

# Ärztliches Handeln im Spannungsfeld von Big Data, Künstlicher Intelligenz und menschlicher Erfahrung

Stellungnahme der Bioethikkommission



# **Ärztliches Handeln im Spannungsfeld von Big Data, Künstlicher Intelligenz und menschlicher Erfahrung**

Stellungnahme der Bioethikkommission

Wien 2020

## **Impressum**

Medieninhaber, Verleger und Herausgeber:  
Geschäftsstelle der Bioethikkommission, Ballhausplatz 2, 1010 Wien  
Wien, 2020. Stand: 18. Mai 2020

**Copyright und Haftung:** Auszugsweiser Abdruck ist nur mit Quellenangabe gestattet, alle sonstigen Rechte sind ohne schriftliche Zustimmung des Medieninhabers unzulässig. Es wird darauf verwiesen, dass alle Angaben in dieser Publikation trotz sorgfältiger Bearbeitung ohne Gewähr erfolgen und eine Haftung der Geschäftsstelle der Bioethikkommission und der Autorin/des Autors ausgeschlossen ist. Rechtsausführungen stellen die unverbindliche Meinung der Autorin/des Autors dar und können der Rechtsprechung der unabhängigen Gerichte keinesfalls vorgreifen.

## Inhalt

<b>1 Präambel</b> .....	<b>5</b>
<b>2 Einleitung</b> .....	<b>7</b>
<b>3 Ärztliches Handeln, Big Data und Künstliche Intelligenz</b> .....	<b>9</b>
3.1 Was macht ärztliches Handeln aus?.....	9
3.2 Psychosoziale Aspekte der Beziehung von Ärztin bzw. Arzt und Patientin bzw. Patient.....	10
3.2.1 Das ärztliche Gespräch.....	10
3.2.2 Die Dynamik der Beziehung von Ärztin bzw. Arzt und Patientin bzw. Patient.....	11
3.2.3 Die Bedeutung von Übertragung, Gegenübertragung sowie von unterschiedlichen Bindungsstilen.....	12
3.3 Rechtliche Verantwortung für ärztliches Handeln.....	14
3.3.1 Allgemeine Grundsätze über die Verantwortlichkeit für ärztliches Handeln.....	15
3.3.2 Intelligente Software als Medizinprodukt.....	16
3.3.3 (Unterlassene) Beachtung neuer Untersuchungsergebnisse.....	17
3.3.4 Verantwortungsvakuum bei Einsatz Künstlicher Intelligenz?.....	18
3.3.5 Datenschutzrechtliche Verantwortung.....	19
3.4 Ökonomische Triebkräfte und Argumente.....	21
3.5 Big Data: Quellen und Qualität.....	23
3.5.1 Was bedeutet „Big Data“?.....	23
3.5.2 Woher kommen die Daten für Big Data-Analysen in der Medizin?.....	25
3.5.3 Quantität und Qualität.....	26
3.5.4 Datenqualität und Künstliche Intelligenz (KI).....	27
3.6 Zentrale Begriffe: Assoziation, Korrelation und Kausalität.....	29
3.6.1 Begriffliche Neubestimmungen.....	29
3.6.2 Das Beispiel der genetischen Assoziationsanalyse.....	30
3.7 Maschinelles Lernen – Algorithmen – technische Grundlagen.....	32
3.7.1 Was ist maschinelles Lernen?.....	33
3.7.2 Datenkomplexität.....	35

3.7.3 Bias.....	36
3.7.4 Begründbarkeit.....	38
<b>4 Fallbeispiele.....</b>	<b>40</b>
4.1 Public Health – Gesundheitsplanung.....	40
4.2 Bildgebende Verfahren.....	41
4.3 Neurologie.....	41
4.4 Dermatologie.....	42
4.5 Ophthalmologie.....	43
4.6 Onkologie.....	43
4.7 Intelligente Aggregation und Visualisierung von Gesundheitsdaten.....	44
<b>5 Ethische Aspekte der Künstlichen Intelligenz.....</b>	<b>46</b>
5.1 Veränderungen der Beziehung von Ärztin bzw. Arzt und Patientin bzw. Patient.....	47
5.2 Wo liegt die Verantwortung bei maschinenbasierten Entscheidungen?.....	49
5.2.1 Ärztliche Handlungsspielräume.....	49
5.2.2 Das Kriterium der Systemkritikalität.....	50
5.2.3 Der Umgang mit Bias und fehlender Transparenz.....	50
5.3 Soziale Gerechtigkeit – die Unvermeidbarkeit distributiver Effekte?.....	52
5.4 Neue Anforderungen an Ärztinnen und Ärzte.....	53
<b>6 Schlussfolgerungen.....</b>	<b>54</b>
6.1 Verbesserung der medizinischen Versorgung durch Einsatz digitaler Technologien.....	55
6.2 Distributive Effekte und Bias.....	56
6.3 Veränderungen der Arzt-Patienten-Beziehung.....	56
6.4 Konsequenzen für Verantwortung und Systemdesign.....	57
6.5 Konsequenzen für die ärztliche Ausbildung.....	58
6.6 Empfehlungen.....	59
<b>7 Literatur.....</b>	<b>61</b>
<b>Abkürzungen.....</b>	<b>72</b>
<b>Mitglieder der Bioethikkommission für das Mandat 2017 bis 2020.....</b>	<b>74</b>

# 1 Präambel

Künstliche Intelligenz ist eines der wichtigen Schlagwörter unserer Zeit. Künstliche Intelligenz kann einerseits ungeahnte Verbesserungen unseres Lebens herbeiführen, andererseits Assoziationen zu einer dystopischen Zukunft, verbunden mit einem Gefühl der Hilflosigkeit hervorrufen. Einige sehen sie als ultimative Heilsversprechung, andere als große Gefahr an. Doch was beinhaltet der Begriff tatsächlich heute? Er wurde vor mehr als einem halben Jahrhundert geprägt und hat ursprünglich ausschließlich „denkende Maschinen“ bezeichnet. Heute betrachten wir Künstliche Intelligenz vor allem im Zusammenhang mit leistungsfähigen Computern.

Wenn auch Künstliche Intelligenz nichts grundlegend Neues darstellt, so hat neben den rasanten Entwicklungen auf dem Gebiet der Computertechnik vor allem die immer größere Verfügbarkeit von hohen Datenmengen – *Big Data* – in den letzten Jahren den Zugang extrem erleichtert und zu gesteigertem Interesse an der Materie geführt. Kaum eine Nation, kaum ein nationales oder internationales Beratungsgremium für (Bio-)Ethik haben es verabsäumt der Digitalisierung und Künstlichen Intelligenz mit der Abwägung von potentielltem Nutzen und Schaden breiten Raum zu widmen.

Auch die Bioethikkommission beim Bundeskanzleramt hat sich in der Vergangenheit mit den ethischen Aspekten dieses Themas intern und in öffentlichen Veranstaltungen auseinandergesetzt. So hat sie bereits im Jahr 2009 eine Stellungnahme zu „Assistiven Technologien – Ethische Aspekte der Entwicklung und des Einsatzes Assistiver Technologien unter Berücksichtigung älterer Menschen“ publiziert.

2016 fand eine öffentliche Kommissionssitzung im Bundeskanzleramt zu „Von Mensch und Maschine: Roboter in der Pflege“ statt. Nationale und internationale Expertinnen und Experten haben hier gemeinsam mit Mitgliedern der Kommission ethische Implikationen und gesellschaftliche Konsequenzen von maschineller Versorgung diskutiert. Diese Kommissionssitzung begründete eine weitere intensive Befassung mit dem Thema, das 2018 mit einer einhelligen Stellungnahme zu „Roboter in der Betreuung alter Menschen“ veröffentlicht wurde. Wenn auch dieses spezielle Thema einen Abschluss gefunden hat, so war eine weiterführende Diskussion zu *Big Data*, Künstlicher Intelligenz und Medizin als ein gemeinsamer Wunsch aller Mitglieder zu betrachten. Im Rahmen der österreichischen Präsidentschaft der EU im zweiten Halbjahr 2018 hat das 23. NEC-Forum (*National Ethics Committees Forum*) – das Zusammentreffen der Vorsitzenden und Leiterinnen bzw. Leiter der Geschäftsstellen der nationalen Bioethikräte und -Kommissionen in Wien unter Teilnahme der *European Group on Ethics in Science and New Technologies* (EGE), sowie Vertreterinnen und Vertretern des Europarates, der UNESCO und der WHO am 19. und 20. September in Wien stattgefunden. Hier haben auch Mitglieder der EGE ihre Stellungnahme zum Thema Künstliche Intelligenz vorgestellt.

Im 1. Halbjahr 2019 wurde das Thema „Digitalisierung & Ethik“ im Rahmen eines gemeinsamen Workshops von Bioethikkommission und dem Beratungsgremium des Bun-

deskanzlers Think Austria beim Pfindstdialog „Geist & Gegenwart“ in Schloss Seggau in der Steiermark unter Einbeziehung internationaler Expertinnen und Experten fortgeführt.

