

ZfME auch  
online  
www.zfme.de

# Zeitschrift für medizinische Ethik

Wissenschaft • Kultur • Religion

67. Jahrgang 2021

Heft 3

## KI in der Medizin

- W. Liedtke/  
M. Langanke Der Einsatz von IT-basierten Decision-Support-Systemen in der  
medizinischen Versorgung aus verantwortungsethischer Sicht
- K. Arshad/B. Schrei- Überblick über ethische und soziale Herausforderungen im  
weis/C. Strodthoff/ Kontext der klinischen Nutzung von KI-Systemen  
B. Bergh
- H.-P. Schlemmer/  
M. Hohenfellner Chancen von KI in der Onkologie am Beispiel der individuali-  
sierten Diagnostik und Behandlung von Prostatakrebs
- I. Schneider Diskriminierungsgefahren und Regulationsansätze bei der  
medizinischen Nutzung von Künstlicher Intelligenz
- T. Weichert Datenschutz im Kontext der medizinischen Nutzung von  
KI-Systemen: heute und morgen
- F. Funer Arzt – Patient – Algorithmus: Ethische und kommunikative  
Erwägungen zu einer KI-gestützten Beziehung
- S. K. Nagel Fragen für eine Zukunft automatisierter Entscheidungsunter-  
stützung in der Medizin

Schwabenverlag

# Zeitschrift für medizinische Ethik

Wissenschaft · Kultur · Religion

Gegründet im Jahre 1954 als Zeitschrift für medizinisch-ethische Grundsatzfragen und bis 1992 erschienen unter dem Titel ARZT UND CHRIST.

---

## Vierteljahresschrift

**Herausgegeben im Auftrag der Görres-Gesellschaft von Franz-Josef Bormann, Alois Joh. Buch und Matthias Volkenandt. Geschäftsführender Herausgeber: Franz-Josef Bormann.**

**Redaktion:** Dr. Florian Funer, Katholisch-Theologische Fakultät, Lehrstuhl für Moralthologie Prof. Dr. F.-J. Bormann, Liebermeisterstr. 12, 72076 Tübingen, Telefon (0 70 71) 297 54 19, Telefax (0 70 71) 2928 66, E-Mail: redaktion@zfme.de

**Wissenschaftlicher Beirat:** Klaus Baumann (Caritaswissenschaft und Christliche Sozialarbeit, Freiburg i. Br.), Hermann Brandenburg (Gerontologische Pflege, Vallendar), Peter Dabrock (Systematische Theologie/Ethik, Erlangen), Andreas Heller (Palliative Care und Organisationsethik, Graz), Dirk Lanzerath (Deutsches Referenzzentrum für Ethik in den Biowissenschaften, Bonn), Karl-Heinz Leven (Geschichte und Ethik der Medizin, Erlangen), Jürgen Manemann (Forschungsinstitut für Philosophie, Hannover), Stephan Sahm (Geschichte und Ethik in der Medizin, Frankfurt a. M.), Urban Wiesing (Ethik in der Medizin, Tübingen)

**Geschäftsführender Herausgeber:** Prof. Dr. Franz-Josef Bormann, Katholisch-Theologische Fakultät, Lehrstuhl für Moralthologie, Eberhard Karls Universität Tübingen, Liebermeisterstr. 12, 72076 Tübingen, Tel.: 0 70 71-29-728 60, Franz-Josef.Bormann@uni-tuebingen.de

**Verlag:** Schwabenverlag AG, Postfach 4280, D-73745 Ostfildern, Telefon (07 11) 4406-0, Telefax (07 11) 4406-177, E-Mail: zfme@schwabenverlag.de, Internet: www.zfme.de, Vorstand: Ulrich Peters.

Manuskripte, redaktionelle Zuschriften und Besprechungsexemplare sind an die Redaktionsadresse zu richten.

**Herstellung:** Satz: Schwabenverlag AG, 73745 Ostfildern. Druck: Stückle Druck und Verlag, Ettenheim.

**Bestellungen** sind zu richten an: Schwabenverlag AG, Postfach 4280, D-73745 Ostfildern, Telefon (07 11) 4406-134, Telefax (07 11) 4406-177, E-Mail: zfme@schwabenverlag.de

**Bezugsbedingungen:** Jahresabonnement einschließlich Onlinezugang € 79,- [D] inkl. MwSt./ Studenten bzw. Abonnenten, die sich in der Ausbildung befinden (Nachweis erforderlich) € 39,50 [D] inkl. MwSt.; jeweils zusätzlich Versandkosten. Das Einzelheft kostet € 18,50 [D] zusätzlich Versandkosten. Bezug der Zeitschrift durch die Post oder den Buchhandel. Die Bezugsdauer verlängert sich jeweils um 1 Jahr, wenn nicht eine Abbestellung schriftlich 6 Wochen vor Bezugsjahresende erfolgt.

**Bankverbindungen:** Postbank Stuttgart IBAN DE28 6001 0070 0004 5597 01, BIC PBNKDEFF; Postsparkasse Wien IBAN AT47 6000 0000 0757 6654, BIC BAWAATWW; Postgiro Basel IBAN CH28 0900 0000 8004 70123, BIC POFICHBEXXX, Kto: 80-47012-3.

**Anzeigen:** Es gilt die Anzeigenpreisliste 2021.

Alle Verlagsrechte sind vorbehalten. Für unverlangt eingesandte Manuskripte wird keine Gewähr übernommen. Gezeichnete Beiträge geben nicht unbedingt die Meinung von Herausgeber und Redaktion wieder.

INGRID SCHNEIDER

## Diskriminierungsgefahren und Regulationsansätze bei der medizinischen Nutzung von KI

### Risks of discrimination and regulatory approaches in the medical use of AI

#### *Zusammenfassung*

Der Beitrag untersucht Diskriminierungsrisiken bei Künstlicher Intelligenz in der Medizin. Hierzu wird eine legitime Differenzierung von ungerechtfertigter Diskriminierung normativ unterschieden. Praxisbeispiele analysieren verschiedene Typen (Geschlecht, ethnischer *Bias*), und Gegenstandsbereiche von Diskriminierungen, darunter Neurowissenschaft, Notfallmedizin und prädiktive ‚Todesalgorithmen‘ und betonen den Gruppencharakter von Klassifizierungen. Der Zusammenhang von sozialer Ungleichheit, Diskriminierungsrisiken und gesellschaftlicher Entsolidarisierung wird problematisiert. Als Regulierungsansätze behandelt der Aufsatz die KI-Ethik und professionelle Selbstregulierung, Zertifizierungsmechanismen sowie den KI-Verordnungsverschlagn der Europäischen Kommission.

#### *Abstract/Summary*

The article examines risks of discrimination in the medical application of artificial intelligence. To this end, legitimate differentiation is normatively distinguished from unjustified discrimination. Using practice examples, different types (gender, race and ethnic bias), and subject areas of discrimination are analysed, including neuroscience, emergency medicine and predictive ‚death algorithms‘. Moreover, the group nature of classification, the connection between socio-economic inequality and discrimination risks as well as social desolidarisation are problematised. As regulatory approaches, the paper deals with AI ethics and professional self-regulation, certification mechanisms, and the European Commission’s AI regulation proposal.

#### *Schlüsselwörter*

Künstliche Intelligenz; Medizin; Diskriminierung; Risiken; Regulierung.

#### *Keywords*

Artificial intelligence; medicine; discrimination; risks; regulation.

## 1. Einleitung

Im Kontext der wissenschaftlichen und gesellschaftlichen Debatte um Künstliche Intelligenz (KI) wird derzeit viel über die Gefahren von Verzerrungen (*Bias*) und der Diskriminierung bestimmter Gruppen diskutiert. Dies betrifft etwa ungerechtfertigte Ungleichbehandlungen bei der Arbeitssuche, bei der Kreditwürdigkeit, im Justizwesen und Strafvollzug, bei staatlichen Sozialleistungen, im Bildungswesen und im Handel.<sup>1</sup> Ob und wie weit es auch im Gesundheitswesen, insbesondere in der klinischen Nutzung von KI-Systemen, möglicherweise zu Diskriminierungsrisiken kommen kann und wie ihnen zu begegnen wäre, ist Gegenstand dieses Beitrags. Hierzu soll zunächst eine Definition von Diskriminierung gegeben und ihre Relevanz für ärztliche KI-Anwendungen bestimmt werden. Anhand von Praxisbeispielen werden verschiedene Typen und Gegenstandsbereiche potentieller Diskriminierungen vorgestellt. Diese werden im nächsten Teil reflektiert und in größere gesellschaftliche Zusammenhänge von sozialer Ungleichheit und gesellschaftlicher Entsolidarisierung gestellt. Am Ende werden unterschiedliche Regulierungsansätze wie die KI-Ethik und professionelle Selbstregulierung, Zertifizierungsmechanismen und der KI-Verordnungsvorschlag der Europäischen Kommission aufgezeigt.

## 2. Zur Definition von Diskriminierung

Der Begriff *Diskriminierung* leitet sich vom lateinischen Verb *discriminare* ab und bedeutet zunächst ‚trennen‘, ‚absondern‘, ‚abgrenzen‘, ‚unterscheiden‘. Im Kontext von Fachtermini der Informatik und KI, insbesondere im angelsächsischen Raum, wird der Begriff auch in dieser wertneutralen Form verwendet. Für den Bereich der KI ist Diskriminierung im Sinne von ‚Unterscheidung‘ und ‚Differenzierung‘ sogar essentiell und gehört zu ihren wesentlichen Funktionen.<sup>2</sup> Hier sollen beispielsweise mittels Mustererkennung pathologische von gesunden Gewebestrukturen in bildgebenden Verfahren unterschieden werden.

Im rechtlichen und gesellschaftlichen Kontext ist Diskriminierung hingegen negativ konnotiert und mit normativen Werturteilen verbunden. Im vorliegenden Beitrag wird Diskriminierung als ungerechtfertigte oder gesellschaftlich unerwünschte Benachteiligung in Handlungen, Praktiken oder gegebenenfalls durch algorithmische Systeme erfolgte Maßnahmen verstanden, die Personen aufgrund der Zuordnung bestimmter Merkmale und bzw. oder der Zugehörigkeit zu bestimmten Gruppen zuteil wird.<sup>3</sup> Zudem beruht die Auferlegung eines Nachteils auf der Gruppenzuschreibung der Betroffenen. ‚Gruppe‘ kann im Sinne einer sozial signifikanten Gruppe verstanden werden oder im Sinne einer Einteilung in verschiedene Kategorien, Klassen oder Segmente im Rahmen von algorithmischen Prozessen erfolgen; insbesondere im letzteren ist die ‚Gruppen‘-Zugehörigkeit häufig dynamisch und in zeitlicher Hinsicht variierend.<sup>4</sup> In Art. 3 Abs. 3 des Grundgesetzes und in Art. 21 der Grundrechte-Charta werden bestimmte geschützte Merkmale aufgeführt, für die rechtliche Diskriminierungsverbote bestehen. Auch in Art. 9 der EU-Datenschutzgrundverordnung (DSGVO) gibt es eine Liste besonders sensibler Kategorien personenbezogener Daten, deren Verarbeitung unter speziellen Schutz gestellt ist. Das Adjektiv ‚ungerechtfertigt‘ verweist bereits darauf,

dass es bei Diskriminierung um Fragen der sachlichen Gründe und ihrer vernünftigen Rechtfertigung sowie um Werturteile geht. Nicht jede Ungleichbehandlung ist als diskriminierend zu bewerten. Im Hintergrund stehen also Fairness- und Gerechtigkeitsmaßstäbe, die allerdings in gesellschaftlichen und kulturellen Kontexten und nach moralisch-ethischen Präferenzen variieren können. Gerechtigkeit kann unterschiedlich ausgelegt werden – beispielsweise als Chancengleichheit (*equality*) oder im Sinne einer Gleichstellungspolitik (*equity*) zum Ausgleich historisch oder sozial bedingter Ungleichheit. Zu unterscheiden ist zwischen absichtlicher und unabsichtlicher Diskriminierung.

Rechtlich unterschieden wird im Allgemeinen Gleichbehandlungsgesetz (AGG) ferner zwischen einer direkten oder unmittelbaren Benachteiligung (§ 3 Abs. 1 AGG), womit eine weniger günstige Behandlung einer Person gegenüber einer anderen in einer vergleichbaren Situation erfasst wird, gegenüber einer mittelbaren Benachteiligung (§ 3 Abs. 2 AGG), die eine Benachteiligung durch scheinbar neutrale Vorschriften, Maßnahmen, Kriterien oder Verfahren, die sich faktisch diskriminierend auswirken, bezeichnet. In den jeweiligen englischsprachlichen Entsprechungen bezieht sich der Begriff des ‚*disparate treatment*‘ auf die Handlung selbst, während ‚*disparate impact*‘ auf die (indirekten bzw. systematischen) Handlungsfolgen verweist.<sup>5</sup>

### 3. KI und Diskriminierung

Landläufig wird davon ausgegangen, dass Menschen Vorurteilen, Stereotypen, Stimmungen und kognitiven Verzerrungen bei der Urteilsbildung unterliegen. Maschinen hingegen gelten als neutral und ihnen wird eine gewisse Objektivität zugesprochen. Optimistische Fachleute erhoffen sich von datengetriebener KI eine bessere und effizientere Medizin und verleihen KI-Systemen epistemische Autorität.<sup>6</sup> Allerdings ist die Objektivitäts-Annahme längst widerlegt, denn auch Maschinen produzieren nicht nur Fehler, sondern können bei ihrer Nutzung auch verschiedene Formen von Ungleichheit und Benachteiligung reproduzieren und sogar verschärfen. Im Kontext von KI wird hierzu seit längerem eine Diskussion um Diskriminierung unter dem Stichwort *Bias* (Verzerrungen) geführt.<sup>7</sup>

