elementes der zusätzlichen privaten Altersvorsorge im Reformszenario verweist das BMAS auf den Koalitionsvertrag für die 20. Legislaturperiode: Das bisherige System der privaten Altersvorsorge solle grundsätzlich reformiert und dazu auch das Angebot eines öffentlich verantworteten Fonds mit einem effektiven und kostengünstigen Angebot mit Abwahlmöglichkeit geprüft werden. Hierbei werde auch der Grundsatz der paritätischen Finanzierung der Altersvorsorge zu beachten sein.

Erweiterung des Versichertenkreises um Beamtinnen und Beamte in der GKV (Tnr. 8.2.2)

Zur Reduzierung des GKV-Beitrags schlage der Bundesrechnungshof vor, dass alle Personen, die ab dem Jahr 2022 neu verbeamtet werden, versicherungspflichtig in der GKV werden.

Beamtinnen und Beamte hätten nach geltendem Recht mit Eintritt der Versicherungsfreiheit (i. d. R. zu Beginn ihrer Beamtenlaufbahn) einmal die Wahlmöglichkeit, eine bestehende Versicherung in der GKV als freiwilliges Mitglied fortzusetzen oder sich für eine Absicherung im Krankheitsfall über die Beihilfe und eine private Restkostenversicherung für den von der Beihilfe nicht umfassten Teil der Krankheitskosten zu entscheiden.

Der Vorschlag des Bundesrechnungshofes gehe über die bisherige Ausgestaltung in einzelnen Bundesländern hinaus, indem eine Versicherungspflicht eingeführt und damit keine Wahlmöglichkeit für Neu-Beamtinnen und -Beamte eröffnet werden solle. Als problematisch wird unter anderem bewertet, dass der Bundesrechnungshof nicht berücksichtige, dass die mit diesem Vorschlag verbundene „Vergreisung" der Beihilfeergänzungstarife in der PKV mit erheblichen Prämiensteigerungen für die Bestandsbeamtinnen und -beamten verbunden sein dürfte. Inwieweit sich diese Verträge dann langfristig überhaupt noch finanziell selbst tragen könnten, sei zumindest fragwürdig.

Abschließend weisen die Bundesministerien darauf hin, dass im Koalitionsvertrag für die 20. Legislaturperiode keine Reform der Absicherung von Beamtinnen und Beamten im Krankheitsfall thematisiert ist.

Beitragszuschläge in der GKV bei vorzeitigem Eintritt in die Rente (Tnr. 8.3.3)

Der Bundesrechnungshof schlage vor, Beitragszuschläge bei vorzeitigem Eintritt in die Rente einzuführen, um geringere Beitragseinnahmen zu kompensieren.

Der Vorschlag widerspräche dem Solidarprinzip der GKV, nach welchem Beiträge gemäß der individuellen wirtschaftlichen Leistungsfähigkeit erhoben werden. Der Gesetzgeber habe insbesondere langjährig Versicherten nach 35 Versicherungsjahren die Möglichkeit eingeräumt, vorzeitig Altersrente in Anspruch zu nehmen. Mit dem allgemeinen Gleichheitssatz wäre es kaum zu vereinbaren, langjährig versicherte Altersrentner schlechter zu behandeln als Regelaltersrentner, die lediglich eine Wartezeit von fünf Jahren erfüllten. Oft führten zudem gesundheitliche, soziale oder familiäre Ursachen oder eine Schwerbehinderung zu einem vorzeitigen Renteneintritt.

Eine Abweichung von den allgemeinen Prinzipien der Beitragsbemessung in der GKV lasse sich so nicht begründen. Um auch bei niedrigen Renten eine an der wirtschaftlichen Leistungsfähigkeit orientierte Beitragserhebung in der GKV sicherzustellen, seien bereits nach geltendem Recht auch Versorgungsbezüge und Arbeitseinkommen beitragspflichtig.

Zuzahlungen in der GKV (Tnr. 8.4.2)

Der Bundesrechnungshof schlage vor, Zuzahlungen in der GKV deutlich zu erhöhen, beispielsweise zu verdoppeln.

Die vom Bundesrechnungshof vorgeschlagene Erhöhung der Zuzahlungen würde eine Verschiebung von der solidarischen Finanzierung der GKV (Beiträge nach wirtschaftlicher Leistungsfähigkeit) hin zu einer stärkeren individuellen Finanzierung je nach Leistungsinanspruchnahme bedeuten. Die häufig mit Zuzahlungen verbundene Erwartung einer Steuerungswirkung hinsichtlich der Leistungsinanspruchnahme sei zudem nicht belegt.