Es gibt heute kaum einen wissenschaftlichen Kongress, der nicht über Künstliche Intelligenz in aktuellen fachlichen Präsentationen berichtet. Öffentliche und private Gelder werden zunehmend auf diesem Gebiet investiert, das große Hoffnungen auf ein besseres Leben in einer zunehmend älter werdenden Gesellschaft setzt. Gerade auch der medizinische Bereich ist einer der großen Hoffnungsträger neuer Technologien. Immer häufiger gehören Künstliche Intelligenz und komplexe, von Ärztinnen und Ärzten selbst im Einzelnen nicht mehr nachvollziehbare Datenanalysen zu Standardverfahren, mit denen ärztliche Entscheidungen getroffen werden.

Was sind aber nun die speziellen ethischen Probleme, die sich in einer solchen Situation in der Medizin ergeben können? Viele Menschen befürchten, dass Probleme entstehen, wenn ärztliches Personal bei der Behandlung von Patientinnen und Patienten im Vertrauen auf das aus Künstlicher Intelligenz entstandene Assistenzsystem eine eigene Einschätzung ganz unterlässt. Zum anderen stellt sich bei einem Auseinanderfallen der Urteile Künstlicher Intelligenz und ärztlicher Erfahrung die Frage nach angemessenen Entscheidungsverfahren. Schließlich stellt sich auch die Frage der ethisch vertretbaren Programmierung ärztlicher Assistenzsysteme, bei denen medizinische und ökonomische Faktoren berücksichtigt werden können, versteckte Diskriminierung durch Algorithmen theoretisch nicht auszuschließen ist und permanent Entscheidungen getroffen werden, die potentiell Entscheidungen über Leben und Tod sein können.

Eine gelingende Medizin ist in ganz besonderer Weise vom wissenschaftlichen Fortschritt abhängig. Dieser Fortschritt kommt aber nicht nur aus dem innersten Bereich der Medizin, sondern aus der Interdisziplinarität der Lebenswissenschaften mit den Geisteswissenschaften. Vor noch 200 Jahren war das ärztliche Handeln sehr limitiert. Erst im ausgehenden 18. und beginnenden 19. Jahrhundert sieht die moderne Medizin ihre Geburtsstunde. Sie reflektiert über sich selbst, es entsteht der „ärztliche Blick“, eine andere Form der Wahrnehmung, hervorgegangen aus einer neuen Struktur zwischen den Wörtern und Dingen, begünstigt durch die Medizin in den breit entstehenden Krankenhäusern, in denen Patientinnen und Patienten von den Ärzten behandelt werden. Dort können sie – anders als in der vorangegangenen Medizin am isolierten Krankenbett – beobachtet und ihre Krankengeschichten verglichen werden. Die neue Beobachtung von Symptomen, die systematischen Methoden zur körperlichen Untersuchung und die möglicherweise folgende postmortale Sektion führten zu dieser neuen Sicht auf Krankheit sowie auf Patientinnen und Patienten. Hier beginnt auch die Standardisierung ärztlicher Kompetenz. In Wien legen die Errungenschaften der großen Ärzte in der I. und II. Wiener Medizinischen Schule über diese Fortschritte Zeugnis ab.

Diese Entwicklung geht seither in einem rasanten Tempo weiter. Ärztinnen und Ärzte mussten sich immer schon und müssen sich fortwährend den neuen Lehren stellen, sich weiterentwickeln, nicht stehen bleiben, sondern neue wissenschaftliche Erkenntnisse im täglichen Leben unter gleichzeitiger vernünftiger Achtung und Anwendung des Bewährten umsetzen. Die Schwierigkeit ist jedoch, dass das gewachsene Wissen heute



bereits so umfangreich ist, dass es dem/der Einzelnen nicht mehr möglich ist und sein kann im Moment der klinischen Entscheidung auf das gesamte vorhandene und relevante Wissen zuzugreifen.

Die Diskussion innerhalb der Bioethikkommission ist allerdings nicht nur auf wissenschaftlich generierte Erkenntnisse aus der Forschung konzentriert, sondern hat vor allem das ärztliche Handeln und seine Verantwortlichkeiten im Fokus. Dabei werden seine ethischen Prinzipien wie der Respekt vor der Autonomie der Patientinnen und Patienten, das Prinzip des Nicht-Schadens, Fürsorge sowie die Priorität des Patientenwohls und die Beachtung des Gerechtigkeitsprinzips zum Thema. Erkenntnisse aus und mit Künstlicher Intelligenz bilden dabei keine Ausnahme.

## 2 Einleitung

Digitale Technologien haben nicht nur die medizinische Forschung revolutioniert (z.B. durch die Möglichkeit zur Analyse von *Big Data*), die Entwicklung ganz neuer Behandlungsmethoden ermöglicht (z.B. personalisierte Medizin) und die Administration des Gesundheitssystems auf neue Beine gestellt (z.B. elektronische Patientenakte), sondern wirken sich auch auf das ärztliche und pflegerische Handeln im unmittelbaren Patientenkontakt aus.

Unter „digitale Technologien“ wird ein breites Spektrum an Werkzeugen und Services verstanden, die elektronische Informations- und Kommunikationstechnologien (IKT) verwenden. Im Gesundheitswesen wird in diesem Zusammenhang mitunter von „*Electronic Health*“ (*E-Health*) oder „*Digital Health*“ gesprochen, zu der z.B. folgende Bereiche zählen: Telemedizin, Teleberatung, *Telecare (Ambient Assisted Living, AAL)*, Telediagnostik, Telelabor, Telemonitoring, elektronische Medikation, digitale Dokumentation/Krankengeschichte, Gesundheitsportale, persönliches Gesundheitsmanagement, *Social Health Networks*, *E-Learning*, elektronische Fakturierung, E-Überweisung. Diese teilweise bereits im Routinebetrieb eingesetzten IKT werden durch neuere Entwicklungen von *Big Data*-Anwendungen, *Machine Learning* und Künstliche Intelligenz (KI/AI) weiter gestaltet. Ein wesentliches Werkzeug digitaler Technologien sind Algorithmen, welche die Grundlage aller Computerprogramme sind „algorithmische Entscheidungssysteme“ bezeichnen den gesamten Prozess von der Datenerfassung bis zur Ableitung eines Ergebnisses.

Erscheinungsformen umfassen dabei insbesondere:

- Die Unterstützung der **Befunderhebung** durch moderne Sensorik, wobei die Befunderhebung dank mobiler Endgeräte, vernetzter Umgebungen und automatisierter Datenverarbeitung auch weit in den häuslichen Bereich von Patientinnen

und Patienten verlagert wird und eine Entwicklung von der anlassbezogenen Maßnahme hin zu kontinuierlichem Monitoring zu beobachten ist.

- Die Unterstützung von **Diagnose und Therapieentscheidungen** durch fortgeschrittene algorithmische Systeme (regelbasierte Programmierung und/oder maschinelles Lernen), wobei ein Spektrum zu verzeichnen ist von an die Patientin bzw. den Patienten gerichteten Triage-Systemen über ärztliche Entscheidungsassistenten-Software bis hin zur algorithmisch determinierten und unmittelbar angewendeten Entscheidung.
- Die Unterstützung pflegerischer oder **medizinischer Eingriffe** durch Robotik-Anwendungen (s. dazu die Stellungnahme der BEK „Roboter in der Betreuung alter Menschen“).

Durch die Einbettung digitaler Technologien in den unmittelbaren Kontakt zwischen Ärztin bzw. Arzt und Patientinnen bzw. Patienten **verändern sich die wechselseitigen Rollen**. Auf der einen Seite machen digitale Technologien Patientinnen und Patienten viel stärker zu aktiven und selbstbestimmten Partnerinnen und Partnern im Prozess der Prävention, Diagnose und Therapie. Auf der anderen Seite geraten Ärztin bzw. Arzt immer stärker in die Rolle von Vermittelnden zwischen der Technik und den Patientinnen bzw. Patienten. Beide Entwicklungen haben unmittelbare Auswirkungen auf die ärztliche Aus- und Fortbildung ebenso wie auf die Ausstattung medizinischer Einrichtungen und die Struktur des Gesundheitssystems, jeweils einschließlich der rechtlichen Rahmenbedingungen. Die Eigenart des Arzt-Patienten-Verhältnisses stellt aber auch erhöhte Anforderungen an die **ethische Gestaltung** digitaler Technologien („*Ethics by Design*“). Dabei gelten die Ausführungen, die im Folgenden zum ärztlichen Handeln gemacht werden, analog auch für den Bereich der Pflege.

Hinzu kommt, dass moderne Medizin zunehmend zur datenintensiven Biomedizin geworden ist, ermöglicht durch eine Reihe technologischer Werkzeuge seit der Sequenzierung des Humangenoms und die Etablierung neuer Wissenschaftszweige, wie etwa die Bioinformatik oder die Datenwissenschaft (*data science*). Dadurch wurde die Komplexität von lebenden Systemen sehr deutlich und z. T. experimentell fassbar, sodass wir heute ein anderes Verständnis von Komplexität im Zusammenhang mit Gesundheit und Krankheit haben als Ende des 20. Jahrhunderts.