KI ist ein Sammelbegriff für ‚Fähigkeiten‘, welche maschinenlernende Systeme aufweisen. Beim Begriff KI in der klinischen Anwendung beziehe ich mich im Folgenden vor allem auf KI-gestützte *Clinical Decision Support Systems* (CDSS), worunter IT-basierte Systeme und Verfahren zur ärztlichen Entscheidungsunterstützung gefasst werden, die in weniger oder stärker automatisiertem Maße ärztliche Handlungen und Entscheidungen vorstrukturieren und Empfehlungen ausgeben. KI-basierte Systeme können rein softwaregestützt in einer virtuellen Umgebung arbeiten (z. B. Sprachassistenten, Bildanalysesoftware, Sprach- und Gesichtserkennungssysteme), oder auch in Hardware-Systeme eingebettet sein (z. B. moderne Roboter). Sie können sich durch sensorische und andere In- und Output-Modalitäten (wie Text, Bilder, Videos, Sprache, Diagramme und Steuersignale) vollziehen. Die Basisfunktionen des Maschinenlernens beruhen auf großen Mengen von Datensätzen (*Big Data*) als sog. Trainingsdaten, die in der Trainingsphase zur mathematischen Modellbildung durch Korrelationen und Mustererkennung führen und eine Vielzahl von Variablen und Relationen untereinander und entsprechende Gewichtungen beinhalten. In der Anwendungsphase werden die erkann-

ten Modelle gespeichert, in Entscheidungsregeln eingebaut und auf neue Entscheidungssituationen übertragen. Diese müssen in iterativen und rekursiven Verfahren validiert werden.<sup>8</sup>

Die klinischen ‚Entscheidungsunterstützungssysteme‘ sind von solchen Systemen abzugrenzen, die automatisiert ‚eigenständige‘ Entscheidungen treffen. Die Rechtmäßigkeit deren klinischen Einsatzes in Deutschland ist unklar (siehe Art. 22 DSGVO). Hardware oder Software gelten als *Automated Decision Making System (ADM System)*, wenn sie durch einen Algorithmus eine Bewertung einer Situation oder eines Menschen vornehmen oder eine Vorhersage über die Wahrscheinlichkeit des Eintretens einer Situation treffen, daraufhin eine Software oder Hardware aktiviert wird, die auf Grundlage der Bewertung oder Prognose eine Entscheidung trifft, deren Handlung unmittelbar das Leben eines Menschen betrifft.<sup>9</sup>

Klinische Anwendungsfelder für CDSS gibt es in diagnostischen, therapeutischen, prognostischen sowie prädiktiven Zusammenhängen. In der Diagnostik betrifft dies beispielsweise die radiologische Bildgebung in Neurologie oder Onkologie, oder die Beurteilung der Malignität von Hautläsionen in der Dermatologie. Bei der Therapie geht es vor allem um Behandlungsunterstützung durch Wissensaggregation. Ein in ethischer Hinsicht kontroverses Feld stellt der Einsatz von CDSS zur klinischen Prognose dar. Dabei sollen mittels künstlicher neuronaler Netzwerke und unter Einbezug sowohl individueller wie bevölkerungsbasierter Daten Risiken für unerwünschte Zwischenfälle (wie etwa kardiovaskuläre Ereignisse) oder die Überlebensdauer von Patienten mit Krebs oder einer terminalen Niereninsuffizienz prognostiziert werden. Diese sollen bestimmte Therapie-Aufnahme oder -Abbruchentscheidungen orientieren. Befürchtet werden dabei zukünftige gesundheitsökonomische Parameter beim Einsatz. Noch einen Schritt weiter geht der Einsatz von CDSS bei der Prädiktion von Krankheiten, welche bei gesunden Menschen ansetzt und Dispositionen und Suszeptibilitäten für bestimmte Erkrankungen aufdecken will.

CDSS können zudem bezüglich ihrer technischen Funktionsweise näher differenziert werden. Sämtliche, auch die komplexeren, Automatisierungsverfahren basieren auf dem Einsatz von Algorithmen. Einerseits bezeichnet dies regelhafte, in einer Software programmierte Handlungsanleitungen (Wenn-Dann-Aussagen), die einem System genau beschreiben, was in einer bestimmten Konstellation zu tun ist. Zum anderen basiert ein wachsender Teil von CDSS auf ‚lernenden‘ Systemen, die nicht nur vorgegebene Handlungsanweisungen ausführen, sondern auch neue Handlungsvollzüge aus eingegebenen Trainingsdaten und deren Verarbeitung ‚lernen‘. Techniken wie maschinelles Lernen, oder auf künstlichen neuronalen Netzwerken beruhendes *Deep Learning* bezeichnen dabei komplexe Verfahren, wie ein KI-System ‚lernt‘, Probleme zu lösen, die nicht genau spezifiziert werden können oder deren Lösungswege nicht durch symbolische Argumentationsregeln genau beschrieben werden können. Dazu zählen das Erkennen von Bildern, Sprache und Texten, das automatisierte Schlussfolgern aus unstrukturierten Daten, das bis hin zu Verhaltens- oder Ergebnisvorhersagen liefert. Bei vielen dieser KI-Systeme können das ‚Wie‘ und ‚Warum‘ eines durch KI gefundenen Lösungsweges und einer automatisierten Entscheidungsempfehlung kaum oder gar nicht von außen nachvollzogen werden. Das Phänomen, dass die internen Verarbeitungsmechanismen zwischen Input und Output eines KI-Systems nicht nachvollziehbar sind, bezeichnet man als ‚*Black-Box*‘ oder Opazität. Neben diesen funktionalen und epistemi-

schen Gründen der Intransparenz solcher Systeme ist ein weiterer Grund, dass diese häufig als Betriebs- oder Geschäftsgeheimnisse oder durch andere Formen geistigen Eigentums geschützt werden. Wenn sich Fehler oder rechtswidrige Praktiken einschleichen, können diese daher nur schwer erkannt und korrigiert werden.<sup>10</sup> Erschwerend für eine Prüfung und Validierung kommt hinzu, dass es sich bei KI nicht um statische, sondern um dynamische Systeme handelt, deren Ergebnisse durch neue Datensätze oder algorithmische Veränderungen fortlaufend variieren können.

Zu Verzerrungen, Voreingenommenheiten und potentiellen Diskriminierungen kann es in unterschiedlichen Phasen der Anwendung von KI-basierten CDSS kommen. Zunächst bereits bei der Datenerhebung und bei den Trainingsdaten, etwa wenn bestimmte Gruppen von Menschen mit bestimmten Charakteristika unter- oder überrepräsentiert werden. Sodann bei der Modellbildung, wenn diese in rekursiven Schleifen auf bestimmte Ziele hin adjustiert werden. Und schließlich bei der Anwendung in bestimmten Praxiskontexten.

Bisher sind viele KI-Systeme unzureichend wissenschaftlich validiert. So waren beispielsweise einer in *The Lancet Digital Health* publizierten Metaanalyse zufolge nur bei weniger als einem Prozent von über 20.000 Bilddiagnostik-Studien, welche die diagnostische Genauigkeit zwischen Menschen und *Deep Learning*-Algorithmen verglichen, die Ergebnisse extern validiert und blieben in der Studiendarstellung ohne qualitative Mängel.<sup>11</sup> Dies zeigt, dass nicht nur größere Studienpopulationen, sondern auch qualitativ bessere, insbesondere prospektive Studien benötigt werden, um die klinische Validität der CDSS in der Praxis beurteilen zu können.<sup>12</sup>

Grob schematisierend kann eine algorithmische Diskriminierung im klinischen Kontext folgende Wirkungen haben: 1.) einen Behandlungsfehler, 2.) eine Unterbehandlung oder 3.) eine Überbehandlung von Menschen, die einer bestimmten, als relevant angesehenen Gruppe bzw. Kategorie mit bestimmten Merkmalen angehören. Die ungerechtfertigte Benachteiligung kann also darin bestehen, dass einem Menschen medizinische Dienste vorenthalten oder von Krankenversicherungen nicht finanziert werden, oder dass bestimmte Leistungen nur gegen einen unfairen Preis oder bei Privatversicherten durch höhere Einstufung bereitgestellt werden. Sie kann aber auch eine medizinisch nicht indizierte und unnötige Behandlung oder eine falsche Prognose oder Prädiktion implizieren.<sup>13</sup>

Im Folgenden soll anhand einiger Anwendungsbereiche und Studien exemplarisch aufgezeigt werden, zu welchen Diskriminierungen es beim Einsatz von KI-basierten CDSS kommen kann. Dies soll auch dazu dienen, genauere Einblicke in verschiedene Problemdimensionen zu erhalten.

#### 4. Beispiele und Gegenstandsbereiche von Diskriminierung in KI-basierten CDSS

Diskriminierung in KI-Systemen kann sich Plough und Holm zufolge auf drei verschiedene Weisen vollziehen: Erstens können Voreingenommenheiten den Algorithmen inhärent sein, welche die KI-Systeme bei ihrer Diagnose und Behandlungsplanung leiten. Zweitens kann es zu Verzerrungen kommen, wenn die Systeme menschliche Entscheidungsfindung imitieren, und zwar schlicht deshalb, weil diese Entscheidungen gelegentlich voreingenommen sein können, z.B. durch das Training der KI-Systeme auf Basis

von Daten, die von Menschen markiert wurden. Ärzte können für Diskriminierung anfällig sein und wenn das System mit einer Reihe medizinischer Fälle trainiert wird, in denen beispielsweise Altersdiskriminierung auftritt, kann das System am Ende diese Voreingenommenheit reproduzieren. Drittens könne die Entscheidungsfindung von KI-Systemen ungewollt zu Verzerrungen führen. Man stelle sich etwa vor, ein KI-System wurde programmiert, Entscheidungen zu favorisieren, die das Ergebnis der Gesundheitsversorgung maximieren. Die bestmögliche Behandlung erfordere dabei ein hohes Maß an Patienten-Compliance. Nun gibt es Hinweise darauf, dass die Patienten in der Altersgruppe der 60- bis 70-Jährigen die höchste Compliance, jene mittleren Alters hingegen die niedrigste aufweisen. Eine KI, die lediglich darauf abzielt, den *Outcome* zu maximieren, kann unter Nutzung der genannten, stark gewichteten Evidenz den Patienten mittleren Alters am Ende nicht die bestmögliche Behandlung empfehlen, sondern eine andere, und somit de facto diese mittleren Altersgruppen diskriminieren.<sup>14</sup> Dieses fiktive Beispiel macht deutlich, dass den Algorithmen vorgegebene oder durch sie ermittelte Effizienzziele nicht notwendigerweise den Standards medizinischer Ethik gehören.

Verschärft wird das Problem der Verzerrungen durch den Black-Box-Charakter vieler KI-Systeme. Wenn Gewichtungen der vielen eingehenden Variablen und die Algorithmen nicht verstanden und nachvollzogen werden können, sind Voreingenommenheiten und Fehler schwer nachzuweisen. Außerdem gibt es meistens einen Zielkonflikt zwischen Genauigkeit (*accuracy*) und Fairness in KI-Systemen; verschiedene Methoden zur Abschwächung von Verzerrungen haben unterschiedliche (negative) Auswirkungen auf die Genauigkeit.<sup>15</sup>

#### 4.1 Gender – Differenzierung und Diskriminierung

Eine statistische Analyse kann beispielsweise aufzeigen, dass ein System unterschiedliche Behandlungsempfehlungen für weibliche und männliche Patienten gibt. Dies kann behandlungsrelevante Gender-Unterschiede für bestimmte Erkrankungen widerspiegeln und damit medizinisch absolut angemessen sein.<sup>16</sup> Es kann aber auch auf mögliche Verzerrung in den Trainingsdaten zurückgehen und eine mögliche Diskriminierung des einen oder des anderen Geschlechts darstellen.

Erwachsene, weiße Männer sind in den bestehenden medizinischen Datensätzen stark überrepräsentiert, zu Ungunsten der Daten von weißen Frauen sowie Kindern und Menschen jeden Alters aus anderen ethnischen Gruppen. Dieser Mangel an Diversität in den Daten führt wahrscheinlich zu verzerrten Algorithmen.<sup>17</sup> Da bei der KI-Entwicklung auf die vorhandenen historischen Datenbestände zurückgegriffen wird, werden genderbezogene Ungleichheiten nicht nur von der analogen in die digitale Welt transferiert, sondern auch von der Vergangenheit in die Zukunft extrapoliert. Um Behandlungsdaten aus Patientenakten für die Modellbildung im Maschinenlernen verfügbar zu machen, müssen diese aufbereitet, interpretiert und standardisiert werden. Damit fließen die Werte und Entscheidungen derjenigen, die Klassifikationen und Interpretationen vornehmen, in die Modellbildung ein. Der geschilderte Zustand von Asymmetrien wird womöglich durch den Gender Gap in der KI-Forschung und -Entwicklung weiter forciert. Denn Algorithmen sind nicht nur das Resultat der Datensätze und Modelle,



sondern auch geprägt von den Menschen, die sie programmieren – und dies sind überwiegend weiße und asiatische Männer.<sup>18</sup>

#### 4.2 Individuum als stratifiziertes Gruppenelement

Es ist wichtig, zu begreifen, dass datengetriebene CDSS Klassen und Kategorien von Patienten und Krankheitstypen bilden, die nach Scores, Rangfolge und Ratings geordnet werden. Sie können durch Korrelationen neue Muster erkennen. Allerdings sind solche Korrelationen nicht als Kausalität misszuverstehen, sondern sie bedürfen einer klinischen Validierung. Wenn die Empfehlungen oder Ergebnisse eines Maschinenlernens systematisch nicht erklärt und nachvollziehbar gemacht werden können (Opazität), stellen sich besondere Fragen an die Validität der Ergebnisse und mögliche Diskriminierungspotentiale. Hypothesenfreie Forschung an Datensätzen kann allenfalls Korrelationen erkennen, denen weiter als Hypothese wissenschaftlich nachgegangen werden kann, es besteht aber die Gefahr von Scheinkorrelationen und statistischen Artefakten ohne echte Relevanz.