Reformen in der Pflegeversicherung (Tnr. 9)

Im Bereich der SPV weisen die Bundesministerien zunächst darauf hin, dass eine Verringerung der Pflegewahrscheinlichkeit ein Instrument zur Kostenbegrenzung sei. Dieses Ziel verfolge auch die Bundesregierung mittels Präventions- und Rehabilitationsmaßnahmen.

Anpassungen bezüglich des Pflegevorsorgefonds (Tnr. 9.4)

Die vom Bundesrechnungshof vorgeschlagene Ausweitung des Pflegevorsorgefonds zur Beitragssatzglättung lehnen die Bundesministerien ab, weil der Glättungseffekt entweder nicht sonderlich groß sei oder die damit verbundene Umverteilung der Beitragsbelastung über eine generationengerechte Belastungsverteilung hinausginge.

13 Abschließende Würdigung des Bundesrechnungshofes

13.1 Allgemeine Vorbemerkungen und Anmerkungen zu den Simulationsberechnungen

Die Simulationsrechnungen sind nach Auffassung des Bundesrechnungshofes durchaus geeignet, die finanziellen Auswirkungen der demografischen Entwicklung auf die Sozialversicherungssysteme darzustellen und den Handlungsbedarf aufzuzeigen. Das dabei gewählte methodische Vorgehen entspricht dem aktuellen Stand der Wissenschaft. Das BMF greift z. B. für die Annahmen zur Erstellung des Tragfähigkeitsberichts auf ebensolche Berechnungen zurück.

Der Bundesrechnungshof stimmt mit den Bundesministerien darin überein, dass die Berechnungen bis zum Jahr 2060 keine Prognosen darstellen. Vielmehr handelt sich um langfristige Vorausberechnungen (Projektionen) unter Berücksichtigung bestimmter Annahmen zur wirtschaftlichen und demografischen Entwicklung. Die gewählten Annahmen bestimmen das Ergebnis der Berechnungen. Der Bundesrechnungshof hat sowohl für die wirtschaftliche als auch für die demografischen Entwicklung plausible und gleichzeitig moderate, also mittlere Annahmen gewählt. Um die Bandbreite der möglichen Ergebnisse bei unterschiedlichen

Annahmen abzubilden, werden in Studien häufig unterschiedliche Szenarien verwandt, die mit einer unteren und oberen Variante einen Korridor aufzeigen. Der Bundesrechnungshof hat sich bewusst gegen die Darstellung eines solchen Korridors entschieden. Denn er geht davon aus, dass aufgrund seiner mittleren Annahmen auch das Ergebnis der Berechnungen im mittleren Bereich eines Korridors liegen wird. Somit sind Übertreibungen aufgrund extrem negativer Annahmen und Verharmlosung aufgrund extrem günstiger Annahmen ausgeschlossen. Das berechnete Basisszenario stellt vielmehr ein mittleres Szenario dar.

Der Hinweis der Bundesministerien, dass die aufgezeigten Berechnungen grundsätzlich keine neuen Erkenntnisse erbracht hätten, die über die bekannten Studien mit ebensolchen Berechnungen hinausgehen, bestätigt nur das Vorgehen des Bundesrechnungshofes. Seine Berechnungen sind im Einklang mit anderen Studien und weder in die eine noch in die andere Richtung extrem.