Die Möglichkeit, „Sinn“ aus großen Datenmengen durch maschinelle, zunehmend auch „lernende“ und evolutive, somit „intelligente“ Prozesse zu gewinnen, macht die Umsetzung der Visionen individualisierter (personalisierter) Medizin ebenso wie die eines patientenzentrierten, hocheffizienten Gesundheitssystems greifbarer als je zuvor, sofern es gelingt, wissenschaftliche Erkenntnisse der Anwendung auf eine den Menschen hilfreiche Weise zugänglich zu machen. Es liegt auf der Hand, dass dies eine immense Herausforderung mit hoher gesellschaftlicher Brisanz und einer Reihe ethischer Fragestellungen ist.

# 3 Ärztliches Handeln, Big Data und Künstliche Intelligenz

Im Mittelpunkt dieser Stellungnahme stehen die Implikationen von *Big Data*, Künstlicher Intelligenz und maschinellem Lernen, sowie ihr Einsatz für Diagnose und Therapieplanung auf ärztliches Handeln. Sie gibt deshalb Raum für grundsätzliche Überlegungen zu ärztlichem Handeln und ärztlicher Verantwortung (auch aus rechtlicher Perspektive), zur Beziehung von Ärztin bzw. Arzt und Patientin bzw. Patient aus psychosozialer Sicht, sowie zum ökonomisch motivierten Bestreben, das Gesundheitswesen effizienter und effektiver zu gestalten.

Ziel der Stellungnahme ist es, Ärztinnen und Ärzten sowie Entscheidungsträgerinnen und Entscheidungsträgern im Gesundheitswesen

1. einen Einblick in die Thematik, ihre technisch-mathematischen, medizinischen und rechtlichen Aspekte zu vermitteln;
2. dabei insbesondere die mit dem Einsatz von *Big Data*, Künstlicher Intelligenz und maschinellem Lernen verbundenen ethischen Fragen zu erläutern; und
3. in einem Empfehlungsteil Handlungsbedarf auf verschiedenen Ebenen aufzuzeigen.

Obwohl auch andere Gesundheitsberufe von dieser Thematik betroffen sind, konzentriert sich diese Stellungnahme auf das ärztliche Handeln.

## 3.1 Was macht ärztliches Handeln aus?

Im Vergleich mit den bereits vertrauten Formen der Evidenzbasierung stellt der Einsatz von *Big Data*, Künstlicher Intelligenz und maschinellem Lernen Ärztinnen und Ärzte vor teilweise neue Anforderungen. Diese betreffen das Sammeln und Dokumentieren von Daten, sowie die Interpretation der computergenerierten Diagnose- und Therapievor schläge. Im Rahmen klinisch orientierter Forschung geht es zudem auch um Teilen und Austausch von Daten.

## 3.2 Psychosoziale Aspekte der Beziehung von Ärztin bzw. Arzt und Patientin bzw. Patient

Ärztliches Handeln beginnt im Wartezimmer oder vielleicht schon am Telefon bei der Terminvereinbarung; eine lange Wartezeit kann bereits die Einstellung der Patientin bzw. des Patienten der Ärztin bzw. dem Arzt gegenüber verändern. Ärztliches Handeln setzt sich fort beim Begrüßen, beim in die Augen schauen, bei der Stimmlage, beim Händedruck und wird mit einer bestimmten Art zuzuhören fortgesetzt: empathisch, rezeptiv, Geborgenheit vermittelnd. Ärztliches Handeln beginnt somit viel früher als gedacht.

Die gute Kommunikation zwischen Ärztin bzw. Arzt und Patientin bzw. Patient ist ein wesentliches Element der klinischen Funktion. Sie ist das Herzstück der Heilkunst und ein zentraler Teil in der Erfüllung der medizinischen Versorgung (Epstein 2017; Street et al. 2009; Balint und Shelton 2002). Sie setzt sich zusammen aus einer guten zwischenmenschlichen Beziehung, gelungenem Informationsaustausch und, damit verbunden, der Fähigkeit und Möglichkeit der Patientin bzw. des Patienten eine Entscheidung zu treffen (Borck 2016; Epstein 2017; Müller-Mielitz 2017; Hehner et al. 2018).

### 3.2.1 Das ärztliche Gespräch

Ärztinnen und Ärzte sind meist umfassend ausgebildet, aber es mangelt ihnen teilweise an dialogischer und psychosozialer Kompetenz. In einer viel zitierten Studie befanden Beckman und Frankel (1984), dass Ärztinnen und Ärzte Patientinnen und Patienten im Gespräch nach durchschnittlich 18 Sekunden unterbrachen. Die Schwierigkeiten von Ärztinnen und Ärzten in der Kommunikation mit Patientinnen und Patienten werden auch durch neue Studien belegt. Beispielsweise verweisen Street und Haidet (2011) darauf, dass Ärztinnen und Ärzte häufig nur ein unzureichendes Verständnis der Einstellungen und Werte ihrer Patientinnen und Patienten entwickeln, wenn diese sich nicht aktiv an der Kommunikation beteiligen.

1. Es gibt viele Hindernisse für eine gelungene Kommunikation. Von ärztlicher Seite ist dies nicht nur das große Arbeitspensum. Ärztinnen und Ärzte arbeiten oft in einem Setting der Angst: die Angst vor unrealistischen Erwartungen und vor der Angst von Patientinnen und Patienten, sowie vor der eigenen Ohnmacht. In dieser Atmosphäre ist es somit oft schwierig, dem Nichtwissen gegenüber offen zu sein und dem Unerwarteten zu begegnen.
2. Schlechte Nachrichten überbringen müssen und das Aushalten-Müssen von (starken) Emotionen ist oft bedrückend und kostet Substanz. Hingegen ist das Teilen-Können von Entscheidungen auf der Basis informierten Konsenses oder bei besonders komplizierten medizinischen Informationen für Ärztinnen und Ärzte aufbauend und zufriedenstellend. Patientinnen und Patienten sind häufig passiv, selten sagen sie ihre Meinung oder beginnen Diskussionen über Aspekte ihrer Behandlung. Zudem haben Ärztinnen und Ärzte häufig Schwierigkeiten, eine solche partizipative Kommunikationssituation herzustellen (Ubel et al.

2017). Gründe dafür sind u. a. „*failing to communicate clearly to patients about decision-relevant information, overwhelming patients with irrelevant information, overlooking when patients' emotions made it hard to engage in choices, and making recommendations before discussing patients' goals*“ (Ubel et al. 2017, 31). Gelungene Kommunikation hilft Patientinnen und Patienten sich an ihre Krankheit und der dazugehörigen Therapie anzupassen. Es gilt daher ihre Entscheidungskompetenz zu stärken.

3. Ärztinnen und Ärzte neigen dazu, ihr eigenes Wissen sowie die Möglichkeiten der Patientinnen und Patienten dieses zu verstehen zu überschätzen (Tversky/Kahneman 1974). So kann der allgemein hohe Bekanntheitsgrad medizinischer Begriffe Ärztinnen und Ärzte dazu verleiten voranzusetzen, dass ihre Patientinnen und Patienten ihre Erklärungen verstehen (Gundling et al. 2019). Hinzu kommt die Tendenz von Ärztinnen und Ärzten den Nutzen dessen, was sie selbst tun zu überschätzen. Dies wird als „therapeutische Illusion“ bezeichnet. Nicht untypisch für ärztliches Handeln ist, was Epstein (2017) als Bestätigungsverzerrung (*confirmation bias*) definiert, nämlich der schwer zu korrigierende Glaube an die Nützlichkeit ihrer eigenen Verordnungen (Epstein/Pro-Publica 2017). Oft wird, was im Moment im Gedächtnis der Ärztinnen und Ärzte verfügbar ist, als Diagnose gewählt.
4. Wir wissen, dass das Reden selbst therapeutisch wirken kann, indem es Ängste reduziert sowie Tröstung und Ermunterung bringt. Aber oft ist die Wirkung nicht direkt, sondern indirekt aus dem Verstehen der Patientinnen und Patienten, ihrem Vertrauen und Einverständnis ersichtlich (Street et al. 2009). Dies führt zu der Erkenntnis, dass die Ärztin bzw. der Arzt selbst ein „Medikament“ ist von dem wir oft nicht wissen, wie häufig es verabreicht werden muss, wie die Dosierung zu sein hat und welche Nebenwirkungen erwartet werden können (s.u.).

### 3.2.2 Die Dynamik der Beziehung von Ärztin bzw. Arzt und Patientin bzw. Patient

In seinem klassischen Buch „Der Arzt, sein Patient und die Krankheit“ identifizierte Balint (1964) sechs wesentliche Kräfte, die die Entwicklung der Arzt-Patienten-Beziehung beeinflussen:

- Die „Basisstörung“, ein Konzept aus der Psychoanalyse, ist eine Reaktion auf Erlebnisse aus der frühen Kindheit. Sie bestimmt die individuelle Reaktion des Einzelnen auf Stress und beeinflusst die Art, wie Personen auf ihre Umgebung reagieren. Ärztinnen und Ärzte, die sich selber kennen können ihre Patientinnen und Patienten besser verstehen. (Die „Basisstörung“ ist bei jedem Menschen vorhanden.)
- Die „apostolische Funktion“ von Ärztinnen und Ärzten nimmt Bezug auf ihren Paternalismus. Ärztinnen und Ärzte tendieren dazu, sobald Patientinnen und Patienten ihre Krankheit präsentiert haben, eine unerschütterlich feste Vorstellung

davon zu entwickeln, was für sie das Richtige ist, was sie hoffen sollen, dulden müssen und wie sie sich zu verhalten haben.