Mittels von Algorithmen und Modellen in KI-Systemen werden Korrelationen und Muster zwischen verschiedenen Variablen identifiziert und Vorhersagen über das Verhalten auf Gruppenebene getroffen. Das heißt bei Patienten, dass sie mit statistisch ‚zusammenpassenden‘, weil ähnliche gemeinsame Merkmale aufweisenden Kategorien in einer Klasse oder einem Segment eingeordnet werden und diese hierarchisch stratifiziert werden. Damit können beispielsweise abgestufte Risikoprofile für bestimmte Erkrankungen oder Krankheitsverläufe gebildet werden. Gleich ob diese Risikoprofile dynamisch oder statisch sind, so bleibt festzuhalten, dass das Individuum stets auf der Grundlage von Verbindungen mit anderen verstanden wird, die durch den Algorithmus identifiziert werden, nicht durch sein tatsächliches individuelles Verhalten. Das Individuum wird daher nicht im Sinne einer Einzelfallentscheidung betrachtet, sondern kommt lediglich als stratifiziertes Gruppenelement in den Blick der KI, was Fehlschlüsse durch Generalisierungen nach sich ziehen kann.<sup>19</sup>

#### 4.3 KI und ethnischer Bias

Ein klinisches KI-Entscheidungsunterstützungs-Programm, das auf Lungenscans trainiert wurde, mag neutral erscheinen, aber sofern die Trainingsdatensätze überwiegend Bilder von Patienten eines bestimmten Geschlechts oder einer bestimmten ethnischen Gruppe enthalten, kann es Gesundheitszustände in anderen Bevölkerungsgruppen übersehen oder falsch diagnostizieren. KI-Programme in der Dermatologie beispielsweise, die Hautkrebs diagnostizieren sollen, können bei Menschen dunklerer Hautfarbe hohe Fehlerraten aufweisen, da sich die Daten aus jahrelanger klinischer Forschung, die zum Trainieren der Programme verwendet werden können, hauptsächlich auf Menschen mit heller Haut konzentrieren.<sup>20</sup> Generell hat auch Gesichtserkennungs-Software wesentlich niedrigere richtige Erkennungsraten bei Frauen und Dunkelhäutigen.<sup>21</sup> Sofern die Entscheidungsprozeduren des Systems lediglich rein mathematisch verstanden werden können, wird es sehr schwierig sein, richtige und angemessene geschlechts- oder

ethnisch bezogene Behandlungsempfehlungen und solche, die problematisch, weil verzerrt sind, auseinander zu halten.

Dieses Beispiel macht deutlich, dass alters-, ethnisch- oder geschlechtsbezogene Unterscheidungen dann eine sachlich angemessene Differenzierung darstellen, wenn sie tatsächlich valide und medizinisch relevant sind. Sofern sie aber lediglich auf Unterrepräsentanzen etwa von Frauen in den Trainingsdaten zurückgehen, können sie etwa bei der Bestimmung von Medikamentendosierungen zu fehlerhaften und potentiell gefährdenden Schlussfolgerungen gelangen.<sup>22</sup>

#### 4.4 ‚Race‘: Biologische und genetische Unterschiede oder Ausdruck sozio-ökonomischer Ungleichheit?

Um die Gesundheitsversorgung möglichst effektiv zu gestalten, kommt in den USA eine KI zum Einsatz, welche Patienten mit besonderem Pflegebedarf identifiziert. Eine vielzitierte Studie von Obermeyer u. a. zeigt allerdings auf, dass Afroamerikaner bei gleicher Krankheitsschwere seltener für die Aufnahme in besondere *Care-Management*-Programme vorgeschlagen wurden als Weiße. Tatsächlich hätte eine KI ohne Bias die doppelte Anzahl von Afroamerikanern mit besonderem Pflegebedarf ausweisen müssen. Die Autoren sehen das Problem darin, dass der Algorithmus dieser KI den Risikoscore der Patienten als die zu erwartenden Kosten der weiteren Behandlung auf Basis früherer Kosten im Gesundheitssystem fasste. Da aufgrund diverser Faktoren in den USA weniger Geld für die Behandlung von Afroamerikanern ausgegeben wird, wurden die Kosten und somit auch das Pflegerisiko für Angehörige dieser ethnischen Gruppe als geringer eingeschätzt.<sup>23</sup>

Sehr kritisch betrachten Vyas u. a. auch die Praxis, statistische ‚Korrekturen‘ für ethnische Zugehörigkeit oder ‚race‘ in medizinische Algorithmen für viele Krankheiten einzuführen, wofür sie eine Reihe von Studien heranziehen. Sie argumentieren, dass trotz zunehmender Beweise, dass ‚Rasse‘ kein verlässlicher Proxy (Ersatzinformation) für genetische Unterschiede ist, sich der Glaube, sie seien es, jedoch zuweilen heimtückisch in der medizinischen Praxis eingebürgert habe. Dabei werde die Diversität in und zwischen den vermeintlichen ethnischen Gruppen negiert. Unterschiede, die als ‚racial‘ wahrgenommen werden, seien oft sozioökonomisch oder genetisch bedingt, und daher sollten Patienten nicht durch ihre Hautfarbe, sondern durch die Analyse ihres Genoms oder Faktoren wie Stress, Ernährung und Gesundheitsgewohnheiten wahrgenommen werden. Die Algorithmen seien auch nicht in der Lage, rassische Mehrdeutigkeiten oder Komplexität zu erkennen.<sup>24</sup> Wenn Ärzte diese Algorithmen verwendeten, um die Risikobewertung zu individualisieren und klinische Entscheidungen zu treffen, bestünde die Gefahr einer rassifizierten Medizin. „Wenn Kliniker jedoch die Rasse in ihre Instrumente einbeziehen, riskieren sie, rassische Ungleichheiten als unveränderliche Fakten zu interpretieren und nicht als Ungerechtigkeiten, die ein Eingreifen erfordern. Forscher und Kliniker müssen zwischen der Verwendung von Rasse in deskriptiven Statistiken, wo sie eine wichtige Rolle in epidemiologischen Analysen spielt, und in präskriptiven klinischen Richtlinien, wo sie Ungleichheiten verschärfen kann, unterscheiden.“ Sie fordern daher die Kliniker auf, Vorsicht walten zu lassen, wenn sie sich in der medizinischen Forschung auf die ‚Rasse‘ berufen: „In vielen Fällen wird diese Beurteilung weitere For-

schung über die komplexen Wechselwirkungen zwischen Abstammung, Rasse, Rassismus, sozioökonomischem Status und Umwelt erfordern. [...] Aber Kliniker können aufmerksam und bewusste Anwender sein. Sie können erkennen, ob die Korrektur die Ungleichheiten eher lindert oder verschlimmert“.<sup>25</sup>

#### *4.5 Fehlrepräsentation, Unter- und Überversorgung: Zum Zusammenhang von sozialer Ungleichheit und Diskriminierungsrisiken*

Die Integration von maschinellem Lernen in klinische Entscheidungshilfen, wie z. B. computergestützte Warnhinweise oder Diagnoseunterstützungen, könnte Ärzten und Pflegekräften gezielte und zeitnahe Informationen bieten, die klinische Entscheidungen verbessern können. Algorithmen des maschinellen Lernens können jedoch vielfältigen Verzerrungen unterliegen. Dazu gehören solche, die auf fehlenden oder lückenhaften Daten, auf falscher Stichprobengröße sowie auf Fehlklassifikationen und Messfehlern beruhen, oder auch aufgrund von fehlerhafter KI entstehen. Eine Verzerrung ist aber auch bei korrekten Daten und fehlerfreiem KI-System möglich, nämlich dann, wenn die genutzte Datengrundlage selbst einseitig geprägt ist. Dies geschieht etwa, wenn bestimmte Gruppen von Menschen (beispielsweise Frauen, Kinder, Geflüchtete und Migranten, ethnische Minoritäten oder Menschen, die aus sozio-ökonomischen Gründen geringeren Zugang zum Gesundheitssystem haben, Menschen mit Behinderungen, Menschen mit seltenen Erkrankungen, trans- und intersexuelle Menschen) zu wenig oder unangemessen in den zugrunde gelegten Daten repräsentiert sind.<sup>26</sup>

Als Zwischenfazit bleibt festzuhalten, dass KI-gestützte CDSS nur so gut sein können wie die Trainingsdaten, auf die sie ihre Berechnungen und Mustererkennungen stützen. KI-Systeme können folglich zu systematisch verzerrten oder falschen Ergebnissen kommen, die bestimmte Gruppen von Menschen benachteiligen, ohne dass dies sachlich gerechtfertigt ist. „Bestehende Ungleichheiten in der Gesundheitsversorgung sollten nicht durch gedankenloses oder übermäßiges Vertrauen in Maschinen verstärkt werden“, warnen Gianfrancesco u. a. und diskutieren Maßnahmen, die ergriffen werden können, um Probleme zu vermeiden.<sup>27</sup>

Dies betrifft auch die Profilbildung bestimmter Segmente von Patienten, die nicht (allein) einer sozial signifikanten Gruppe (z. B. nach Alter, Geschlecht, ethnischer Zugehörigkeit) zugeordnet werden, sondern die als weniger gesund, mit höheren Gesundheitsrisiken behaftet oder weniger ‚compliant‘ eingestuft werden. Auch für diese besteht die Gefahr, dass sich strukturelle Diskriminierung und Stigmatisierung bei der medizinischen Behandlung fortsetzt oder sogar durch KI-basierte CDSS verstärkt wird.<sup>28</sup> Gerade in Gesundheitssystemen, die nicht wie in der Bundesrepublik durch eine allgemeine Pflicht-Krankenversicherung die Risiken von Erkrankungen abdecken, sind sozioökonomische Disparitäten in der Gesundheitsversorgung besonders ausgeprägt. Diese schlagen sich aber auch in der Datenbasis nieder und können zu einer Perpetuierung von strukturellen Ungleichheiten beitragen, sofern keine Korrekturen veranlasst werden.

Doch auch in Deutschland bestehen bekanntlich erhebliche soziale Unterschiede in Gesundheit, Mortalität und Lebenserwartung, wie die Sozialmedizin feststellt. Menschen mit niedrigem Einkommen, Berufsstatus und Bildungsniveau haben ein erhöhtes

Risiko für chronische Erkrankungen wie Diabetes und chronische Bronchitis sowie für krankheitsbedingte funktionelle Einschränkungen in ihrer Alltagsgestaltung und Lebensqualität.<sup>29</sup> Die sozialen Disparitäten zeichnen sich auch im individuellen Gesundheitsverhalten und bei verhaltensbezogenen Risikofaktoren wie Rauchen, Bewegungsmangel, Übergewicht und Bluthochdruck deutlich ab. Letztlich kumulieren sie in einer verkürzten Lebenszeit in sozial benachteiligten Bevölkerungsgruppen. Daten des sozio-ökonomischen Panels (SOEP) der Jahre 1992 bis 2016 zufolge sterben 13 % der Frauen und 27 % der Männer aus der niedrigsten Einkommensgruppe bereits vor Vollen- dung des 65. Lebensjahres, während dies in der höchsten Einkommensgruppe lediglich auf 8 % der Frauen und 14 % der Männer zutrifft. Die Differenz zwischen der niedrigsten und höchsten Einkommensgruppe beträgt bezogen auf die mittlere Lebenserwartung bei Geburt bei Frauen 4,4 Jahre und bei Männern 8,6 Jahre.<sup>30</sup> Dies bedeutet umgangs- sprachlich auf den Punkt gebracht: Wohlhabende leben länger und sterben gesünder.

Gleichzeitig stellte der Sachverständigenrat im Gesundheitswesen (SVR) schon 2000/2001 eine Überversorgung bei privat versicherten Patienten und eine Unterversorgung bei weniger gesunden gesetzlich Versicherten fest. Auch in jüngeren Gutachten kam der SVR zu dem Schluss, dass es in Deutschland weiterhin Über-, Unter- und Fehl- versorgung sowie erhebliche Steuerungsdefizite gibt. Überversorgung bedeutet, dass Patienten Leistungen erhalten, die für sie unnötig sind und Folgebehandlungen und Komplikationen ihnen sogar schaden können.<sup>31</sup>

Wenn solche Ungleichgewichte im Gesundheitssystem nicht kritisch reflektiert wer- den, sondern ungefiltert und unkorrigiert in die Trainingsdaten und Modellbildung der KI-Systeme eingehen, dann besteht das Risiko, dass diese aus der Vergangenheit für die Zukunft als Standard fortgeschrieben werden. So hält auch der SVR in seinem jüngsten Gutachten zur Digitalisierung für das Gesundheitswesen fest, maschinelles Lernen be- nötige „valide Trainingsdatensätze, die Daten in ausreichender Menge und Qualität ent- halten. Fehlerhafte, unvollständige oder verzerrte Datensätze führen zu irrigen Resulta- ten der Lernverfahren. Zudem können Ergebnisse zuverlässig nur auf solche Populationen übertragen werden, die in ihrer Struktur im Trainingsdatensatz abgebildet sind.“ Daher benötigt man eine „repräsentative, strukturierte und kontrollierte deutsche Gesund- heitsdatenbasis“.<sup>32</sup>

Die Gefahr einer Überdiagnostik und Überbehandlung nach CDSS-Empfehlung kann jedoch andere Gründe haben, wenn durch KI etwa mögliche Anzeichen für eine Erkrankung aufgespürt werden, deren Manifestation aber statistisch gesehen sehr un- wahrscheinlich und bzw. oder sehr unplausibel ist, von Ärzten aber eine entsprechende invasive Diagnostik oder Therapie vorsichtshalber, etwa auch zur eigenen Absicherung, veranlasst wird.<sup>33</sup>

#### *4.6 Riskante Fehlentscheidungen aufgrund fehlenden Kontextwissens*

Ein oft zitiertes Beispiel für Fehlentscheidungs-Empfehlungen bei CDSS stellt das auf einem neuronalen Netz basierende KI-System dar, welches das Risiko der Sterblichkeit für Patienten mit Lungenentzündung vorhersagen, und damit Entscheidungen, diese entweder ambulant oder stationär zu behandeln, unterstützen sollte. Das System ermit- telte eine höhere Überlebenschance für Patienten mit Asthma, was allerdings konträr

zum medizinischen Erfahrungswissen stand. Die Verzerrung im Trainingsdatensatz ergab sich daraus, dass Patienten mit Lungenentzündung und einem bestehenden Asthmaleiden direkt auf stationäre Intensivpflege-Stationen gebracht worden waren. Dadurch wurden für Patienten mit Lungenentzündung und Asthma bessere Behandlungsergebnisse erzielt, als für solche, die ‚lediglich‘ eine Lungenentzündung hatten. Obwohl das System also diesbezüglich korrekte Prognosen traf, kam es zu falschen Empfehlungen, weil die genannten Kontextinformationen nicht in das CDSS einbezogen wurden. Hätten Behandelnde sich allein auf das CDSS verlassen, hätten sie die am stärksten gefährdeten Patienten fälschlicherweise zur ambulanten Behandlung nach Hause entlassen.<sup>34</sup> Dieses Fallbeispiel verdeutlicht die Probleme, die entstehen, wenn relevante, aber nicht in den entsprechenden Datensätzen repräsentierte Kontextfaktoren bei Entscheidungen unbeachtet bleiben. Sie sind gegebenenfalls als indirekte ungerechtfertigte Benachteiligung zu charakterisieren.