Der Bundesrechnungshof bleibt bei seiner Einschätzung, dass in der Sozialversicherung eine „Politik nach Kassenlage“ praktiziert wird. Zu dieser Einschätzung kommt er nicht auf der Basis seines Berichtes, sondern vielmehr aufgrund der Analyse der politischen Entscheidungen der letzten Jahre. Die „Politik nach Kassenlage“ ist im Gegenteil ein Grund, warum die im Bericht dargestellten Herausforderungen so groß sind und der Handlungsbedarf so dringlich. Wie die Bundesministerien zutreffend darlegen, sind die finanziellen Risiken durch die demografische Entwicklung lange bekannt. In der Wirtschaftskrise zu Beginn der 2000er Jahre wurden auch deshalb Reformen in den Sozialversicherungssystemen beschlossen, die kurz- und auch langfristig deren Finanzlage verbessern sollten. In der anschließenden Phase des Wirtschaftsaufschwungs beschloss die Politik dann eine Vielzahl von Leistungsverbesserungen in allen Sozialversicherungszweigen und nahm zudem Reformen zur langfristigen finanziellen Stabilisierung der Sozialversicherungszweige zurück. Noch immer sind die Sozialversicherungssysteme nicht grundlegend und ausreichend an die finanziellen Herausforderungen der demografischen Entwicklung angepasst worden. Finanzierungslücken in den Sozialversicherungssystemen werden derzeit vielmehr nur kurzfristig und kurzsichtig durch höhere Bundeszuschüsse geschlossen. Dies geht einher mit einer höheren Verschuldung, wodurch die finanziellen Lasten in die Zukunft und auf die jungen Generationen verschoben werden. Nach Auffassung des Bundesrechnungshofes sollten die politischen Entscheidungsträger zeitnah darlegen, wie die Sozialversicherungssysteme langfristig finanzierbar, generationengerechter und damit zukunftsfest gestaltet werden können.

13.2 Abschließende Würdigung zu den Vorausberechnungen

Die Bundesministerien haben in ihrer Stellungnahme dargelegt, dass die vom Bundesrechnungshof für die GRV dargestellten Entwicklungen nicht grundsätzlich von anderen Berechnungen dieser Art abweichen. Der Bundesrechnungshof sieht sich in seiner Analyse insofern bestätigt.

Sowohl der medizinisch-technische Fortschritt als auch die demografische Entwicklung führen zu steigenden Ausgaben in der GKV. In der Vergangenheit lag die Steigerungsrate der

Pro-Kopf-Ausgaben in der GKV u. a. aufgrund des medizinisch-technischen Fortschrittes im Durchschnitt über dem Wachstum des Pro-Kopf-BIP. Der Bundesrechnungshof hat sich deshalb dazu entschieden, die Annahmen zur Entwicklung der altersspezifischen Pro-Kopf-Ausgaben an der durchschnittlichen Entwicklung des Pro-Kopf-BIP zu orientieren. Um den Einfluss des medizinisch-technischen Fortschritts zumindest grob abzubilden, berücksichtigt er einen Aufschlag zur Zuwachsrate des Pro-Kopf-BIP. Als Ergebnis zeigt sich, dass rund die Hälfte des Anstiegs der Ausgabenquote in der GKV bis zum Jahr 2060 auf den so berücksichtigten medizinisch-technischen Fortschritt und die andere Hälfte des Anstieges auf die demografische Entwicklung zurückzuführen ist. Dieser „demografische Effekt" entsteht deshalb, weil die Pro-Kopf-Ausgaben der GKV ab einem Alter von 60 Jahren mit zunehmendem Alter stetig steigen. Durch die demografische Entwicklung – insbesondere den Eintritt der „Babyboomer" in den Ruhestand – werden sich die Anzahl der Personen im Alter über 60 Jahren und damit die Ausgaben in der GKV deutlich erhöhen. Diese stärker steigenden Ausgaben treffen auf weniger stark steigende Einnahmen, weil die GKV-Beiträge nur auf die niedrigeren Renten der größeren Anzahl der im Ruhestand befindlichen GKV-Versicherten erhoben werden. Nach der Analyse des Bundesrechnungshofes werden diese finanziellen Auswirkungen der demografischen Entwicklung somit keineswegs „moderat" ausfallen.

Dass die Ausgabenquote der GKV in den Modellrechnungen des Bundesrechnungshofes stärker steigt als dies in anderen Studien, z. B. im Tragfähigkeitsbericht der Fall ist, ist mit einer anderen Modellspezifikation und mit etwas anderen Annahmen zur Entwicklung der Pro-Kopf-Ausgaben zu begründen.

Bei der SPV stimmt der Bundesrechnungshof den Bundesministerien zu, dass auch hier die Entwicklung mit einer hohen Unsicherheit verbunden ist. Daher ist es nach Auffassung des Bundesrechnungshofes sachgerecht, auch für die SPV mittlere Annahmen für die zukünftige Entwicklung zu wählen und von konstanten altersspezifischen Pflegewahrscheinlichkeiten auszugehen. Die Pflegewahrscheinlichkeit steigt ab dem Alter von 80 Jahren deutlich an. Da der Anteil der über 80-jährigen Personen stetig zunimmt und die durchschnittliche Lebenserwartung ansteigt, steigen die projizierten Ausgaben nur aufgrund der Änderung der Altersstruktur ebenfalls deutlich. Für die Entwicklung der Leistungsausgaben pro Kopf hat der Bundesrechnungshof wieder aus seiner Sicht plausible und in der Höhe mittlere Fortschreibungsraten angenommen. Dass diese ggf. nicht mit den gesetzlich vorgesehenen Fortschreibungsraten der Leistungspauschalen übereinstimmen, ist damit begründet, dass im Modell nicht die Leistungspauschalen, sondern die Pro-Kopf-Leistungsausgaben fortgeschrieben werden.