- Die gegenseitige „Investmentfirma“: Es kommt also dazu, dass Ärztinnen bzw. Ärzte und Patientinnen bzw. Patienten sich gegenseitig „erziehen“. Balint nennt das eine „gegenseitige Investmentfirma“. Wenn dies glückt, wird daraus eine gute, vertrauensvolle Beziehung, die für beide Parteien nützlich ist.
- Die Rolle von Ärztinnen und Ärzten als therapeutisches Agens: Das „Medikament Arzt“ ist immer im Einsatz, doch wir wissen sehr wenig über dessen pharmakologische Wirkung und Nebenwirkung oder ihre Dosierung. Jedenfalls wissen wir heute, dass der Placebo Effekt eine wesentliche Rolle spielt.
- Die „tiefere Diagnose“: Die gegenseitige „Investment Firma“ ist die Grundlage dafür, dass Ärztinnen und Ärzte mehr über Patientinnen und Patienten erfahren und auf diese Art und Weise eine „tiefere gesamtheitlichere Diagnose“ erstellen können.
- Die „Anonymitätsklausel“: Patientinnen und Patienten, die mehrere undefinierte Krankheitsbilder zeigen, werden häufig zur Abklärung zu verschiedenen Spezialistinnen und Spezialisten geschickt. Wenn keinem dieser Spezialistinnen und Spezialisten eine Abklärung der Symptome glückt, kommt es häufig zu einer Spirale von „entwarnenden“ Briefen und erneuten Arztbesuchen, wenn weitere Symptome auftauchen. Balint (1964) spricht von einer „Anonymitätsklausel“, einem geheimen Einverständnis, bei dem sich keiner der beteiligten Ärztinnen und Ärzte in der gebotenen Weise für die betreffenden Patientinnen und Patienten zuständig fühlt.

Wir können also feststellen, dass die Beziehung von Ärztin bzw. Arzt und Patientin bzw. Patient eine leicht störbare ist, und so wundert es nicht, dass diese Beziehungen nicht selten gespannt, unglücklich und unerfreulich sind.

### 3.2.3 Die Bedeutung von Übertragung, Gegenübertragung sowie von unterschiedlichen Bindungsstilen

Wissenschaftliche Studien haben gezeigt, dass ein signifikanter Prozentsatz der physisch kranken Menschen unter starken Ängsten und Depressionen leiden: Da sind gute Übertragungs- und Gegenübertragungsmechanismen – eine geglückte Bindung – von besonderer Wichtigkeit. Vorausgeschickt sei dieser Betrachtung allerdings, dass es ein aus guten Gründen verbrieftes Recht von Patientinnen und Patienten auf Selbstbestimmung gibt, welches auch das Recht auf Nichtwissen sowie auf Unvernunft miteinschließt. Es mag gute Gründe geben, eine Therapie abzulehnen oder abubrechen, sodass derartige Entscheidungen nicht notwendigerweise dem Bindungsstil von Patientinnen und Patienten zuzuschreiben sind.

Als **Übertragung** wird die Beziehung von Patientinnen und Patienten zur Ärztin und zum Arzt auf der Grundlage der früheren Beziehungen zu wichtigen Bezugspersonen angesehen. Wenn Patientinnen und Patienten die vorgeschlagene Therapie nicht annehmen, wenn sie vorzeitig entlassen werden wollen, kann dies unter Umständen darauf

zurückzuführen sein, dass bei den handelnden Menschen die Atmosphäre nicht stimmt. Man muss dann in Betracht ziehen, dass eventuell auf Grund frühkindlicher Erfahrungen die Beziehung zu wesentlichen Bezugspersonen gestört gewesen sein könnte.

Als **Gegenübertragung** bezeichnen wir Gefühle und Reaktionen der Ärztinnen und Ärzte ihren Patientinnen und Patienten gegenüber. Auch sie ist von frühen Beziehungen beeinflusst und kann die Beziehung von Ärztin und Arzt und Patientin und Patient sowohl positiv als auch negativ beeinflussen. Ärztinnen und Ärzte müssen daher über ein gewisses Ausmaß an Introspektion verfügen, um die eigenen Gefühle und Widerstände zu erkennen und zu verstehen. Dazu muss ärztlicherseits sowohl das Verhalten von Patientinnen und Patienten, aber auch das eigene Empfinden (inklusive schwer aushaltbarer Gefühle wie Wut, Hilflosigkeit oder besonderer emotionaler Betroffenheit) beobachtet und hinterfragt werden.

An dieser Stelle ist es hilfreich, einen Blick in die Ergebnisse der Bindungsforschung zu werfen. Auf der Grundlage der von John Bowlby und Mary Ainsworth (2003) entwickelten Bindungstheorie haben Bartholomew und Horowitz (1991) vier Prototypen des erwachsenen Bindungsverhaltens definiert:

- Der **sichere Bindungsstil** ist bei Menschen zu beobachten, die warme, empathische Bezugspersonen hatten, die ihnen ein grundlegendes Gefühl der Sicherheit vermitteln konnten. Die Patientin bzw. der Patient wirkt im Kontakt aufgeschlossen, offen und kooperativ.
- Der **unsicher-ambivalente (*preoccupied*) Bindungsstil** ist durch ein abhängiges Verhalten der Patientin bzw. des Patienten, einen stärkeren Ausdruck von Gefühlen und einem intensiven Bedürfnis nach emotionaler Nähe charakterisiert. Ein chronisches Gefühl der Verletzbarkeit oder Gefahr lässt im medizinischen Setting diese Patientinnen und Patienten ihren körperlichen Symptomen eine sehr große Aufmerksamkeit widmen. Es kann daher für Ärztinnen und Ärzte schwer werden, harte Fakten in der Krankengeschichte zu erheben.
- Der **unsicher-vermeidende (*dismissive*) Bindungsstil**: Diese Menschen erlebten in ihrer Kindheit, dass sie viel zu wenig umsorgt und beschützt wurden. Diese Patientinnen und Patienten reagieren in Situationen, die ein gewisses Ausmaß an Intimität und Abhängigkeit verlangen, eher zurückhaltend, abweisend und eher misstrauisch, weil sie sich ausgeliefert fühlen.
- Der **unsicher-desorganisierte (*fearful*) Bindungsstil**: In der Kindheit dieser Menschen wurden seine Bedürfnisse von den Bezugspersonen abgewiesen bzw. es wurde ihnen mit Härte begegnet. Im Krankheitsfall kommt es häufig zu einer unsicher-ambivalenten Antwort der Ärztin bzw. dem Arzt gegenüber. Die Patientinnen und Patienten signalisieren Leidensdruck, lehnen aber medizinische Behandlung ab. Diese Patientinnen und Patienten brauchen einen sehr behutsamen Vertrauensaufbau und Gelassenheit der behandelnden Ärztin bzw. des behandelnden Arztes.

Manchmal müssen Ärztinnen und Ärzte organisiert, wissenschaftlich und entscheidungsfreudig sein. Dann wiederum müssen sie mehr auf ihre eigene Reaktion der Patientin bzw. dem Patienten gegenüber achten, z. B. etwas Überraschendes, eine atmosphärische Spannung, Nichtwissen, Konflikte, die mit den derzeitigen Problemen der Patientin bzw. des Patienten zu tun haben, wahrnehmen. Die Möglichkeit zu häufigen kurzen Kontakten in Kombination mit einem Gefühl des langzeitigen Engagements erlaubt es Patientinnen bzw. Patienten, die Ärztin bzw. den Arzt im eigenen Rhythmus, Tiefe und Tempo ins Vertrauen zu ziehen und erlaubt auch der Ärztin bzw. dem Arzt, die notwendige Grenze ihres bzw. seines Engagements zu ziehen. Die „ideale“ Konsultation erlaubt das Entstehen eines besonderen Augenblicks des Einvernehmens.

Zusammenfassend geht es sowohl im emotionalen Bereich als auch in Bezug auf die Umsetzung heutiger ärztlicher Ethik darum, eine vertrauensvolle und kreative, menschliche Beziehung/Bindung herzustellen.

### 3.3 Rechtliche Verantwortung für ärztliches Handeln

Ziel jedweden ärztlichen Handelns ist es, entsprechend dem Stand der medizinischen Wissenschaft und ärztlichen Erfahrung zum Wohl der Kranken und Gesunden zu agieren (vgl. § 49 ÄrzteG).