#### 4.7 Stimm- und Spracherkennung als Diagnostik: Neurodiskriminierung

Eine Fülle von Studien versucht, mittels Stimmdiagnostik und Maschinenlernen ‚Stimmen-Biomarker‘ für eine Vielzahl von Erkrankungen zu identifizieren, darunter Demenz, Depression, Parkinson, Autismus-Spektrum-Störungen und sogar Herzkrankheiten.<sup>35</sup> Auch hier kann es zu Störungen und Fehlern kommen, etwa wenn das KI-System Menschen mit Dialekt oder fremdsprachigem Akzent falsch oder pathologisierend diagnostiziert.<sup>36</sup> *Winterlight Labs*, ein in Toronto ansässiges Unternehmen, programmierte ein KI-Tool zur Identifizierung von Personen mit Alzheimer-Erkrankung anhand kurzer Sprachproben. Offensichtlich war die Technologie nur für englische Muttersprachler eines bestimmten kanadischen Dialekts ausgelegt. Sie fehlinterpretierte Pausen, falsche Aussprache, Unklarheiten und Unsicherheiten von Nicht-Muttersprachlern als Anzeichen für kognitiven Verfall.<sup>37</sup>

Wenn in einer weiteren Studie behauptet wird, KI-Algorithmen könnten bei der Analyse des Gebrabbers von zehn Monate alten Säuglingen mit 80-prozentiger Genauigkeit eine später diagnostizierte Autismus-Spektrum-Störung bei den Kindern vorhersagen<sup>38</sup>, so stellen sich hier allerdings eine Fülle von Fragen nicht nur zur Validität der Erkenntnisse, sondern auch zur Rechenschaftslegung gegenüber der Öffentlichkeit. Denn solche prädiktiven Diagnosen können stigmatisierend wirken und einen Schatten auf die Zukunft der Kinder legen, zumal, solange keine gezielte Therapie entsprechender Erkrankungen besteht.<sup>39</sup>

Andere Studienteams haben stimmliche Biomarker identifiziert, die mit Psychosen, Suizidalität und bipolaren Störungen in Verbindung gebracht werden. Solche prognostischen Diagnosen von psychischen Erkrankungen, insbesondere im Kinder- und Jugendalter, können nicht nur medizinische Implikationen haben, sondern die schulische und berufliche Laufbahn enorm beeinflussen. Auch können sie schwerwiegende Eingriffe in die Privatsphäre darstellen. Denn durch Maschinenlernen gestützte Vorhersagen bergen Gefahren für Personen, die gar keinen Bezug zu den zugrundeliegenden Trainingsdaten oder dem Modell selbst haben, sondern als kollektive Schädigungspotenziale zu qualifizieren sind, wenn sie Eingang in breite Anwendungen, gegebenenfalls sogar ohne Wissen und Zustimmung der Anwendenden, finden. Hier bestehe das Risiko einer „Neuro-

diskriminierung<sup>40</sup>, insbesondere von vulnerablen Gruppen. Zum Beispiel gibt es KI-gestützte Modelle, die beanspruchen, Angst und Depressionen bei Kindern allein auf der Grundlage statistischer Muster in der Stimme des Kindes erkennen zu können. Mit diesem Modell können über Sprachaufnahmen sensible diagnostische Daten erhoben werden.<sup>41</sup> Jedes Kind, das in sprachgestützte Assistenten wie etwa Siri oder Alexa spricht, kann potentiell in seiner Privatsphäre betroffen sein. Solange es keine entsprechende Regulierung gibt, besteht auch die Gefahr, dass Versicherungen oder Arbeitgeber die Sprachanalysensysteme ohne ausdrückliche Zustimmung nutzen oder damit persönliche Gesundheitsinformationen einholen, was einer möglichen Diskriminierung von Kunden oder Mitarbeitern Tür und Tor öffnen könnte.<sup>42</sup>

#### 4.8 KI und Priorisierung in der Notfallmedizin: ‚Triage‘?

Einen weiteren *Use Case* zur KI-gestützten Priorisierung und Allokation von Gesundheitsleistungen bilden die Notfallambulanzen. Ein geplanter Gesetzentwurf des Bundesgesundheitsministeriums zum verbindlichen und standardisierten Einsatz eines Ersteinschätzungssystems für Notfallpatienten, das algorithmenbasierte „Triage“-Entscheidungen<sup>43</sup> vorsieht, hat diesbezüglich von der Deutschen Interdisziplinären Vereinigung für Intensiv- und Notfallmedizin (DIVI) und anderen Fachgesellschaften harte Kritik erfahren. Es gebe derzeit „kein wissenschaftlich validiertes Ersteinschätzungssystem, das automatisiert die Behandlungsdringlichkeit feststellen“ könne. Die automatisierte Priorisierungsentscheidung und „jede Fehleinschätzung aufgrund eines bloßen Abfragemoduls“ könne „fatale Folgen für den einzelnen Patienten haben“.<sup>44</sup> Auch der 124. Deutsche Ärztetag 2021 lehnt diesen Gesetzentwurf ab und betont, das Patienten-Vertrauen „würde massiv erschüttert, wenn Patienten zukünftig ohne ärztliche Abklärung ihrer Beschwerden allein aufgrund eines Software-Algorithmus abgewiesen werden können“. Die Steuerung in die richtige Versorgungsebene könne mit solchen Systemen nicht sicher gewährleistet werden. Patienten müssten sich aber darauf verlassen können, dass ihre individuellen Beschwerden ärztlich bewertet werden und die Patientensicherheit im Vordergrund steht.<sup>45</sup> Eine Gruppe von Wissenschaftlern, die in einem offenen Brief ihre Sorge artikulieren, verweisen darauf, dass die diesen Algorithmen zugrundeliegenden Parameter direkt oder indirekt auch durch die Behandlungskosten mitbestimmt werden. Insofern führten politische Entscheidungen über medizinische Kostendeckelungen zwangsläufig zu Parameteränderungen, die darüber bestimmten, „wieviele Menschen letztlich von Triage-Entscheidungen betroffen sind und früher sterben werden, auch, wenn das nicht beabsichtigt war.“ Neben den Behandlungskosten spielten aber auch Daten zur gesundheitlichen Verfassung einer Person eine Rolle, die zum Teil durch Verhalten, Arbeit und Umwelt, zum Teil aber auch genetisch bedingt seien. Sie befürchten, es sei „im Ergebnis zu erwarten, dass die Kombination von Kostenoptimierung und daten-basierter, algorithmen-gestützter Medizin verstärkt zu eugenischen Selektions-Effekten führt“.<sup>46</sup>

Auch wenn man nicht zu solch markigen Worten und Einschätzungen greifen mag, so bleibt doch festzuhalten, dass viele KI-Systeme inzwischen für Priorisierungsentscheidungen eingesetzt werden. Bei der Unterscheidung zwischen dringenden und weniger dringenden Fällen in Notfallambulanzen oder auf Intensivstationen sind medizinische

Unterschiede für eine Einschätzung der Dringlichkeit relevant, die noch nicht an sich eine Diskriminierung darstellen. Sie bergen jedoch ein Diskriminierungspotential, sofern sie auf verzerrten Datensätzen oder Algorithmen beruhen, oder wenn sie bestimmte Gruppen von Menschen systematisch niedriger bewerten.

Diesbezüglich wurde dem Bundesverfassungsgericht 2020 eine Klage von Menschen mit Behinderung vorgelegt, welche die sog. Triage in Notstandssituationen bei fehlenden Intensiv-Beatmungs-Betten wie in der Covid-19-Pandemie verhandelt. Die Kläger sind in Sorge, dass Prognose-Scores wie etwa der *Clinical Frailty Score* sie aufgrund ihrer behinderungsbedingten Beeinträchtigungen als mit einem schlechteren allgemeinen Gesundheitsstatus qualifizieren könnte. Daher würden sie unter Umständen wegen geringerer medizinischer Erfolgsaussichten als sie Menschen ohne Behinderung haben, trotz medizinischer Indikation und Einwilligung keine lebensrettende Behandlung erhalten, sondern lediglich eine ihr Sterben begleitende palliative Versorgung.<sup>47</sup> Der *Clinical Frailty Score* stellt ein – allerdings ethisch umstrittenes – klinisches Instrument dar, anhand dessen die langfristigen erwartbaren Überlebenschancen eines Patienten nach erfolgreicher Behandlung abgeschätzt werden.<sup>48</sup> Solche Scores sind keineswegs immer automatisierte, KI-basierte Entscheidungssysteme, sondern können grundsätzlich auch manuell oder mit einfachen Computersystemen ausgeführt werden. Sie beruhen jedoch auf Datensätzen früherer Notfallpatienten und statistischer Wahrscheinlichkeiten der Erfolgsaussichten des Überlebens. Diese können nicht nur in Notfallambulanzen oder Intensivstationen eingesetzt werden, sondern auch in anderen klinischen Bereichen, wie das folgende Beispiel zeigt.

#### 4.9 Todesalgorithmen

Von CDSS erstellte Prognosen zur Wahrscheinlichkeit des Überlebens bzw. Versterbens eines Patienten sollen auch ärztliche Entscheidungen zur Überleitung in die palliativmedizinische Pflege unterstützen. Solche Prognosen zum Todeszeitpunkt individueller Patienten gibt es sowohl bei akademisch entwickelten CDSS, wie etwa eines von der Stanford-Universität, das keine primär kommerziellen Interessen verfolgt. Es existieren aber auch kommerzielle Produkte, etwa des Start-Ups *Aspire Health, Inc.*, das von der Wagniskapital-Tochter „GV“ der Google-Mutter Alphabet finanziert und in 26 US-Bundesstaaten angeboten wird.<sup>49</sup> Dieses bietet auch palliativmedizinische Pflegedienstleistungen an. *Aspire Health* betont auf seiner Website, dass sich durch ihr CDSS auch Kosten einsparen ließen, wenn man mit einer infausten Prognose rechnen müsse.<sup>50</sup> Wie das CDSS über die Todesprognose entscheidet und inwiefern die Berechnung des Algorithmus für die behandelnden Ärzten transparent ist, bleibt offen: „Eine Validierung ihres Algorithmus bleibt *Aspire Health* bisher schuldig. Es gibt keine öffentlichen Studien darüber, wie hoch die Kosteneinsparungen tatsächlich sind, inwiefern eine Identifikation von Palliativpatienten durch den Algorithmus tatsächlich dazu beiträgt, den Zugang zur Versorgung zu verbessern, oder wie gut der Algorithmus auch die ‚richtigen‘ Patienten identifiziert“.<sup>51</sup>

Mögliche Ungleichbehandlungen oder Fehler lassen sich nicht eruieren, da *Aspire Health* die genaue Funktionsweise des Algorithmus nicht preisgibt. Hier wie auch in anderen Fällen besteht die Gefahr einer sich selbst erfüllenden Prophezeiung: Wenn

ein System, das auf Ergebnisdaten z. B. von Krebspatienten trainiert ist, eine schlechte Prognose vorhersagt, wird palliativ statt kurativ behandelt und damit die Empfehlung der CDSS verstärkt. Im Folgenden sollen einige Aspekte der aufgeworfenen Gegenstandsbereiche noch einmal auf einer anderen Ebene reflektiert werden.

#### *4.10 Zwischenfazit: Ebenen der Risiken von KI: Individuen, Gruppen und staatliche Institutionen*

Zusammenfassend lässt sich anhand der Beispiele und Gegenstandsbereiche erkennen, dass die epistemischen, normativen und übergreifenden ethischen Risiken, die mit dem Einsatz von KI in der Gesundheitsversorgung verbunden sind, nicht nur auf der individuellen, sondern auch auf der Gruppen-, institutionellen- und gesellschaftlichen Ebene ihre Wirkung zeitigen.

Morley und Floridi weisen diesbezüglich auf drei Ebenen des Schadenspotentials von KI-Systemen hin: Zum einen könnten diagnostische Bilderkennungssysteme oder klinische Entscheidungsunterstützungssoftware Einzelpersonen Schaden zufügen, etwa wenn seitens des Algorithmus der Schweregrad der Symptome einer Person falsch eingeschätzt wird. Zweitens können einzelne soziale Gruppen sowie die Gesellschaft als Ganze negativ davon betroffen sein, wenn voreingenommene Trainingsdatensätze zu unverhältnismäßig besseren oder schlechteren Gesundheitsergebnissen für verschiedene Gruppen von Menschen führen.

Auf der dritten, systemischen Ebene geht es um Institutionen des Gesundheitswesens, wenn Diskriminierungen, Verzerrungen und Undurchschaubarkeiten von KI-Systemen zu Problemen der Patientensicherheit führen, die von den Aufsichtsbehörden nicht angemessen behandelt werden können, was einen Verlust des öffentlichen Vertrauens zur Folge haben könnte.<sup>52</sup> Da die Diskriminierungs- und Schadenspotentiale zudem schwer aufzudecken und auf ihre jeweilige Ursache zurückzuführen sind, ist es jedoch schwierig, Verantwortlichkeiten und Haftung für diese Schäden zuzuschreiben und damit auch, deren erneutem Auftreten vorzubeugen.