Zu der finanziellen Gesamtsituation der Sozialversicherung und deren Risiken für den Bund verweisen die Bundesministerien darauf, dass der vom BMF veröffentlichte Tragfähigkeitsbericht eine größere Bandbreite möglicher Entwicklungen (Varianten) berücksichtige, um den steigenden Unsicherheiten stärker Rechnung zu tragen. Die Schuldenstandsquote und die Tragfähigkeitslücke seien zudem stark von der Zinsentwicklung abhängig. Der Bundesrechnungshof hat dargelegt, warum er auf die Betrachtung unterschiedlicher Varianten verzichtet hat. Grundlage seiner Bestandanalyse ist ein moderates – mittleres – Szenario (Basisszenario). Dies gilt auch für die Annahme zur Zinsentwicklung. Damit weisen auch die

zinssensiblen Tragfähigkeitsindikatoren eher mittlere Werte auf. Ihre Funktion, den finanzpolitischen Handlungsbedarf aufzuzeigen, verlieren sie nicht.

13.3 Abschließende Würdigung zu den Reformszenarien

Der Bundesrechnungshof hat auf der Grundlage des Basisszenarios die Auswirkungen möglicher Reformansätze berechnet. Dies soll Möglichkeiten und Grenzen von Reformen in den Sozialversicherungszweigen aufzeigen. Da das gesamte Sozialversicherungssystem und auch der Bundeshaushalt berücksichtigt werden, können ein Gesamteffekt der Reformen simuliert und Wechselwirkungen zwischen den Sozialversicherungszweigen erfasst werden. In anderen Studien wurde dieser umfassende Ansatz bisher kaum verfolgt.

Der Bundesrechnungshof bleibt bei seiner Auffassung, dass ein politisches Programm zur Verminderung der finanziellen Risiken in den Sozialversicherungszweigen die Einzelmaßnahmen aufzeigen sollte. Zu dieser Darstellung gehört auch aufzuzeigen, wie sich die finanziellen Auswirkungen langfristig darstellen.

Der Bundesrechnungshof hat mit der Analyse der Reformszenarien bezweckt, die mit der demografischen Entwicklung einhergehenden finanziellen Belastungen gleichmäßiger auf die Generationen zu verteilen. Die Bundesministerien lehnen einige der Reformoptionen u. a. mit dem Hinweis auf einen Verstoß gegen das Solidarprinzip ab. Bei anderen Vorschlägen verweisen sie darauf, dass ähnliche Maßnahmen im Koalitionsvertrag der Regierungsparteien vom 7. Dezember 2021 bereits geplant seien.

Der Bundesrechnungshof gibt zu bedenken, dass die Solidarität nicht nur für die derzeit in den Sozialversicherungen Versicherten gelten sollte, sondern auch für die künftigen Generationen. Der Bundesrechnungshof bleibt deshalb bei seiner Auffassung, dass ein ausgewogenes generationengerechtes Reformkonzept auch Leistungskürzungen und höhere Belastungen enthalten sollte. Denn Leistungskürzungen und höhere Belastungen für die Einen bedeuten Entlastungen für die Anderen. Entlastet werden sollen vor allem junge und künftige Generationen. Dies kann dadurch erreicht werden, dass man den künftigen Beitragssatzanstieg und die Verschuldung des Bundes begrenzt.

Die im Koalitionsvertrag dargestellten Reformvorhaben wie die Reaktivierung des Nachhaltigkeitsfaktors, die Umgestaltung der staatlich geförderten zusätzlichen Altersvorsorge oder die Einbeziehung der nicht obligatorisch abgesicherten Selbstständigen in der GRV sind deshalb zu begrüßen. Bisher handelt es sich jedoch lediglich um Absichtserklärungen. Die Reformvorhaben sind noch nicht umgesetzt. Der Bundesrechnungshof empfiehlt, zeitnah entsprechende Gesetze zu verabschieden. Zudem sollten auch konkrete Berechnungen vorgelegt werden, in welchem Umfang sich diese Vorhaben auf die langfristige finanzielle Entwicklung in den Sozialversicherungszweigen und die Verschuldung des Bundes auswirken.