Wird entgegen solchem Handeln *lege artis* einer Patientin bzw. einem Patienten ein **Schaden** zugefügt, kann dies zu einer **zivil- und allenfalls auch strafrechtlichen Haftung** führen, sowohl des **Individuums** als auch unter Umständen der **juristischen Person**. Dabei wird stets im Nachhinein (**ex post**) – allenfalls unter Heranziehung gutachterlicher Expertise – seitens der Justiz **beurteilt**, ob ärztliches Handeln **im Zeitpunkt ex ante fachlich korrekt** durchgeführt und alles Erforderliche zur Risikovermeidung gesetzt wurde. Eine Zusicherung der Justiz im Vorhinein dahingehend, dass ein konkretes Handeln unter allen Umständen auch rechtlich korrekt ist, geschieht nicht. Dies ist bei technologieunterstützten Operationsmethoden oder computerunterstützter Diagnoseerstellung nicht anders als bei einer klassischen medizinischen Untersuchung und Behandlung. Vor diesem Hintergrund soll im Folgenden zunächst auf die allgemeinen Haftungsgrundsätze für ärztliches Handeln eingegangen werden und daran anschließend auf die Haftung infolge (gesetzter oder unterlassener) Heranziehung neuer wissenschaftlicher Methoden.

Im Rahmen von *Big Data* ist die behandelnde Ärztin bzw. der behandelnde Arzt auch ein **Knotenpunkt**, bei der/dem sowohl **Daten gesammelt** als auch an die/den **Daten übermittelt** und durch die bzw. den Daten **weitergegeben** werden. Diese Daten werden in weiterer Folge gespeichert, sodass infolge der Sensibilität dieser Daten eine erhebliche rechtliche Verantwortung auf der Ärztin bzw. dem Arzt dahingehend lastet, wie sie/er mit diesen Daten umzugehen hat. Darüber hinaus werden von der Ärztin bzw. vom Arzt verschiedene **Daten angefordert**, was die skizzierte Verantwortung für die Wahrung und Verarbeitung von Daten noch unterstreicht. Die Regelungen zum rechtlichen Umgang mit Gesundheitsdaten sind als solche keineswegs neu, und die mit dem



Datenschutz einhergehende Verantwortung ist in den vergangenen Jahren, insbesondere im Zusammenhang mit der Datenschutzgrundverordnung (DSGVO), bereits deutlich unterstrichen worden. Gleichzeitig muss ein breites Bewusstsein über das große Potential von – jeweils professionell – aufgenommenen, kuratierten und vernetzten Daten in der Medizin geschaffen werden.

### 3.3.1 Allgemeine Grundsätze über die Verantwortlichkeit für ärztliches Handeln

Die **individualrechtliche Verantwortlichkeit** basiert – sowohl im zivil- als auch strafrechtlichen Bereich – auf dem Vorwurf der sozialen Missbilligung eines Verhaltens, die grundsätzlich im Verschulden zum Ausdruck kommt. Daneben gibt es auch eine verschuldensunabhängige Haftung, wie etwa die Produkthaftung oder auch die Gefährdungshaftung. **Juristische Personen** haften im Zivilrecht nach den gleichen Grundsätzen, wobei ihnen das Verschulden ihrer sogenannten Organe und Machthaber (etwa Auswahl- oder Überwachungsverschulden), im vertraglichen Kontext (etwa eines Behandlungsvertrags) sogar das Verschulden jedes Gehilfen zugerechnet wird. Im strafrechtlichen Bereich ist die Verantwortlichkeit von juristischen Personen (Verbandsverantwortlichkeit) im Wesentlichen strafrechtsakzessorisch von einem vorwerfbaren Verhalten der konkret handelnden Person (Individualhaftung) abhängig. Darüber hinaus muss auch hier eine Art Organisationsverschulden hinzukommen, als entweder die Tat zu Gunsten des Verbandes begangen worden sein muss oder durch die Tat Pflichten verletzt worden sind, die den Verband treffen (vgl. § 3 Abs. 1 VbVG). Ob ein Verhalten juristisch nicht korrekt war, ist zunächst danach zu beurteilen, ob dessen Verbot in einem Gesetz oder in einer Verordnung eine positivrechtliche Anerkennung gefunden hat (**Verletzung einer Rechtsvorschrift**). Fehlt es an derartigen formalrechtlichen Vorgaben, werden subsidiär berufsspezifische Sorgfaltsregeln herangezogen (**Missachtung der *leges artis***). In diesem Zusammenhang kann Leitlinien von Fachgesellschaften oder anderen „Sachverständigen-Konsens-Standards“ wesentliche Bedeutung zukommen (s. allgemein zur Thematik von Leitlinien und strafrechtliche Haftung Birklbauer 2019). Fehlen zu einem Tätigkeitsbereich solche Standards, kommt als dritte Begründungsmöglichkeit für ein sorgfaltswidriges Verhalten das Fehlen des gefährlichkeitsvermeidenden Verhaltens in Betracht, das ein durchschnittlicher bzw. sorgfältiger Anderer aus dem Verkehrskreis des Täters gesetzt hätte (**Verhalten eines durchschnittlichen bzw. sorgfältigen Anderen**).

Auf die für den Sorgfaltsmaßstab beschriebene Heranziehung von Fachstandards und Leitlinien verweist für den ärztlichen Bereich letztlich auch der bereits erwähnte **§ 49 Abs. 1 ÄrzteG**. Demnach hat eine Ärztin bzw. ein Arzt „unter Einhaltung der bestehenden Vorschriften und der fachspezifischen Qualitätsstandards ... das Wohl der Kranken und den Schutz der Gesunden zu wahren“. Es bildet somit ein Indiz für einen Sorgfaltsverstoß, wenn die konkrete Ärztin bzw. der konkrete Arzt, z. B. bei der Erstellung der medizinischen Indikation oder auch im Rahmen einer konkreten Behandlung, vom allgemeinen Standard, wie er in der (internationalen) Kollegenschaft akzeptiert ist, abgewichen ist. Der für eine Berufsgruppe geltende Standard ist somit quasi der Maß-

stab für eine rechtliche Verantwortung. Umgekehrt kommt **Leitlinien** gleichsam eine **Sorgfaltsvermutung** zu. Die Ärztin bzw. der Arzt, die/der entsprechend den Leitlinien behandelt, agiert sorgfaltsgemäß, solange sie/er nicht Anhaltspunkte hat, dass die in den Leitlinien empfohlenen Grundsätze mittlerweile nicht mehr den *leges artis* entsprechen. Insofern lastet auch eine gewisse Verantwortung auf den Fachgesellschaften, die von ihnen zusammengefassten Leitlinien an den aktuellen fachlichen Gegebenheiten zu orientieren. Gerade im Zusammenhang mit Forschungsmethoden, bei denen es infolge der komplexen Datenstruktur schwierig ist, deren Qualität zu durchblicken, ist dies eine beträchtliche Herausforderung.

Vor diesem Hintergrund kommt z. B. medizinischen Computerprogrammen, die als **Medizinprodukt** eingestuft werden, eine gewisse Vertrauensvermutung zu, ist doch dieser Einstufung ein rechtlich etabliertes Konformitätsbewertungsverfahren vorausgegangen (vgl. 3.3.2).

### 3.3.2 Intelligente Software als Medizinprodukt

Die Heranziehung „Künstlicher Intelligenz“ durch Ärztinnen und Ärzte erfolgt nicht im rechtsfreien Raum oder auf rein experimenteller Basis. Vielmehr gilt Software, einschließlich jeglicher Künstlichen Intelligenz als **Medizinprodukt**, wenn sie vom Hersteller speziell für einen oder mehrere der in der Definition von Medizinprodukten genannten medizinischen Zwecke bestimmt ist. Dagegen ist Software für allgemeine Zwecke, auch wenn sie in Einrichtungen des Gesundheitswesens eingesetzt wird (z. B. Büro-Software) grundsätzlich ebenso wenig ein Medizinprodukt wie es Lifestyle-Produkte (z. B. Fitness-App) sind.

Medizinprodukte dürfen nur dann in Verkehr gebracht und auch grundsätzlich nur dann von der Ärztin bzw. vom Arzt verwendet werden, wenn sie das vorgeschriebene **Konformitätsbewertungsverfahren** durchlaufen haben und die „Grundlegenden Anforderungen“ erfüllen. Die Grundlegenden Anforderungen sind bisher im jeweiligen Anhang I der Richtlinien 90/385/EWG (Aktive implantierbare Medizinprodukte), 98/79/EG (In-vitro-Diagnostika) und 93/42/EWG (Sonstige Medizinprodukte) festgelegt. Mit Wirkung ab 26. Mai 2020 werden diese Richtlinien (und die jeweiligen staatlichen Umsetzungsakte) durch die Verordnung (EU) 2017/745 über Medizinprodukte (*Medical Device Regulation*, MDR) sowie durch die Verordnung (EU) 2017/746 ersetzt.

Welches Konformitätsbewertungsverfahren durchzuführen und in welchem Umfang dabei eine unabhängige Prüf- und Zertifizierungsstelle (sog. „Benannte Stelle“) zu beteiligen ist, hängt vom potentiellen Risiko der Produkte ab. So sieht die MDR (in Übereinstimmung mit der Richtlinie 93/42/EWG) eine **Differenzierung der Produkte in vier Klassen** vor (I, IIa, IIb, III), wobei die Klassifizierung nach den im Anhang VIII festgelegten Kriterien erfolgt. Die Konformitätsbewertung erfolgt dabei grundsätzlich gemäß Anhang IX zur MDR auf der Grundlage eines Qualitätsmanagementsystems und einer Bewertung der technischen Dokumentation.