#### *5. Dilemmata und Paradoxien: Mehr Daten, mehr Fairness – oder höhere Risiken?*

Aus der Analyse heraus, dass KI-Anwendungen bestehende Ungleichheiten fortsetzen oder gar verstärken könnten, wird oftmals die Forderung erhoben, die Datenlage zu verbessern und Diversität schon im KI-Entwicklungsprozess mitzubedenken. Denn manche gesellschaftlichen Minoritäten, die stigmatisierungs- und diskriminierungsgefährdet sind, wie etwa Transgender- oder Intersexpersonen, machen nur ca. 0,5–4,6 % der Bevölkerung aus.<sup>53</sup> Andere, wie etwa Geflüchtete ohne Papiere, Menschen ohne Krankenversicherung und andere sozial Benachteiligte begeben sich seltener in ärztliche Behandlung. Die Forderung, solche Minoritäten in die Datensätze für das Training des Maschinenlernens aufzunehmen und dort auch insgesamt mehr soziodemographische Informationen zu Ethnizität, Migrationsstatus, Einkommen, Bildungsgrad, psychischen Belastungen, etc. einzubringen, ist allerdings zumindest als zweischneidig zu bewerten. Denn etwa die Forderung nach Nicht-Diskriminierung von LGBTQ+-Menschen würde



eine sehr differenzierte und granulare Aufnahme entsprechender Daten in die medizinischen Datensätze verlangen. Somit geht sie zumindest diskursiv eine Koalition ein mit Experten, die das Mantra vertreten, „je mehr Daten, desto bessere und innovativere KI“, und die Zahlen würden dann „für sich selbst sprechen“. <sup>54</sup> Solche Forderungen befördern den Datenhunger der KI-Medizin und wecken Begehrlichkeiten auf den Zugriff auf elektronische Kranken- und Behandlungsakten. <sup>55</sup> Sofern diese äußerst sensitiven Daten nicht gut geschützt sind und sie gegebenenfalls zentralisiert gespeichert werden, sind sie einerseits anfällig für *Hacking*-Angriffe und *Ransomware*-Attacken. In Finnland beispielsweise wurden 2020 über 30.000 Behandlungsakten eines Psychotherapiezentrums gehackt und die betroffenen Patienten mit der Forderung nach Zahlungen erpresst. Seit Ende Januar 2021 sind die Akten mit persönlichen Angaben, Therapieprotokollen und Diagnosen frei im Internet verfügbar. Arbeitgeber könnten somit diese Daten herunterladen und Bewerber auf mögliche psychische Erkrankungen überprüfen, oder neugierige Bürger die Daten nach Prominenten oder Familienmitgliedern durchsuchen. <sup>56</sup>

Zum anderen kann es zu *Spill-over*-Effekten in andere Bereiche kommen und Daten können etwa für Zwecke der Strafverfolgung eingesetzt werden. In den USA wurden zwei jahrzehntelang gesuchte Serienmörder mittels kriminalpolizeilichen Zugriffs auf die DNA-Profile aus Ahnenforschungs-Datenbanken und privaten *Direct-to-Consumer-Gentest*-Firmen überführt. <sup>57</sup> Relevante Daten können aber auch zu anderen Zwecken missbraucht werden, wie etwa zum Ausschluss aus bestimmten Berufsfeldern (klassisch diesbezüglich war etwa für lange Zeit der Ausschluss offen Homosexueller aus dem Militär). Historisch gesehen zeigen jedenfalls Beispiele, wie etwa kirchliche Daten aus Taufregistern später zur Identifizierung und Verfolgung jüdischer Menschen eingesetzt wurden. Dies zeigt die Herausforderungen, Dilemmata und Paradoxien auf, denen sich stellen muss, wer Auswege aus den aufgezeigten Problemen der Unterrepräsentation bestimmter vulnerabler Gruppen sucht.

## 6. Potentielle soziale Folgen von Bias: Inklusion, Exklusion und Entsolidarisierung

Die eingangs genannten rechtlichen Diskriminierungsverbote sind auch für KI-basierte medizinische Entscheidungssysteme relevant, sofern sie einen Gesundheitsbezug haben und eine sachlich nicht gerechtfertigte Benachteiligung von Individuen hervorrufen. Der Nachweis einer entsprechenden ungerechtfertigten Benachteiligung wird allerdings von den Betroffenen schwer zu führen sein, zumal angesichts des *Black-Box*-Charakters von maschinell lernenden Systemen. Auch sind rein erkenntnissuchende Analysen durch die Diskriminierungsverbote nicht untersagt. <sup>58</sup>

Zu den sozialen Folgen von KI-basierten Entscheidungen können Inklusions- und Exklusionsprozesse zählen. Diese sind nicht zwingend negativ oder diskriminierend, sondern können eine sinnvolle Differenzierung darstellen. So können etwa in der Pharmakogenetik Patienten, bei denen aufgrund von Biomarkern keine Wirkung eines Therapeutikums valide vorhersagbar ist, von dem Zugang dazu ausgeschlossen werden. Bei anderen kann hingegen biomarkergestützt eine besonders gute therapeutische Reaktion erfolgen. Das wohl prominenteste Beispiel für eine solche Genotypisierung in Hinblick auf die Pharmakodynamik ist die antitumorale Therapie des metastasierten HER2-positiven Mammakarzinoms mit Trastuzumab (Herceptin). Auch Dosisanpassungen zur

besseren Wirksamkeit und Verträglichkeit von Medikamenten je nach genetischen Typen metabolischer Aktivität (CYP-Enzym-Polymorphismen) können mithilfe von pharmakogenetischen Tests erkannt und die daraus gewonnenen Informationen über Arzneimittelwirkung und -verstoffwechslung in die Therapie mit eingebaut werden.<sup>59</sup> Hierbei handelt es sich nicht um eine Diskriminierung, sondern um eine sachlich gerechtfertigte Differenzierung. In der so genannten Präzisionsmedizin wird in wachsendem Maße KI in Diagnostik, Therapie und Prädiktion zum Einsatz gelangen, die Patienten in unterschiedlicher Weise stratifiziert, allerdings auch ökonomische Fragen der Finanzierbarkeit aufwirft.

Wie eingangs bereits hingewiesen, setzt der Einsatz von CDSS bei der Prädiktion von Krankheiten bei gesunden Menschen an. Dabei steht die Beurteilung individueller Risikofaktoren wie Blutdruck, EKG, Body-Mass-Index, Lebensstil, Biomarker aus dem *Genomics*-Bereich oder die Vorhersage individueller Reaktionsweisen auf Medikamente (Pharmakogenetik) im Mittelpunkt. Sie kann auch bei der Prämienfestsetzung bei Abschluss einer privaten Versicherung oder der Feststellung der gesundheitlichen beruflichen Eignung eingesetzt werden. Diese prädiktive Medizin will Dispositionen und Empfänglichkeiten für bestimmte Erkrankungen aufdecken.<sup>60</sup> Dies geschieht jedoch auf der Basis oft schwer interpretierbarer statistischer Wahrscheinlichkeiten, die weniger auf Methoden der deduktiven Kausalitätsermittlung als auf induktiven Kausalitätsannahmen aus Korrelationsanalysen beruhen und damit zu bekannten Fehlschlüssen führen können. So können etwa handelsübliche *Smartwatches* mit telemetrischen Funktionen zur Ermittlung atrialer Fibrillationen als Prädiktoren klinisch bedeutsamer Herz-Arhythmien ausgestattet werden.<sup>61</sup> Weltweit breiten Einsatz haben Multiparameter-Genests an Speichelproben an subjektiv Gesunden für die (vorgebliche) Ermittlung individueller Risikoprofile gefunden.<sup>62</sup> Viele solcher Ansätze zeichnen sich dadurch aus, dass die zur Risikoermittlung erhobenen Rohdaten nicht aus einer bestehenden Arzt-Patienten-Beziehung heraus, sondern als *Direct-to-consumer*-Angebote im häuslichen Kontext prospektiver Patienten generiert werden. Dies lässt Defizite u. a. hinsichtlich Aufklärung, Datenschutz und Qualitätssicherung erwarten. Oft besteht auch eine Kluft zwischen den Diagnosemöglichkeiten und präventiven therapeutischen Optionen. Diese prädiktiven Prozesse laufen häufig auf eine Individualisierung von Risiken hinaus. Der medizinische Blick richtet sich insbesondere auf im Körper aufspürbare genetische und *Omic*s-Biomarker und auf den Lebensstil der Patienten. Der im Zusammenhang mit KI-basierten Systemen gebrauchte Begriff der ‚Individualisierung‘ oder ‚Personalisierung‘ von Medizin ist oft irreführend, denn es geht immer um eine Einordnung von Individuen in Kategorien, Klassen und *Scores*, also um eine hierarchische Stratifizierung der Patienten. Dies kann zu Stigmatisierung und (Gruppen-)Diskriminierung führen. Wenn sozialmedizinische Risiken individualisiert werden, kann Krankheit auf eine individuelle Schuldzuschreibung, unter Vernachlässigung von sozialen und umweltbedingten Determinanten, hinauslaufen. Daher ist vor einer Aufkündigung des Solidarprinzips des europäischen Versicherungsmodells<sup>63</sup> und einer weiteren „Benachteiligung der bereits Benachteiligten“<sup>64</sup> zu warnen. Wenn dem nicht entgegengewirkt wird, kann die Individualisierung von Risiken durch KI-basierte Risikoprofil-Bildungen zu einer Entsolidarisierung der Gesellschaft und damit einer Vertiefung sozialer Ungleichheit und sozialer Spaltung beitragen.

## 7. Regulationsansätze

### 7.1 KI- Ethik

Wie mehr Transparenz, Rechenschaftspflicht und demokratische Kontrolle über Algorithmen und KI-basierte klinische Entscheidungsunterstützungssysteme erlangt werden kann, bleibt eine offene Frage, die derzeit unter dem Stichwort ‚KI-Ethik‘ in verschiedensten Anwendungsbereichen diskutiert wird. Inzwischen bestehen eine Fülle von Expertenkommissions-Berichten, Kodizes und Selbstverpflichtungen aus Politik, Wirtschaft und Zivilgesellschaft zur KI-Ethik.<sup>65</sup> Fairness und Schutz vor Diskriminierung werden in vielen Wertekatalogen als Prinzipien und Orientierungsmaßstäbe aufgeführt. Das ‚Wie‘ deren Implementierung bleibt allerdings weitgehend offen. Entsprechend verweisen viele Reports vor allem auf Transparenz, Rechenschaftspflicht und Verfahrensgerechtigkeit als zentrale Verfahrens-Aspekte, um sicherzustellen, dass die Eingaben in ein algorithmisches System akkurat, geeignet und angemessen und die Ergebnisse valide und fair sind. Auch internationale Standardorganisationen wie das *Institute of Electrical and Electronics Engineers* (IEEE) befassen sich intensiv mit KI-Ethik.

### 7.2 Professionelle Selbstregulierung

Innerhalb der Informatik, der Philosophie und der Rechts- und Sozialwissenschaften gibt es durchaus ein gewisses Problembewusstsein zu den epistemischen und ethischen Fragen der KI-Systeme. So besteht eine eigene Fach-Community mit jährlichen Konferenzen zu Fairness, Rechenschaftslegung (*accountability*) und Transparenz beim Maschinellen Lernen (FAT-ML). Diese von der *Association for Computing Machinery* (ACM) gesponserte Konferenz konzentriert sich auf Themen wie algorithmische Transparenz, Fairness beim maschinellen Lernen, Vorurteile und Ethik aus einer multidisziplinären Perspektive.<sup>66</sup> Die Konferenz-Community umfasst Interessierte aus Informatik, Statistik, Sozialwissenschaft, Philosophie und Rechtswissenschaft. Seit 2020 hat sich die Fachgemeinschaft etwas aufgespalten, mit einem Teil (FORC – *Foundation of Responsible Computing*), der vor allem genuin Informatik-technische Fragen bearbeiten möchte, während der andere Teil (ACM FAccT 2021) stärker interdisziplinär mit Wissenschaftlern aus den *Science and Technology-Studies* (STS) und *Data Justice*-Projekten zusammenarbeiten will. Letzterer Zweig wehrt sich gegen technizistische Verkürzungen von Fairness- und Gerechtigkeitsfragen und betont die zwangsläufigen inhärenten normativen Entscheidungen von Software-Entwicklung, denen man nicht mathematisch-technizistisch ausweichen dürfe.<sup>67</sup> Andere artikulieren demokratierelevante Fragen von Selbstregulierung und datenökonomischen Entscheidungen.<sup>68</sup> Seit längerem bestehen daneben Ansätze eines werte-sensitiven Designs von Software und KI-Systemen.<sup>69</sup> Diese sind bisher aber kaum in der Praxis verbreitet.

### 7.3 Aufsicht, Transparenz und Zertifizierung

Neben der Selbstregulierung von Wissenschaft und Medizin werden bestehende und noch zu schaffende Mechanismen zur Kontrolle von KI-Systemen diskutiert. Hierzu zählen Kennzeichnungspflichten von automatisierten Entscheidungsprozessen, und Schulungen für Anwender. Zu mehr Transparenz und Qualitätssicherung soll die Offenlegung von Codes und Trainingsdaten beitragen, zumindest für qualifizierte Stellen, und damit die diskriminierungsrelevante Beforschung von Algorithmen und maschinell lernenden Modellen ermöglichen.<sup>70</sup>

Als Anforderungen zur technischen Gestaltung und Einbettung von KI werden Rechenschaftspflichten und Privatsphäre-Assessments sowie begleitende Technikfolgeabschätzungen wie auch ethische und soziale *ex-ante Impact-Assessments* vorgeschlagen.<sup>71</sup> Ethische Standards und die Berücksichtigung von Rechtmäßigkeitsanforderungen sollen die Risiken von KI einhegen. Als Kontrollinstanzen sind Zulassungskontrollen von KI-Systemen, etwa als Medizinprodukte, bereits implementiert. Ähnlich der Pharmakovigilanz sollten auch Fehlfunktionen oder Unzulänglichkeiten von CDSS, die Leib oder Leben der Patienten erheblich gefährden können, dem Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM) gemeldet werden. Die Ausgestaltung des Haftungsregimes (Verschuldens- oder Gefährdungshaftung) für KI-Systeme auf EU-Ebene dürfte erheblichen Einfluss auf die weitere Entwicklung dieser Systeme nehmen.

Es ist zu fordern, dass vor ihrem breiten Einsatz in der Routineversorgung KI-basierte CDSS hinsichtlich ihrer Nutzen- und Schadens-Potenziale systematisch evaluiert werden sollten, einschließlich ethischer, rechtlicher und sozialer Implikationen. Eine besondere Bedeutung kommt dabei der prospektiven Validierung der Systeme im Rahmen kontrollierter und randomisierter klinischer Studien zu, die bisher für KI-Systeme in der Medizin noch nicht etabliert sind. Dabei gilt es, einer Stigmatisierung und Diskriminierung von Patienten und Bürgern durch eine für alle Bevölkerungsgruppen repräsentative Datenbasis, eine rigorose und robuste Validierung der Modelle, Trainingsdaten und Ergebnisse sowie durch geeignete politische und rechtliche Maßnahmen des Grundrechtsschutzes entgegenzuwirken.