Ein Gesamtkonzept, wie die Sozialversicherungssysteme zukunftsfest gemacht werden sollen, bietet der Koalitionsvertrag indes nicht. Der Bundesrechnungshof bleibt daher bei seiner

Kernforderung, die Sozialversicherungssysteme an die sich verändernde Altersstruktur der Bevölkerung anzupassen. Dazu sollte zunächst die Situation der Sozialversicherungszweige transparent dargelegt werden. Gesetzesvorhaben sollten eine Gesetzesfolgenabschätzung für einen langfristigen Zeitraum enthalten. Zudem sollten die politisch Verantwortlichen auf der Basis der transparenten Bestandsanalyse ein politisches Programm für alle Sozialversicherungszweige erarbeiten, in dem die Maßnahmen und deren finanzielle Auswirkungen aufgeführt sind. Dabei kann selbstverständlich auf Leistungskürzungen verzichtet werden. Dies führt aber zwangsläufig dazu, dass die Sozialversicherungsbeitragssätze und/oder die Verschuldung des Bundes höher liegen und künftige Generationen stärker belastet werden. Damit werden die finanziellen Lasten der demografischen Entwicklung auf künftige Generationen abgewälzt. Es ist nach Auffassung des Bundesrechnungshofes die Aufgabe der politisch Verantwortlichen darzulegen, wie die ansteigenden finanziellen Lasten in den Sozialversicherungszweigen auch in Zukunft finanziert werden sollen und welche Bevölkerungsgruppen und Generationen diese Finanzierungslasten zu tragen haben.

Ahrendt Dr. Weber Wyrwoll

Beglaubigt: Hupertz, Tarifbeschäftigte

Wegen elektronischer Bearbeitung ohne Unterschrift und Dienstsiegelabdruck.

Deutscher Bundestag
Ausschuss f. Gesundheit

Ausschussdrucksache
20(14)124

21.06.2023

Bundes rechnungshof

Bundesrechnungshof · Adenauerallee 81 · 53113 Bonn

Frau Dr. Kirsten Kappert-Gonther, MdB
Stellvertretende Vorsitzende
des Ausschusses für Gesundheit des
Deutschen Bundestages

nur per E-Mail:
gesundheitsausschuss@bundestag.de

21. Juni 2023

Zeichen: IX 4 - 2020 - 0523

0151 74213781

poststelle@brh.bund.de
www.bundesrechnungshof.de
Zentrale: 0228 99 721-0

Bericht nach § 88 Absatz 2 BHO über die Prüfung „Demografische Entwicklung: Finanzrisiken des Bundes aus seiner Beteiligung an der Finanzierung der Sozialversicherungen“

Sehr geehrte Frau Vorsitzende,

am 23. Mai 2022 haben wir den o. g. Bericht an den Haushaltsausschuss des Deutschen Bundestages, den Ausschuss für Arbeit und Soziales des Deutschen Bundestages sowie den Ausschuss für Gesundheit des Deutschen Bundestages übersandt.

Der Haushaltsausschuss des Deutschen Bundestages hat diesen Bericht in seiner 20. Sitzung am 22. Juni 2022 beraten und förmlich zur Kenntnis genommen. Seit diesem Zeitpunkt erreichten uns vielfach Bitten, diesen Bericht zu veröffentlichen.

Nach § 96 Absatz 4 BHO kann der Bunderechnungshof diesen Bericht erst veröffentlichen, wenn er im Parlament abschließend beraten wurde. Dies ist zwar beim Haushaltsausschuss des Deutschen Bundestages der Fall, nicht aber bei dem von Ihnen geleiteten Ausschuss. Eine abschließende Beratung im Sinne von § 96 Absatz 4 BHO liegt nur dann vor, wenn dieser Bericht in dem von Ihnen geleiteten Ausschuss nicht mehr behandelt wird.

Wir bitten Sie daher freundlich um Mitteilung, ob dies zutrifft und der Bericht nicht mehr im Ausschuss für Gesundheit behandelt wird.

Mit freundlichen Grüßen

Ahrendt Dr. Weber Wyrwoll