Die MDR enthält auch umfangreiche Bestimmungen über Software. Solche Software, die ein Medizinprodukt steuert oder dessen Anwendung beeinflusst, wird derselben Klasse zugerechnet wie das Produkt. Ist die Software von anderen Produkten

unabhängig, so wird sie für sich allein klassifiziert. Software, die für die **Kontrolle von physiologischen Prozessen** bestimmt ist gehört zur Klasse IIa, es sei denn, sie ist für die Kontrolle von vitalen physiologischen Parametern bestimmt, deren Änderung zu einer unmittelbaren Gefahr für die Patientin bzw. den Patienten führen könnte (dann Klasse IIb). Software, die dazu bestimmt ist, Informationen zu liefern, die zu **Entscheidungen für diagnostische oder therapeutische Zwecke** herangezogen werden, gehört grundsätzlich zur Klasse IIa. Kann durch die Entscheidung eine schwerwiegende Verschlechterung des Gesundheitszustands einer Person bewirkt oder ein chirurgischer Eingriff erforderlich werden, wird Software der Klasse IIb zugeordnet. Wenn die betreffende Entscheidung Auswirkungen haben kann, die den Tod oder eine irreversible Verschlechterung des Gesundheitszustands einer Person zur Folge haben kann, wird sie der Klasse III zugeordnet. Sämtliche **andere Software** wird der Klasse I zugeordnet.

### 3.3.3 (Unterlassene) Beachtung neuer Untersuchungsergebnisse

Zur Frage des Zusammenhangs zwischen Abweichungen von einem definierten Standard und einer Fahrlässigkeitshaftung kann auch auf die für den **Off-Label-Use** von Behandlungsmethoden oder verwendeten Medikamenten erarbeiteten Grundsätze zurückgegriffen werden (dazu Kopetzki 2008; Mayrhofer 2014). Die fehlende Zulassung eines Medikaments für eine bestimmte Behandlung bedeutet nicht zwingend, dass dessen Verabreichung keine Behandlung *lege artis* darstellen würde und deshalb auch nach § 49 Abs. 1 ÄrzteG unzulässig wäre. Im Gegenteil kann die Anwendung eines solchen Medikaments sogar geboten sein, sofern die Verabreichung des Medikaments nach dem jeweiligen Stand der medizinischen und pharmazeutischen Wissenschaft indiziert und Erfolg versprechend ist. Allerdings trifft die Ärztin oder den Arzt, die bzw. der ein Medikament „*off label*“ anwendet, eine höhere Begründungslast, weil sie bzw. er die Anwendung nicht „einfach“ mit der Zulassung rechtfertigen kann, sondern andere Quellen anführen muss. Hier spielen neben wissenschaftlichen Publikationen wiederum Leitlinien von Fachgesellschaften eine wichtige Rolle, die durch die breite Verfügbarkeit von Anwendungsdaten (*Real World Data*) entscheidend unterstützt wird.

Hat Software kein entsprechendes Konformitätsbewertungsverfahren durchlaufen, will die Ärztin bzw. der Arzt aber dennoch eine diagnostische und/oder therapeutische Entscheidung auf algorithmische Berechnungen stützen, ist Vorsicht angezeigt. Das Heranziehen neuer Behandlungsmethoden oder -ergebnisse kann im Einzelfall eine (zivil- oder straf-)rechtliche Verantwortlichkeit begründen, wenn darin eine Sorgfaltswidrigkeit zu sehen ist. Werden beispielsweise neue, noch **nicht hinreichend erforschte Methoden angewendet** und erleidet die Patientin bzw. der Patient dadurch einen Schaden, ergibt sich eine rechtliche Verantwortlichkeit, wenn sich solche Behandlungen im Bereich des Experimentellen befinden, denen eine hinreichende Evidenzbasierung fehlt. Eine Haftung könnte allenfalls dann ausscheiden, wenn die Patientin bzw. der Patient unter hinreichender Kenntnis des vollen Risikos in die möglichen Folgen eingewilligt hat, wobei im strafrechtlichen Bereich die rechtfertigende Einwilligung durch das Sittenwidrigkeitskorrektiv begrenzt ist (vgl. § 90 StGB).

Auf der anderen Seite kann eine rechtliche Haftung aber auch dadurch begründet werden, dass der Patientin bzw. dem Patienten eine **neue, erfolgsversprechende Behandlungsmethode vorenthalten** wird, weil sich die Ärztin oder der Arzt mit dieser noch nicht auseinandergesetzt hat und sich damit auch nicht auseinandersetzen will. Eine Haftung würde nur dann entfallen, wenn die entsprechende Behandlungsmethode nicht zugänglich ist bzw. diese auch nicht organisiert werden kann, denn niemand kann zu einem Tun verpflichtet werden, das er faktisch nicht anbieten kann.

Für die umfassende und eingehende Information von Patientinnen und Patienten über mögliche Behandlungen, welche die Basis für seine Zustimmung oder Ablehnung bilden kann, ist es erforderlich, dass die **Ärztin bzw. der Arzt** selbst sich **hinreichend fortgebildet** hat, um derartige Möglichkeiten überhaupt aufzeigen und erklären zu können. Gerade beim Heranziehen von algorithmischer Unterstützung bei Diagnosen und Behandlungen wird es in diesem Kontext erforderlich sein, dass die behandelnde Ärztin bzw. der behandelnde Arzt hinreichende Kenntnis über das Zustandekommen der entsprechenden Daten, des dahinterstehenden Algorithmus usw. hat bzw. dass sie/er auf die Qualität von publizierten Untersuchungen und den verwendeten Methoden vertrauen kann. Vor diesem Hintergrund kommt etwa auch dem einer Publikation vorausgehenden Review-Prozess besondere Bedeutung zu. Dabei bildet den Sorgfaltsmaßstab wiederum grundsätzlich die bzw. der **mit Durchschnittswissen und Durchschnittkenntnissen ausgestattete Ärztin bzw. Arzt**, die/der auch ihrer/seiner regelmäßigen Fortbildungsverpflichtung nachgekommen ist. Bei allen rechtlichen Haftungsüberlegungen ist jedoch in diesem Kontext zu bedenken, dass der juristische Haftungsmaßstab nicht überdehnt werden darf. Sorgfaltswidrigkeit bedeutet Nichteinhaltung der „gebotenen“ Sorgfalt, und nicht Vernachlässigung der „möglichen“ Sorgfalt.

### 3.3.4 Verantwortungsvakuum bei Einsatz Künstlicher Intelligenz?

Im Zusammenhang mit der zivilrechtlichen Haftung bei Einsatz Künstlicher Intelligenz wird vielfach ein Verantwortungsvakuum befürchtet. Die Ärztin bzw. der Arzt (bzw. der Träger der Krankenanstalt, bei welcher die Ärztin bzw. der Arzt beschäftigt ist) haftet nämlich im Rahmen der Verschuldenshaftung nur für eigenes Verschulden und für Verschulden von Menschen, die bei der Erfüllung der aus dem Behandlungsverhältnis fließenden Pflichten eingesetzt werden (§ 1313a ABGB). Kommt es dagegen nicht zum Einsatz von Menschen, sondern von Maschinen, wird an sich nur für **eigenes Verschulden der Ärztin bzw. des Arztes gehaftet**, welches etwa in einer sorgfaltswidrigen Auswahl, Bedienung, Aktualisierung oder Überwachung der betreffenden Software bestehen kann. Hat sich die Ärztin bzw. der Arzt einer anerkannten und zertifizierten Software bedient, diese vorschriftsgemäß eingesetzt und musste sie/er auch nichts von Fehlleistungen der Software wissen, haftet die Ärztin bzw. der Arzt (bzw. der Träger der Krankenanstalt) an sich **nicht für Fehlleistungen der Software als solcher**. Daher wird von einer im Vordringen befindlichen Meinung gefordert, § 1313a ABGB analog auf Fehlleistungen intelligenter Software zu erstrecken (i. d. S. *Expert Group on Liability and New Technologies, Report on Liability for Artificial Intelligence and other Emerging Digital Technologies – New*

*Technologies Formation* 2019, Empfehlungen Nr. [18] und [19]; Gutachten der Datenethikkommission der deutschen Bundesregierung 2019, Empfehlung Nr. 74).

Auch die Haftung des Herstellers der Software ist alles andere als gesichert. Bei der verschuldensunabhängigen Haftung nach dem **Produkthaftungsgesetz** (PHG) ist umstritten, ob diese überhaupt für reine Software gilt, oder nur für körperliche Gegenstände, in denen Software eingebettet ist. Zudem wird nur für Fehler gehaftet, die dem Produkt im Zeitpunkt des Inverkehrbringens anhafteten, womit Fehler, die erst durch spätere Updates oder das Ausbleiben von Sicherheitsupdates entstehen, nicht erfasst sind. Auch kann sich ein Hersteller der Haftung entziehen, wenn er nachweist, dass der Fehler nach dem Stand von Wissenschaft und Technik nicht erkennbar war, als das Produkt in Verkehr gebracht wurde. Die **Verschuldenshaftung des Herstellers** nach dem ABGB ist gleichfalls mit Schwächen behaftet, zumal der Nachweis einer Sorgfaltswidrigkeit und/oder der Kausalität vielfach nicht gelingt.