Bei der Entwicklung von Maßnahmen zur Erklärbarkeit und Nachvollziehbarkeit von KI-gestützten CDSS, aber auch bei der Einhegung von Risiken und der Regulierung von CDSS, welche die benannten technischen, ethischen, rechtlichen und sozialen Spannungsfelder als Gestaltungsfrage begreift und konstruktive Lösungsansätze bietet, besteht noch erheblicher Forschungsbedarf. Instruktiv hierzu sind Ansätze, die nicht nur bestehende Diskriminierung verhindern, sondern ‚bias-transformierende‘ Fairnessmetriken beim maschinellen Lernen sowie andere rechtliche und gesellschaftliche Praktiken entwickeln wollen, um historisch und sozial bedingte Ungleichheiten zu überwinden und eine substantielle, nicht nur formale Gleichstellung zu erreichen.<sup>72</sup>

Darüber hinaus werden präventive oder begleitende Aufsichtsmechanismen vorgeschlagen. Leitlinien unter anderem von medizinischen Fachgesellschaften, Standards, Zertifizierungen und betriebliche oder externe KI-System-Beauftragte analog zu den Datenschutzbeauftragten könnten eine Rechenschaftskette zwischen Herstellern, Betreibern und Anwendern etablieren. Dies könnte dazu beitragen, die Entwicklung und Anwendung dieser KI-Systeme auf die Interessen und das Wohl der Patienten und Bürger auszurichten und Schäden zu vermeiden. Auch ein Nachteilsausgleich wäre denkbar.

Eine breite, partizipative gesellschaftliche Auseinandersetzung darüber ist notwendig, denn der Diskurs sollte nicht allein Experten und Unternehmen überantwortet werden. Dabei darf durchaus auch thematisiert werden, dass Technik nicht immer eine Lösung, sondern manchmal ein Problem darstellt, und dass für etliche Probleme *low-tech*-Lösungen und gute Arbeitsplätze hilfreich sind.

#### 7.4 Die KI-Regulierungsinitiative der Europäischen Kommission

In die Frage, welche Entwicklungspfade KI nehmen soll, spielen nicht zuletzt auch internationale Standortkonkurrenzen und geopolitische Fragen – insbesondere nach der Rolle Europas gegenüber den USA und China – hinein.<sup>73</sup> Dies hat die Europäische Kommission (EU-Kommission) in ihrem Weißbuch „Zur Künstlichen Intelligenz – ein europäisches Konzept für Exzellenz und Vertrauen“ aufgenommen, indem sie dort hervorhebt, dass sich der Einsatz der KI in Europa an „unseren Werten und Grundrechten wie Menschenwürde und Schutz der Privatsphäre“ orientieren muss.<sup>74</sup> KI-Systeme sollen mensch-zentriert und vertrauenswürdig sein. Ihre Auswirkungen sollen nicht nur auf das Individuum, sondern auch auf die gesamte Gesellschaft abgeschätzt und Risiken vermieden werden. In diesem Zusammenhang betont die EU-Kommission, dass die menschliche Aufsichts- und Kontrollfunktion gegenüber KI-gestützten Systemen immer gewährleistet sein muss. In der Europäischen Union (EU) wurde durch die Hochrangige Expertengruppe der EU-Kommission (HLEG) ein Konsens über sieben Kernanforderungen hergestellt: Die ethischen Kriterien des Vorrangs menschlichen Handelns, der technischen Robustheit und Sicherheit, der Wahrung von Datensouveränität und Transparenz, der Berücksichtigung von Vielfalt, Nichtdiskriminierung und Fairness, des gesellschaftlichen Wohlergehens und der Rechenschaftspflicht sind zu beachten und umzusetzen.<sup>75</sup> Wie die EU-Kommission betont, müssen diese Kernanforderungen entsprechend auf den verschiedenen Verantwortungsebenen eingelöst werden, um personales und Systemvertrauen zu gewährleisten. Diese Kernanforderungen sind auch in den von der EU-Kommission am 21. April 2021 veröffentlichten Verordnungsentwurf COM(2021) 206 zum Umgang mit Künstlicher Intelligenz eingegangen.<sup>76</sup>

Dieser Vorschlag der EU-Kommission zu KI-Regulierung ist eine eigene Analyse wert, die an dieser Stelle nicht geleistet werden kann. Einige Schlaglichter sollen genügen. Der Verordnungsentwurf will KI im Einklang mit der europäischen Grundrechte-Charta, insbesondere dem Recht auf Menschenwürde (Art. 1), der Achtung des Privatlebens und des Schutzes personenbezogener Daten (Art. 7 und 8), der Nichtdiskriminierung (Art. 21) und der Gleichstellung von Frauen und Männern (Art. 23) regulieren. Für Risiken wird eine vierstufige Risiko-Skala vorgeschlagen, wovon die in Stufe Vier als „Hoch-Risiko“ eingeordneten KI-Systeme besonders beaufsichtigt und reguliert werden sollen. Gesundheit und Sicherheit sind elementare Maßstäbe. Dabei wird auch das Ziel einer „Minimierung des Risikos von fehlerhaften oder voreingenommenen KI-gestützten Entscheidungen in kritischen Bereichen“ formuliert. Unter dieser Hochrisiko-Gruppe werden explizit „KI-Systeme, die zur Disposition oder zur Festlegung von Prioritäten bei der Disposition von Notfalleinsatzdiensten eingesetzt werden“ eingestuft, „da sie Entscheidungen in sehr kritischen Situationen für das Leben und die Ge-

sundheit von Personen und deren Eigentum treffen“ (Erwägungsgrund 37). Diskriminierungsrisiken im klinischen Medizinbereich und der Forschung selbst, wie im vorliegenden Aufsatz diskutiert, werden nicht explizit genannt. Hingegen geht es um Nicht-Diskriminierung im wirtschaftlichen und wettbewerbsrechtlichen Sinne, wenn statuiert wird, „der europäische Gesundheitsdatenraum [wird] den diskriminierungsfreien Zugang zu Gesundheitsdaten und das Training von Algorithmen der künstlichen Intelligenz auf diesen Datensätzen erleichtern, und zwar auf eine die Privatsphäre wahrende, sichere, zeitnahe, transparente und vertrauenswürdige Weise und mit einer angemessenen institutionellen Steuerung. Einschlägige zuständige Behörden, einschließlich sektoraler Behörden, die den Zugang zu Daten bereitstellen oder unterstützen, können auch die Bereitstellung hochwertiger Daten für das Training, die Validierung und das Testen von KI-Systemen unterstützen“ (Erwägungsgrund 45). Hiervon werden jedoch wiederum Abstriche gemacht, wenn es heißt: „Unter bestimmten Bedingungen kann die rasche Verfügbarkeit innovativer Technologien für die Gesundheit und Sicherheit von Personen und für die Gesellschaft als Ganzes entscheidend sein. Daher sollten die Mitgliedstaaten ausnahmsweise aus Gründen der öffentlichen Sicherheit oder des Schutzes von Leben und Gesundheit natürlicher Personen sowie des Schutzes des gewerblichen und kommerziellen Eigentums das Inverkehrbringen oder die Inbetriebnahme von KI-Systemen, die keiner Konformitätsbewertung unterzogen wurden, zulassen können“ (Erwägungsgrund 47).

In welcher Form und mit welchen Änderungen diese Verordnung COM(2021)206 in der EU letztlich verabschiedet wird, bleibt abzuwarten.

## 8. Fazit

Wie die bisherige Diskussion gezeigt hat, bietet der Einsatz von KI in der Medizin Chancen, ist aber auch mit vielen Herausforderungen verbunden, die etwa die Transparenz und Repräsentativität der Datengenerierung, die klare Zuschreibung von Verantwortlichkeiten, auch in Bezug auf Haftungsfragen, den Schutz von Autonomie, Vertrauen und Privatheit, die Qualität der Kommunikation zwischen Ärzten und Patienten sowie institutionelle Strukturen betreffen.<sup>77</sup> Wie gezeigt wurde, können KI-basierte klinische Entscheidungsunterstützungssysteme sinnvolle Differenzierungen liefern und damit potentiell Krankheiten besser erkennen und behandeln helfen. Die Validität und Nachvollziehbarkeit algorithmisch basierter Entscheidungen bedarf jedoch der Verbesserung, damit sie nicht einer schleichenden Ökonomisierung der Krankenversorgung anheimfällt. Allerdings werfen die KI-Anwendungen, insbesondere bei der Prognose und Prädiktion, ein erhebliches Diskriminierungspotential auf. Sofern dem nicht aktiv entgegengewirkt wird, können sich bestehende Ungleichheiten und Asymmetrien durch KI nicht nur fortsetzen, sondern sogar verschärfen. Dies stellt eine Gestaltungsaufgabe dar, die auf die Überwindung vielfältiger Formen sozialer Ungleichheit gerichtet werden sollte. Vor diesem Hintergrund stellt eine frühzeitige Abwägung von Chancen und Risiken der KI in der Medizin eine wichtige sozialetische und politische Aufgabe im Hinblick auf die Gesundheitsversorgung der Zukunft dar. Letztlich geht es um die Frage, in welcher Gesellschaft wir leben wollen.

## ANMERKUNGEN

- <sup>1</sup> Vgl. C. O'NEIL, *Angriff der Algorithmen*, München 2017.
- <sup>2</sup> T. HAGENDORFF, *Maschinelles Lernen und Diskriminierung: Probleme und Lösungsansätze*, in: Österreichische Zeitschrift für Soziologie 44 (2019), S1, 53–66.
- <sup>3</sup> C. ORWAT, *Diskriminierungsrisiken durch Verwendung von Algorithmen*, Baden-Baden 2019, 24.
- <sup>4</sup> Vgl. L. TAYLOR/L. FLORIDI/B. VAN DER SLOOT (Hg.), *Group Privacy. New Challenges of Data Technologies*, Dordrecht 2017.
- <sup>5</sup> Vgl. S. BAROCAS/A. SELBST, *Big Data's Disparate Impact*, in: California Law Review 104 (2016), 671–732.
- <sup>6</sup> Vgl. E. J. TOPOL, *High-performance medicine: the convergence of human and artificial intelligence*, in: Nature Medicine 25 (2019), 44–56.
- <sup>7</sup> Vgl. B. FRIEDMAN/H. NISSENBAUM, *Bias in Computer Systems*, in: B. Friedman (Hg.), *Human Values and the Design of Computer Technology*, Cambridge 1997, 21–40.
- <sup>8</sup> Vgl. M. CHRISTEN/C. MADER/J. ČAS u. a., *Wenn Algorithmen für uns entscheiden. Chancen und Risiken der künstlichen Intelligenz*, Zürich 2020, 67–114.
- <sup>9</sup> K. ZWEIG, *Wo Maschinen irren können. Arbeitspapier*, Bertelsmann Stiftung 2018.
- <sup>10</sup> Vgl. F. PASQUALE, *The Black Box Society*, Cambridge 2015.
- <sup>11</sup> X. LIU/L. FAES/A. U. KALE u. a., *A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis*, in: The Lancet Digital Health 1 (2019), 6, e271–e297.
- <sup>12</sup> M. KIESER, *Big Data in der klinischen Forschung: Vieles ist noch Wunschdenken*, in: Deutsches Ärzteblatt 116 (2019), 35/36, A-1534.
- <sup>13</sup> Grundsätzlich ist festzuhalten, dass es bei allen KI-Systemen inhärent zu falsch-negativen und falsch-positiven Ergebnissen kommt, die u. a. Ergebnis von Klassifikationsfehlern sind. Sicherlich machen auch Menschen Fehler und können andere diskriminieren. Gleichwohl sind Fehler bei KI-Systemen insofern nicht vergleichbar, als sie erheblich mehr skalieren können. Zum anderen gibt es bei CDSS zunächst keine regulierenden Mechanismen, in welche diese eingebettet sind, wie etwa Ärzteteams, Ethik-Codizes und soziale Kontrollmechanismen (vgl. T. PLOUG/S. HOLM, *The right to refuse diagnostics and treatment planning by artificial intelligence*, in: Medicine, Health Care and Philosophy 23 (2020), 107–114, 110). Qualitäts- und Sicherheitsprobleme sowie Fehler können beispielsweise entstehen, wenn sich die Anwendungsbedingungen sehr von den Trainingsdaten der CDSS unterscheiden. Sofern bestimmte Gruppen von Patienten in den Trainingsdaten unterrepräsentiert sind, können die Spezifität und Sensitivität der CDSS hinsichtlich dieser Patientengruppe erheblich niedriger sein (vgl. R. CHALLEN/J. DENNY/M. PITT u. a., *Artificial intelligence, bias and clinical safety*, in: BMJ Quality Safety 28 (2019), 3, 231–237). Vergleichsstudien zeigen zudem, dass die Ergebnisse menschlichen Urteilens und Einschätzens keineswegs dem Outcome digitaler Maschinen unterlegen sein müssen. Grundsätzlich ist zu berücksichtigen, dass Menschen andere Fehler machen als Maschinen. Maschinenlernenden Systemen fehlt das Kontextwissen von Ärzten, wodurch es zu erheblichen Fehlschlüssen kommen kann (vgl. F. CABITZA/R. RASOINI/G. F. GENSINI, *Unintended consequences of machine learning in medicine*, in: JAMA 318 (2017), 6, 517–518).
- <sup>14</sup> PLOUG/HOLM (Anm. 13), 149.
- <sup>15</sup> Vgl. K. SO, *A primer on AI fairness*, 2019, online unter: [www.towardsdatascience.com/artificial-intelligence-fairness-and-tradeoffs-ce11ac284b63U](http://www.towardsdatascience.com/artificial-intelligence-fairness-and-tradeoffs-ce11ac284b63U) (Zugriff am 28.6.2021).
- <sup>16</sup> Vgl. A. KAUTKZY-WILLER, *Gendermedizin*, Böhlau 2012.
- <sup>17</sup> A. BLASIMME/E. VAYENA, *The Ethics of AI in Biomedical Research, Patient Care and Public Health*, in: M. Dubber/F. Pasquale/S. Das (Hg.), *The Oxford Handbook of Ethics of Artificial Intelligence*, Oxford/New York 2020.
- <sup>18</sup> Vgl. T. GEBRU, *Race and Gender*, in: M. Dubber/F. Pasquale/S. Das (Hg.), *The Oxford Handbook of Ethics of AI*, 2020, 251–269; K. CRAWFORD, *The Atlas of AI*, Yale 2021.
- <sup>19</sup> ORWAT (Anm. 3), 86 ff.; T. WEICHERT, *Big Data im Gesundheitsbereich*, ABIDA – Assessing Big Data, Gutachten, 2018, 104, online unter: [www.abida.de/sites/default/files/ABIDA%20Gutachten-Gesundheitsbereich.pdf](http://www.abida.de/sites/default/files/ABIDA%20Gutachten-Gesundheitsbereich.pdf) (Zugriff am 28.6.2021).

- <sup>20</sup> Vgl. W. GLAUSER, *AI in health care: Improving outcomes or threatening equity?*, in: CMAJ 192 (2020), 1, E21–E22; A. S. ADAMSON/A. SMITH, *Machine Learning and Health Care Disparities in Dermatology*, in: JAMA Dermatology 154 (2018), 1247.
- <sup>21</sup> L. GOODE, *Facial recognition software is biased towards white men, researcher finds*, online unter: [www.theverge.com/2018/2/11/17001218/facial-recognition-software-accuracy-technology-mit-white-men-black-women-error](http://www.theverge.com/2018/2/11/17001218/facial-recognition-software-accuracy-technology-mit-white-men-black-women-error) (Zugriff am 28.6.2021).
- <sup>22</sup> Vgl. A. RIEDER/B. LOHFF (Hg.), *Gender Medizin*, Wien 2008.
- <sup>23</sup> Z. OBERMEYER u. a., *Dissecting racial bias in an algorithm used to manage the health of populations*, in: Science 366 (2019), 6464, 447–453.
- <sup>24</sup> D. A. VYAS/L. G. EISENSTEIN/D. S. JONES, *Hidden in Plain Sight – Reconsidering the Use of Race Correction in Clinical Algorithms*, in: New England Journal of Medicine 83 (2020), 9, 874–882.
- <sup>25</sup> VYAS/EISENSTEIN/JONES (Anm. 24), 880.
- <sup>26</sup> Vgl. R. BAUMGARTNER, *Künstliche Intelligenz in der Medizin: Diskriminierung oder Fairness?*, in: G. Bauer/M. Kechaja/S. Engelmann/L. Haug (Hg.), *Diskriminierung und Antidiskriminierung*, Bielefeld 2021, 149–164.
- <sup>27</sup> M. A. GIANFRANCESCO/S. TAMANG/J. YAZDANY/G. SCHMAJUK, *Potential Biases in Machine Learning Algorithms Using Electronic Health Record Data*, in: JAMA Internal Medicine 178 (2018), 11, 1544–1547.
- <sup>28</sup> A. RAJKOMAR/M. HARDT/M. D. HOWELL u. a., *Ensuring Fairness in Machine Learning to Advance Health Equity*, in: Annals of Internal Medicine 169 (2018), 12, 866–872.
- <sup>29</sup> BUNDESMINISTERIUM FÜR ARBEIT UND SOZIALES, *Sechster Armuts- und Reichtumsbericht (Entwurf)*, 2021, 378 ff., online unter: [www.armuts-und-reichtumsbericht.de](http://www.armuts-und-reichtumsbericht.de) (Zugriff am 28.6.2021).
- <sup>30</sup> T. LAMPERT/J. HOEBEL/B. KUNTZ u. a., *Gesundheitliche Ungleichheit in verschiedenen Lebensphasen*, Gesundheitsberichterstattung des Bundes, gemeinsam getragen von RKI und Destatis, Robert Koch-Institut, Berlin 2017; DIES., *Soziale Unterschiede in der Mortalität und Lebenserwartung in Deutschland – Aktuelle Situation und Trends*, in: Journal of Health Monitoring 4 (2019), 1, 3–15.
- <sup>31</sup> SVR (SACHVERSTÄNDIGENRAT ZUR BEGUTACHTUNG DER ENTWICKLUNG IM GESUNDHEITSWESEN), *Gutachten 2018: Bedarfsgerechte Steuerung der Gesundheitsversorgung*, online unter: [www.svr-gesundheit.de/fileadmin/Gutachten/Gutachten\\_2018/Kurzfassung\\_2018.pdf](http://www.svr-gesundheit.de/fileadmin/Gutachten/Gutachten_2018/Kurzfassung_2018.pdf) (Zugriff am 28.6.2021).
- <sup>32</sup> SVR, *Gutachten 2021: Digitalisierung für Gesundheit*, 228, online unter: <https://www.svr-gesundheit.de/> (Zugriff am 28.6.2021).
- <sup>33</sup> Vgl. C. ASCHWANDEN, *Artificial Intelligence Makes Bad Medicine Even Worse*, 1.10.2020, online unter: [www.wired.com/story/artificial-intelligence-makes-bad-medicine-even-worse/](http://www.wired.com/story/artificial-intelligence-makes-bad-medicine-even-worse/) (Zugriff am 28.6.2021).
- <sup>34</sup> R. CARUANA u. a., *Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission*, in: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1721–1730, 1721.
- <sup>35</sup> Vgl. E. ANTHES, *KI erkennt Krankheit: Alexa, habe ich Covid-19?* 2020, online unter: [www.spektrum.de/news/kuenstliche-intelligenz-unterscheidet-stimme-von-gesunden-und-kranken/1777593](http://www.spektrum.de/news/kuenstliche-intelligenz-unterscheidet-stimme-von-gesunden-und-kranken/1777593) (Zugriff am 28.6.2021).
- <sup>36</sup> CH. VILLONGCO/F. KHAN, *„Sorry I Didn’t Hear You.“ The Ethics of Voice Computing and AI in High Risk Mental Health Populations*, in: AJOB Neuroscience 11 (2020), 2, 105–112.
- <sup>37</sup> G. STARKE/E. DE CLERCQ/B.S. ELGER, *Towards a pragmatist dealing with algorithmic bias in medical machine learning*, in: Medical Health Care and Philosophy (2021), 1–9; S. HOFFMAN/A. PODGURSKI, *Artificial Intelligence and Discrimination in Health Care*, in: Yale Journal of Health Policy, Law, and Ethics 19 (2020), 3, 17.
- <sup>38</sup> F. B. POKORNY u. a., *Earlier Identification of Children with Autism Spectrum Disorder: An Automatic Vocalisation-Based Approach*, in: Proceedings Interspeech (2017), 309–313.
- <sup>39</sup> KI-Systeme in der Pädiatrie sind ein besonders schwieriges Thema, da sie eine sehr vulnerable Gruppe betreffen. Da Kinder sich in Wachstum und Entwicklung befinden und diese bei Gleichaltrigen oft unterschiedlich verläuft, müssten bei Entscheidungen im Vergleich zu Erwachsenen zusätzliche Daten berücksichtigt und entsprechende Datenbanken und Algorithmen erweitert werden. Dies wäre allerdings wiederum mit ethischen und rechtlichen (stellvertretende Einwil-



ligung bei Minderjährigen, gegebenenfalls konstante Überwachung der Kinder) sowie technischen Herausforderungen (Schwierigkeiten der Standardisierung) verbunden.

- <sup>40</sup> M. IENCA/K. IGNATIADIS, *Artificial Intelligence in Clinical Neuroscience. Methodological and Ethical Challenges*, in: *AJOB Neuroscience* 11 (2020), 2, 77–87, 83.
- <sup>41</sup> C. SONG/V. SHMATIKOV, *Auditing Data Provenance in Text-Generation Models*, 2019, online unter: [www.arxiv.org/pdf/1811.00513.pdf](http://www.arxiv.org/pdf/1811.00513.pdf) (Zugriff am 28.6.2021).
- <sup>42</sup> Vgl. ANTHES (Anm. 35).
- <sup>43</sup> Der Begriff ‚Triage‘ sollte meines Erachtens Katastrophenfällen und absoluten Ausnahmesituationen wie dem Mangel an Beatmungsbetten auf Intensivstationen in einer Pandemie vorbehalten bleiben. Die Popularisierung, und sei es auch nur zur Skandalisierung zwecks Schaffens öffentlicher Aufmerksamkeit, wie im oben zitierten Fall, trägt m.E. diskursiv zu einer Normalisierung von etwas bei, an das sich eine Gesellschaft normativ gesehen niemals gewöhnen darf. Der Begriff wurde in diesem Aufsatz beibehalten, weil die genannte Autorengruppe ihn so verwendet hat.
- <sup>44</sup> TH. BESCHORNER u. a., *Angesichts von Triage und Todesalgorithmen: Ist die heutige datengetriebene Medizin mit der Verfassung vereinbar?*, Policy-Paper, 2021, online unter: [www.researchgate.net/publication/350870973\\_Angesichts\\_von\\_Triage\\_und\\_Todesalgorithmen\\_Ist\\_die\\_heutige\\_daten-getriebene\\_Medizin\\_mit\\_der\\_Verfassung\\_vereinbar/link/6076fcf3a020ab5f2175a375/download](http://www.researchgate.net/publication/350870973_Angesichts_von_Triage_und_Todesalgorithmen_Ist_die_heutige_daten-getriebene_Medizin_mit_der_Verfassung_vereinbar/link/6076fcf3a020ab5f2175a375/download) (Zugriff am 28.6.2021).
- <sup>45</sup> ÄRZTETAG 2021, *Reform der Notfallversorgung: Schlüssiges Gesamtkonzept statt Stückwerk*, Beschluss des 124. Deutschen Ärztetags (Online) 1 – 04, TOP I Gesundheits-, Sozial- und ärztliche Berufspolitik: Leitantrag zu Lehren aus der COVID-19-Pandemie.
- <sup>46</sup> BESCHORNER u. a. (Anm. 44).
- <sup>47</sup> Vgl. BÄK, *Stellungnahme der Bundesärztekammer als sachkundige Dritte im Verfahren bezüglich der Verfassungsbeschwerde gegen die staatlichen Maßnahmen zur Bewältigung der durch das COVID-19 ausgelösten Pandemie und die Untätigkeit der Bundesregierung, Vorkehrungen zu treffen, die Beschwerdeführenden vor Benachteiligung wegen ihrer Behinderung und in Zusammenhang mit ihrem Alter im Rahmen der gesundheitlichen Versorgung zu schützen* (1 BvR 1541/20), Berlin, 10.12.2020.
- <sup>48</sup> Vgl. BÄK, *Präzisionsmedizin: Bewertung unter medizinisch-wissenschaftlichen und ökonomischen Aspekten*, Stellungnahme, in: *Deutsches Ärzteblatt* 117 (2020), 22/23, A-1184.
- <sup>49</sup> A. LOBE, *Der Algorithmus schlägt die letzte Stunde*, in: F.A.Z. vom 8.1.2017, online unter: [www.faz.net/aktuell/feuilleton/debatten/ein-algorithmus-soll-ueber-leben-und-tod-entscheiden-14606357.html](http://www.faz.net/aktuell/feuilleton/debatten/ein-algorithmus-soll-ueber-leben-und-tod-entscheiden-14606357.html) (Zugriff am 28.6.2021).
- <sup>50</sup> C. ORWAT/A. KOLLECK, *Mögliche Diskriminierung durch algorithmische Entscheidungssysteme und maschinelles Lernen*, TAB-Hintergrundpapier Nr. 24, 2020, 46f.
- <sup>51</sup> B. BRISEÑO, *Optimierung der Palliativversorgung: Wenn Algorithmen den Tod vorhersagen*, in: *Algorithmenethik*, 21.3.2018, online unter: [www.algorithmenethik.de/2018/03/21/optimierung-der-palliativversorgung-wenn-algorithmen-den-tod-vorhersagen/](http://www.algorithmenethik.de/2018/03/21/optimierung-der-palliativversorgung-wenn-algorithmen-den-tod-vorhersagen/) (Zugriff am 28.6.2021).
- <sup>52</sup> J. MORLEY/L. FLORIDI, *An Ethically Mindful Approach to AI for Health Care*, 2020, online unter: [www.ssrn.com/abstract=3830536](http://www.ssrn.com/abstract=3830536) (Zugriff am 28.6.2021).
- <sup>53</sup> BAUMGARTNER (Anm. 26), 157, 154.
- <sup>54</sup> C. ANDERSON, *The End of Theory: The Data Deluge Makes the Scientific Method Obsolete*, in: *Wired*, 2008 online unter: [www.wired.com/2008/06/pb-theory/](http://www.wired.com/2008/06/pb-theory/) (Zugriff am 28.6.2021) dagegen kritisch: D. BOYD/C. CRAWFORD, *Critical questions for Big Data*, in: *Information, Communication & Society* 15 (2012), 5, 662–679; S. EPSTEIN, *Inclusion: The Politics of Difference in Medical Research*, Chicago 2007.
- <sup>55</sup> Siehe SVR 2021 (Anm. 32).
- <sup>56</sup> J. BALLWEBER, *Daten in der Forschung. Zum Wohle der Patient:innen*, 14.5.2021, online unter: [www.netzpolitik.org/2021/daten-in-der-forschung-zum-wohle-der-patientinnen/](http://www.netzpolitik.org/2021/daten-in-der-forschung-zum-wohle-der-patientinnen/) (Zugriff am 28.6.2021).
- <sup>57</sup> E. ROSENBAUM, *5 biggest risks of sharing your DNA with consumer genetic-testing companies*, in: *CNBC* am 16.6.2018, online unter: [www.cNBC.com/2018/06/16/5-biggest-risks-of-sharing-dna-with-consumer-genetic-testing-companies.html](http://www.cNBC.com/2018/06/16/5-biggest-risks-of-sharing-dna-with-consumer-genetic-testing-companies.html) (Zugriff am 28.6.2021).
- <sup>58</sup> WEICHERT (Anm. 19), 103.