Daher wird verstärkt gefordert, eine **Gefährdungshaftung** für den Einsatz derartiger Systeme einzuführen. Kurzfristig wurde – namentlich vom Europäischen Parlament (Entschließung vom 16. Februar 2017 zu zivilrechtlichen Regelungen im Bereich Robotik, Empfehlung Nr. 59 f.) – sogar ins Spiel gebracht, Robotern und „autonomer“ Software selbst Rechtspersönlichkeit zuzugestehen; dieser Gedanke wird allerdings inzwischen der überwiegenden Auffassung nach zu Recht verworfen (s. etwa *Open Letter to the European Commission, Artificial Intelligence And Robotics* 2018; Gutachten der Datenethikkommission der deutschen Bundesregierung 2019, Empfehlung Nr. 73). Nunmehr scheint das Europäische Parlament eher ein zweispuriges Haftungskonzept von (überarbeiteter) Produkthaftung einerseits und einer neu einzuführenden KI-Haftung andererseits zu befürworten (Entwurf vom 27.4.2020 für einen Bericht des Rechtsausschusses, 2020/2014(INL)).

### 3.3.5 Datenschutzrechtliche Verantwortung

Beim Umgang mit Daten von Patientinnen bzw. Patienten unterliegt die Ärztin bzw. der Arzt den Vorgaben der **Datenschutzgrundverordnung** (DSGVO), ergänzt durch das Datenschutzgesetz (DSG). Von Bedeutung für Ärztinnen und Ärzte ist ferner das **Gesundheitstelematikgesetz** 2012 (GTelG 2012). Im Bereich der medizinischen Forschung ergibt sich die Rechtslage zusätzlich aus einer Reihe **weiterer Vorschriften**, insbesondere des Forschungsorganisationsgesetzes (FOG).

Die **elektronische Gesundheitsakte** (ELGA) stellt eine elektronische Vernetzung der ELGA-Gesundheitsdaten von Patientinnen und Patienten her, die verteilt im Gesundheitswesen entstehen. Mit ELGA werden in Österreich stationäre Einrichtungen wie z. B. Spitäler, niedergelassene Ärztinnen und Ärzte sowie Apotheken und Pflegeeinrichtungen flächendeckend vernetzt. Im GTelG 2012 ist klar geregelt, wer auf ELGA-Gesundheitsdaten zugreifen darf: Es sind dies neben der Patientin bzw. dem Patienten selbst ausschließlich jene Ärztinnen und Ärzte und ELGA-Gesundheitsdiensteanbieter, die tatsächlich gerade die betreffende Patientin bzw. den betreffenden Patienten behandeln oder betreuen. Dagegen dürfen etwa Behörden, Betriebsärztinnen bzw. Betriebsärzte

oder jene Ärztinnen und Ärzte, die durch die Patientinnen und Patienten vom Zugriff ausgeschlossen wurden, die Daten nicht einsehen. Die ersten über ELGA verfügbar gemachten Daten sind ärztliche und pflegerische Entlassungsbriefe der öffentlichen Krankenhäuser, Laborbefunde, Radiologiebefunde und Medikationsdaten, d. h. die verordneten Medikamente sind in die e-Medikationsdatenbank einzutragen. Apotheken sind ebenso ab einem bestimmten Zeitpunkt verpflichtet, die Abgabe von verschreibungspflichtigen sowie von wechselwirkungsrelevanten Medikamenten einzutragen. Aus dem GTelG 2012 ergeben sich eine Fülle von weiteren Anforderungen an die Sammlung, Speicherung und Übermittlung von Daten, welche Ärztinnen und Ärzte zu beachten haben.

Bei den meisten Daten, die von Ärztinnen und Ärzten erhoben und verarbeitet werden, handelt es sich um **Gesundheitsdaten** und damit um „besondere Kategorien von Daten“ im Sinne von Art 9 DSGVO, deren Verarbeitung verschärften Anforderungen unterliegt. Jede Verarbeitung solcher Daten – und sei es nur deren Einsichtnahme am Bildschirm – bedarf einer **Rechtsgrundlage**. Dabei kommt etwa die ausdrückliche Einwilligung der bzw. des Betroffenen, die Notwendigkeit für die Erfüllung des Behandlungsvertrags oder die Notwendigkeit zur Erfüllung einer gesetzlichen Verpflichtung (wie beispielsweise nach dem GTelG 2012 oder dem ÄrzteG) in Betracht. In der Praxis erfolgt die Verarbeitung der Daten zumeist zur Erfüllung des Behandlungsvertrags oder zur Erfüllung der gesetzlich vorgeschriebenen Dokumentationspflicht, weshalb eine ausdrückliche Einwilligung der Patientinnen und Patienten meist gar nicht erforderlich ist. Dennoch bedarf es ausführlicher **Information** der Patientinnen und Patienten, u. a. über die durchgeführten Verarbeitungen und deren Zwecke sowie über die **Rechte**, die Patientinnen und Patienten bezüglich ihrer personenbezogenen Daten zustehen. Dazu kann z. B. auch ein Lösungsanspruch gehören.

Verantwortlich für die gesetzeskonforme Datenverarbeitung ist in einer Ordination immer die/der jeweilige (selbständig) behandelnde Ärztin bzw. Arzt, bei einer Krankenanstalt grundsätzlich deren Träger und die für ihn handelnden Organe. Die Verletzung datenschutzrechtlicher Vorschriften kann zu einer verwaltungsstrafrechtlichen und zivilrechtlichen Haftung der verantwortlichen Person führen. Nach der DSGVO drohen bekanntermaßen hohe Geldstrafen.

Das Bestreben nach bestmöglicher medizinischer Qualität dank Einsatz moderner digitaler Technologien steht vielfach in einem **Spannungsverhältnis** zum Bestreben nach Datenschutz. Insbesondere jene Technologien, deren Einsatz weit in den häuslichen Bereich hineinragt (z. B. Datenerhebung zu medizinischen Zwecken mithilfe von Smart Watches, Smart Phones oder Smart Home-Geräten) bringen vielfach ein enormes Eingriffspotenzial mit sich. Während es viele Fälle gibt, in denen der Schutz von Leben und Gesundheit aufgrund einer Abwägung Vorrang genießt gegenüber der digitalen Selbstbestimmung, kann von einem absoluten Vorrang nicht gesprochen werden. Ein hinreichender Datenschutz ist nicht nur grundrechtlich geboten, sondern auch eine wichtige Maßnahme zur Herstellung von Vertrauen und Akzeptanz der Patientinnen und Patienten im Zusammenhang mit der Datenverarbeitung in der Medizin und damit ein wichtiger Hebel für die weitere Digitalisierung (in) der Medizin.

### 3.4 Ökonomische Triebkräfte und Argumente

Digitale Technologien im Gesundheitswesen werden auch aus wirtschaftlicher Perspektive als systemverändernd angesehen. Sie beeinflussen die Gestaltung der Versorgungsprozesse, Risikoprofilbildung, Prävention, Diagnostik, Therapieansätze, *Public Health*, Pflege und Betreuung sowie Forschung (Jannes et al. 2018, 15ff.).

Das ökonomische Bestreben geht dahin, das Gesundheitswesen durch den Einsatz digitaler Technologien effektiver und effizienter zu gestalten, insgesamt also die Kosten-Effektivität der Gesundheitsversorgung zu erhöhen (Müller-Mielitz 2017), und daraus eine entsprechende Rendite zu erlösen:

- Eine potentielle Steigerung der Effektivität wird darin gesehen, dass digitale Technologien dabei helfen, Krankheitsursachen, -zusammenhänge und wirksame Behandlungsansätze besser zu verstehen. Dadurch sollen Fehlversorgung minimiert und Kosten reduziert werden.
- Eine Effizienzsteigerung durch digitale Technologien wird dadurch erhofft, dass diese Entwicklungen mit einem günstigeren Input-Output-Verhältnis erfolgen (insbesondere durch eine Aufwandsreduktion durch besseres Zuschneiden von Leistungen auf die Bedürfnisse individueller Patientinnen und Patienten).
- Die Rendite, die potentielle Effektivitäts- und Effizienzsteigerungen bergen, werden von verschiedenen Stakeholdern in unterschiedlichem Ausmaß erlöst.

Diese Aussichten führen zu einer massiven Investitionsbereitschaft (*Forbes Insights Team* 2019; Taylor 2015; Hipp et al. 2018) in digitale Infrastrukturen und andere digitale Technologien. So investieren beispielsweise pharmazeutische Unternehmen in zunehmendem Maß in Apps für die Aufklärung und Schulung im Umgang mit Medikamenten (Taylor 2015, 10). Es treten aber auch neue Akteure im Gesundheitswesen auf: Technologieunternehmen wie Google, Apple und Microsoft sind dabei, zu maßgeblichen Gestaltern im Gesundheitswesen zu werden. Mit ihnen verbunden sind zahlreiche Start-ups, App-Entwicklerinnen bzw. Entwickler und kleinere Technikfirmen, die im Gesundheitsmarkt mit teilweise disruptiven Innovationen auftreten (Kaltenbach et al. 2016; Hipp et al. 2018).