- <sup>59</sup> Vgl. R. KOLLEK/G. FEUERSTEIN/M. SCHMEDDERS/J. VAN AKEN, *Pharmakogenetik: Implikationen für Patienten und Gesundheitswesen*, Baden-Baden 2004; L. KOVAR /S. SCHMIDT/H. DERENDORF, *Maßgeschneidert*, 2018, online unter: [www.deutsche-apotheker-zeitung.de/daz-az/2018/daz-48-2018/massgeschneidert](http://www.deutsche-apotheker-zeitung.de/daz-az/2018/daz-48-2018/massgeschneidert) (Zugriff am 28.6.2021).
- <sup>60</sup> Vgl. M. WILDNER, *Diagnose, Prognose, Prädiktion*, in: *Gesundheitswesen* 76 (2014), 4, 185–186.
- <sup>61</sup> T. INUI u. a., *Use of a Smart Watch for Early Detection of Paroxysmal Atrial Fibrillation: Validation Study*, in: *JMIR Cardiology* 4 (2020), 1, e14857.
- <sup>62</sup> M. BATES, *Direct-To-Consumer Genetic Testing: Is the Public Ready for Simple, At-Home DNA Tests to Detect Disease Risk?*, in: *IEEE Pulse* 9 (2018), 6, 11–14.
- <sup>63</sup> WEICHERT (Anm. 19), 106.
- <sup>64</sup> BAROCAS/SELBST (Anm. 5), 675.
- <sup>65</sup> Siehe zum Überblick C. RUDSCHIES/I. SCHNEIDER/J. SIMON, *Value Pluralism in the AI Ethics Debate – Different Actors, Different Priorities*, in: *The International Review of Information Ethics* 29 (2021), online unter: [www.informationethics.ca/index.php/irrie/article/view/419](http://www.informationethics.ca/index.php/irrie/article/view/419) (Zugriff am 26.6.2021); DIES., *Heterogeneity of AI Ethics Guidelines Examined: Varying Natures, Actors, and Perceptions*, in: Hoeren/Pinelli (Hg.), *Künstliche Intelligenz – Ethik und Recht* (im Erscheinen).
- <sup>66</sup> vgl. B. LEPRI/E. OLIVER/A. LETOUZÉ u. a., *Fair, Transparent, and Accountable Algorithmic Decision-making Processes: The Premise, the Proposed Solutions, and the Open Challenges*, in: *Philosophy & Technology* 31 (2017), 611–627.
- <sup>67</sup> Vgl. L. TAYLOR, *Global Data Justice*, in: *Communications of the ACM* 62 (2019), 6, 22–24; L. TAYLOR/N. PURTOVA, *What is Responsible and Sustainable Data Science?*, in: *Big Data & Society* 2 (2019), 1–6; O. KEYES/J. HUTSON/M. DURBIN, *A Mulching Proposal*, in: S. Brewster/G. Fitzpatrick/A. Cox/V. Kostakos (Hg.), *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, New York 2019, 1–11.
- <sup>68</sup> Vgl. I. SCHNEIDER, *Bringing the state back in: Big Data-based capitalism, disruption, and novel regulatory approaches in Europe*, in: I. Schneider/A. Rudinow Saetnan/N. Green (Hg.), *The Politics of Big Data: Big Data, Big Brother?*, New York 2018, 129–175.
- <sup>69</sup> Vgl. B. FRIEDMAN/P. H. KAHN, *Value Sensitive Design and Information Systems*, in: P. Zhang/D. Galletta (Hg.), *Human-Computer Interaction in Management Information Systems: Foundations*, New York 2006, 348–372.
- <sup>70</sup> Zu den Potentialen und Grenzen der Transparenz vgl. M. ANANNY/K. CRAWFORD, *Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability*, 2018, online unter: [journals.sagepub.com/doi/pdf/10.1177/1461444816676645](http://journals.sagepub.com/doi/pdf/10.1177/1461444816676645) (Zugriff am 28.6.2021).
- <sup>71</sup> C. D. RAAB, *Information privacy, impact assessment, and the place of ethics*, in: *Computer Law & Security Review* 37 (2020), 105404.
- <sup>72</sup> S. WACHTER/B. MITTELSTADT/C. RUSSELL, *Bias preservation in machine learning: the legality of fairness metrics under EU non-discrimination law*, in: *West Virginia Law Review* (2021) (im Erscheinen).
- <sup>73</sup> I. SCHNEIDER, *Democratic Governance of Digital Platforms and Artificial Intelligence?*, in: *JeDEM – EJournal of EDemocracy and Open Government* 12 (2020), 1, 1–24.
- <sup>74</sup> EUROPÄISCHE KOMMISSION, *Zur Künstlichen Intelligenz – ein europäisches Konzept für Exzellenz und Vertrauen*, Weißbuch, COM(2020) 65 final vom 19.2.2020, 2.
- <sup>75</sup> HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE (HLEG), *Ethics guidelines for trustworthy AI*, Brüssel 2019.
- <sup>76</sup> EUROPEAN COMMISSION, *Proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*, COM(2021) 206 final vom 21.04.2021.
- <sup>77</sup> Vgl. ZEKO, *Entscheidungsunterstützung ärztlicher Tätigkeit durch Künstliche Intelligenz*, Stellungnahme der Zentralen Kommission zur Wahrung ethischer Grundsätze in der Medizin und ihren Grenzgebieten bei der Bundesärztekammer (wird im Herbst 2021 veröffentlicht).

# Autorenverzeichnis

DR. MED. ERIK ALLEMMEYER, geb. 1968, Facharzt für Chirurgie, spezielle Viszeralchirurgie, Proktologie und Notfallmedizin, leitender Arzt; Proktologie, Kontinenz- und Beckenbodenchirurgie, Niels-Stensen-Kliniken, Franziskus-Hospital Harderberg, Alte Rothenfelder Straße 23, D – 49124 Georgsmarienhütte.

erik.allemeyer@niels-stensen-kliniken.de

KFEEL ARSHAD, M. SC., Wissenschaftlicher Mitarbeiter, Christian-Albrechts-Universität zu Kiel, Institut für Medizinische Informatik und Statistik, Sektion Medizinische Informatik, Universitätsklinikum Schleswig-Holstein, Kaistr. 101, D – 24114 Kiel.

kfeel.arshad@uksh.de

DR. MED. BJÖRN BERGH, Direktor des Instituts für Medizinische Informatik und Statistik mit Universitätsprofessur für Medizinische Informatik, Leiter der Sektion Medizinische Informatik; Christian-Albrechts-Universität zu Kiel, Universitätsklinikum Schleswig-Holstein, Kaistr. 101, D – 24114 Kiel.

bjoern.bergh@uksh.de

THOMAS BUCHSCHUSTER, MAG. THEOL., geb. 1994, Wissenschaftlicher Mitarbeiter am Lehrstuhl für Moralthologie; Eberhard Karls Universität Tübingen, Katholisch-Theologische Fakultät, Liebermeisterstr. 12, D – 72076 Tübingen.

thomas.buchschuster@uni-tuebingen.de

DR. MED. FLORIAN FUNER, MAG. THEOL., M. A., geb. 1993, Wissenschaftlicher Mitarbeiter am Lehrstuhl für Moralthologie, Redaktion der Zeitschrift für medizinische Ethik; Eberhard Karls Universität Tübingen, Katholisch-Theologische Fakultät, Liebermeisterstr. 12, D – 72076 Tübingen.

florian.funer@uni-tuebingen.de

DR. MED. MARKUS HOHENFELLNER, geb. 1958, Ärztlicher Direktor; Urologische Universitätsklinik, Im Neuenheimer Feld 420, D – 69120 Heidelberg.

DOMINIK KEMMER, M. A., M. Sc., geb. 1989, Wissenschaftlicher Mitarbeiter, Organisationsberater, Coach; Friedrich-Alexander-Universität Erlangen Nürnberg, Lehrstuhl für Psychologie im Arbeitsleben, Nägelsbachstr. 49c, D – 91052 Erlangen.

dominik.kemmer@distinkt.de

DR. MED. CARSTEN JOHANNES KRONES, geb. 1966, Professor, Facharzt für Chirurgie, Viszeralchirurg, Chefarzt, Mitglied des erweiterten Vorstands des Berufsverbands der Deutschen Chirurgen (BDC); Marienhospital Aachen, Klinik für Allgemein-, Viszeral- und Minimalinvasive Chirurgie, Zeise 4, D – 52066 Aachen.

carsten.krones@marienhospital.de

DR. THEOL. MARTIN LANGANKE, geb. 1972, Professor für Ethik; Evangelische Hochschule Rheinland-Westfalen-Lippe, Protestant University of Applied Sciences, Fachbereich Soziale Arbeit, Immanuel-Kant-Str. 18–20, D – 44803 Bochum.

langanke@evh-bochum.de

DR. THEOL. WENKE LIEDTKE, geb. 1981, Wissenschaftliche Mitarbeiterin; Evangelische Hochschule Rheinland-Westfalen-Lippe, Protestant University of Applied Sciences, Fachbereich Soziale Arbeit, Immanuel-Kant-Str. 18–20, D – 44803 Bochum.

liedtke@evh-bochum.de

DR. PHIL. SASKIA K. NAGEL, Universitätsprofessorin; RWTH Aachen, Angewandte Ethik, Theaterplatz 14, D – 52062 Aachen.

saskia.nagel@humtec.rwth-aachen.de

JULIA PETERSEN, M. SC., geb. 1990, Wissenschaftliche Mitarbeiterin; Bundesanstalt für Arbeitsschutz und Arbeitsmedizin, Fabricestr. 8, D – 01099 Dresden.

petersen.julia@baua.bund.de

SEBASTIAN RITZI, Akademischer Mitarbeiter; Universität Heidelberg, Institut für Gerontologie, Berghheimerstraße 20, D – 69115 Heidelberg.

sebastian.ritzl@gero.uni-heidelberg.de

DR. MED., DIPL.-PHYS. HEINZ-PETER SCHLEMMER, geb. 1961, Universitätsprofessor und Abteilungsleiter der Radiologie am DKFZ; Im Neuenheimer Feld 280, D – 69120 Heidelberg.

h.schlemmer@dkfz.de

DR. PHIL. INGRID SCHNEIDER, Professorin für Politikwissenschaft, Fachbereich Informatik, Ethik in der Informationstechnologie; Universität Hamburg, Vogt-Kölln-Straße 30, D – 22527 Hamburg.

ingrid.schneider@uni-hamburg.de

DR. SC. HUM. BJÖRN SCHREIWEIS, Professor für Medizininformatik am Institut für Medizinische Informatik und Statistik, Sektion Medizinische Informatik; Christian-Albrechts-Universität zu Kiel, Universitätsklinikum Schleswig-Holstein, Kaistr. 101, D – 24114 Kiel.

bjoern.schreiweis@uksh.de

DIPL.-KFM. ARNE SIEWERT, Ex. MBA, geb. 1982, Manager, Forensic Services, Pricewaterhouse Coopers GmbH, Wirtschaftsprüfungsgesellschaft, Friedrich-Eber-Anlage 37-39, D – 60327 Frankfurt a. M. arnesiewert@gmx.de

DR. MED. CLAAS STRODTHOFF, Assistenzarzt für Anästhesiologie; Universitätsklinikum Schleswig-Holstein, Campus Kiel, Klinik für Anästhesiologie und operative Intensivmedizin, Arnold-Heller-Str. 3, D – 24105 Kiel.

Claas.strodthoff@uksh.de

DR. PHIL. WERNER VOGD, geb. 1963, Universitätsprofessor und Inhaber des Lehrstuhls für Soziologie; Universität Witten/Herdecke, Fakultät für Gesundheit (Department für Humanmedizin), Lehrstuhl für Soziologie, Alfred-Herrhausen-Straße 50, D – 58448 Witten.

werner.vogd@uni-wh.de

DR. MED. KARL-HEINZ WEHKAMP, Professor, affiliertes Mitglied im SOCIUM Forschungszentrum; SOCIUM Forschungszentrum Ungleichheit und Sozialpolitik, Universität Bremen, Unicom-Gebäude, Mary-Somerville-Straße 3, D – 28359 Bremen.

karl.wehkamp@uni-bremen.de

DR. JUR. THILO WEICHERT, geb. 1955, Datenschutzexperte; Waisenhofstr. 41, D – 24103 Kiel.

thilo.weichert@t-online.de

# Zeitschrift für medizinische Ethik

Wissenschaft · Kultur · Religion

## Rückschau auf Themen der Hefte 2021

- Reform der Pflegeberufe (1/2021)  
Verantwortlich für die allgemeine Pflege. Hintergründe, Ziele und zentrale Innovationen der Pflegeberufereform (Gertrud Hundeborn) – Interprofessionalität in der Pflegeausbildung: (inter-)nationale Reformen, Erfahrungen und Perspektiven (Michael Ewers/Bettina Glunde/Doreen Herinek) – Vorbehaltsaufgaben und Heilkundeübertragung: Pflege in neuer Verantwortung? (Thomas Klie) – Palliative Versorgung: die Bedeutung, die (interprofessionellen) Perspektiven zu wechseln und zu erweitern (Annette Riedel) – Kollektive Interessenvertretung von Altenpflegekräften aus Geschlechter-Perspektive (Diana Auth) – Die (Alten-)Pflege braucht eine Zukunft: Ein dringender Appell an die deutsche Politik (Hermann Brandenburg/Michael Bossle/Maik H.-J. Winter) – Alt und suchtkrank? Medizinethische Aspekte der Vernachlässigung von Substanzabhängigkeit im Alter (Selma Kadi/Hans-Jörg Ehni) – Die Deklaration von Genf (WMA) und ökonomische Aspekte: Zeit für eine Ergänzung des Gelöbnisses? (Thomas Kapitza)
- COVID-19-Pandemie (2/2021)  
Die Corona-Pandemie verstehen (Walter Schaupp) – Der vulnerable Mensch. Menschsein zwischen Autonomie und Angewiesenheit (Elisabeth Gräß-Schmidt) – Alter und COVID-19 (Andreas Kruse) – Pandemien als Herausforderung für die Gerechtigkeit (Franz-Josef Bormann) – Impfen in der Pandemie als moralische Duldungs- und Handlungspflicht (Wolfram Henn) – „Vaccine for all“ – Prüfstein globaler Bioethik (Stephan Sahm) – Die Regulierung der Pandemie als Grundrechtsproblem (Stephan Rixen) – Besuchsverbote in Pflegeeinrichtungen während der Corona-Pandemie. Infektionsschutz und Persönlichkeit (Bernhard Opolony) – Medizin und ärztliche Kunst (Marcus Knaup) – Das Konzept eines partizipativen Biomedizinrechts aus verfassungsrechtlicher Sicht (Josef Franz Lindner)

## Vorschau auf die nächsten Hefte:

- Ärztliche Suizidassistenz
- Klinische Herausforderungen der Demenz



SCHWABENVERLAG

Senefelderstraße 12 • 73760 Ostfildern bei Stuttgart  
Telefon (07 11) 44 06-134  
zfme@schwabenverlag.de  
www.zfme.de