Für das deutsche Gesundheitswesen wird das Steigerungspotenzial der Kosten-Effektivität auf insgesamt 12% des Gesamtaufwands eingeschätzt, wenn es zu einer konsequenten Ausnützung bereits vorhandener digitaler Technologien käme (Hegner/Biesdorf et al. 2018), davon 47% in der stationären Versorgung (Hegner/Liese et al. 2018). Entsprechende Daten zur österreichischen Situation liegen derzeit nicht vor. Folgende Technologien werden für diese potentielle Verbesserung der Kosten-Effektivität als relevant bewertet (gereiht nach ihrem Impact; Hegner/Biesdorf et al. 2018; Hegner/Liese et al. 2018):

- Elektronische Datenübermittlung statt Papier: einheitliche elektronische Patientenakte, elektronisches Rezept, Krankenhaus-interne Kommunikation der Mitarbeiterinnen und Mitarbeiter, elektronische Überweisungen;
- Onlineinteraktion: Teleberatung, Fernüberwachung chronisch kranker Patientinnen und Patienten, elektronische (Algorithmen-basierte) Triage an der Schnittstelle von intra- und extramuraler Versorgung;
- Arbeitsabläufe: digitale Vernetzung des Pflegepersonals (Zugang zu Patienteninformationen über intra-/extramurale Grenzen hinweg, z. B. in der häuslichen Pflege), Barcode-basierte Verabreichung von Medikamenten, RFID-Tracking, Überwachung von Vitalparametern, Roboter für Krankenhauslogistik, Automatisierung einfacher Prozessschritte, elektronische Überweisungen;
- Entscheidungsunterstützung: Leistungs-Dashboards (d. h. digitale Informationssysteme, die intern Auskunft über die Versorgungsschritte und den Outcome durch Behandlungsteams geben und helfen, Verbesserungsmöglichkeiten zu identifizieren), Steuerung der Patientenströme, elektronische (Algorithmen-basierte) Triage, klinische Entscheidungsunterstützung (d. h. bei Therapieentscheidungen), Entscheidungsunterstützung im Rahmen von genetischen Tests;
- Patientenselbstbehandlung: Tools für das Management chronischer Erkrankungen, medizinische Chatbots, Tools zur Vorbeugung von Erkrankungen, patientenunterstützte Netzwerke, digitale Diagnosetools, Virtual Reality Applikationen (z. B. für Schmerzbehandlung);
- *Patient-Self-Service*: elektronische Terminvereinbarung.

Der Großteil des Kosten-Effektivitäts-Steigerungspotenzials wird derzeit im Prozessmanagement klinischer, administrativer und logistischer Bereiche gesehen (*Forbes Insights Team* 2019). Durch den Einsatz digitaler Technologien wird in diesen Bereichen eine Verringerung der Transaktionskosten angestrebt. Die unmittelbare Gesundheitsversorgung (Prävention, Diagnostik, Therapie, Rehabilitation) spielt im Vergleich dazu noch eine untergeordnete Rolle. Es wird jedoch damit gerechnet, dass sich dies in Zukunft ändert (Price 2019; Hipp et al. 2017, 8 ff.; Hipp/Schlude et al. 2017, 17 ff.; Prainsack 2019, 11–16).

Die ökonomisch motivierten Triebkräfte, die hinter der Nutzung von *Big Data* und maschinellem Lernen für Diagnose und Therapieplanung stehen, bringen einige der ethischen Überlegungen nochmals „auf den Punkt“, vor allem was die sich öffnenden Spannungsfelder betrifft:

- Konzentration auf das klinische Kerngeschäft: Verringerung der Transaktionskosten durch digitale Technologien setzt Ressourcen für das medizinische Kerngeschäft frei.
- Chancengerechtigkeit beim Zugang: Verstärkung oder Verringerung von gesundheitlicher Ungleichheit durch digitale Technologien („digitale Kluft“; Fischer 2017, 147).



- Effizienzsteigerung der strukturellen Gesundheitsversorgung: Unterstützung der Versorgungsforschung durch Anwendung von *Big Data*-Analysen (Jannes et al. 2018, 26).
- Effizienz auf Kosten von Effektivität: mangelnde Differenzierung zwischen Korrelation und Kausalität; Förderung von Automatismen (Jannes et al. 2018, 25).
- Verantwortung beim und zum Einsatz: Pflicht zur Auseinandersetzung mit digitalen Technologien durch Gesundheitsdienstleister im Rahmen ihres Sorgfaltsmaßstabs (Gründinger et al. 2019, 12); erhöhte Transaktionskosten und dadurch Verzögerung bei der Entwicklung und Anwendung digitaler Technologien durch mangelhafte Zuschreibung und Verteilung von Verantwortung (Jannes et al. 2018, 29).
- Risikostratifizierung in der Krankenversicherung: Auswirkungen von Datenverarbeitung hinsichtlich von Vorerkrankungen und Therapie-Adhärenz (Gründinger et al. 2019, 17).
- Wettbewerb der Algorithmen: Tendenz zur Monopolbildung bei datenbasierten Diensten; Gestaltung von Daten-Sharing-Pflichten (Gründinger et al. 2019, 41).

## 3.5 Big Data: Quellen und Qualität

### 3.5.1 Was bedeutet „Big Data“?

Kritikerinnen und Kritiker wenden ein, dass der derzeitige Fokus auf „*Big Data*“ fälschlicherweise suggeriert, dass es sich hierbei um völlig neue Zugänge handle. Schon die Verwendung des Begriffs „*Big Data*“ verstellt demnach den Blick darauf, dass ärztliches Handeln schon seit vielen Jahrzehnten – und spätestens seit der Einführung der evidenzbasierten Medizin – auf der systematischen Auswertung von Daten aufbaut, entweder direkt (indem Ärztinnen und Ärzte Hilfsmittel wie systematische *Reviews* zu Rate ziehen) oder indirekt (über klinische Richtlinien). Diese Kritik ist ernst zu nehmen. Gleichzeitig muss festgehalten werden, dass die rapide Weiterentwicklung diagnostischer und therapeutischer Technologien und Instrumente insbesondere seit der Jahrtausendwende, und die „digitale Revolution“ im Allgemeinen etablierte Praktiken, Standards und Möglichkeiten stark verändert haben.

Eine allgemeingültige Definition von *Big Data* gibt es nicht. Folgende Definitionen aus unterschiedlichen Disziplinen und Kontexten gehören zu den etabliertesten:

- *Big Data* als **inkrementales Phänomen**: *Big Data*-Datensätze sind größer als jene, für die die bisherigen Datenbanken und Softwaresysteme gebaut wurden. Sie sind größer hinsichtlich ihres Volumens, ihrer Velozität (d. h. der Geschwindigkeit, mit der sie generiert werden und sich durch Systeme bewegen) und ihrer Variabilität (e. g. Laney 2011).
- *Big Data* als **computationales Problem**: Eng verbunden mit der vorangegangenen Definition bezieht sich diese Definition auf Datensätze, die so groß sind, dass sie die bisherigen Kapazitäten zur Erfassung, Speicherung und Analyse über-

steigen (McKinsey 2011). Mit „*Big Data*“ sind nach dieser Definition alle Datensätze gemeint, deren Speicherung und Verarbeitung mit derzeitigen Mitteln nicht bewerkstelligt werden kann.

- *Big Data* als **technosoziales Phänomen**: Nach dieser Definition bezeichnet der Begriff „*Big Data*“ nicht bloß Datensätze und die dazugehörigen Technologien, sondern die datenmäßige Erfassung immer weiterer Teile unserer Körper und unseres persönlichen und sozialen Lebens („Datafizierung“).
- *Big Data* als **methodologischer Zugang**: nach dieser Definition sind „*Big Data*“-Ansätze jene, die „hypothesenfrei“ in großen Datensätzen nach Korrelationen suchen (z. B. „welche Korrelationen zwischen genetischen Markern und einem Phänotyp lassen sich in den Daten von 300.000 Patientinnen und Patienten finden?“ (Antes 2016)). Damit setzen sie sich von herkömmlichen Zugängen, die mit einer bestehenden Arbeitshypothese oder einer konkreten Forschungsfrage an die Datenanalyse herangingen, ab. Wenn Korrelationen gefunden sind, kann weitergeforscht werden, ob es sich um zufällige Korrelationen handelt oder ob diese durch einen Kausalzusammenhang mit einander verbunden sind; d. h. die Arbeitshypothese wird nach der Datenanalyse, nicht vorher, generiert.
- *Big Data* bezeichnet **alle Daten, die ein Phänomen ausmachen**. Dies ist die „radikalste“ Definition von *Big Data* und jene, die sich vergleichsweise am wenigsten durchgesetzt hat. Von „*Big Data*“ würde man nach dieser Definition nur dann sprechen, wenn man z. B. den ganzen Körper einer Person durch bildgebende Verfahren erfasst hat und nicht nur das Organ, in dem sich ein zu behandelnder Tumor befindet. Ebenso würden Anhängerinnen und Anhänger dieser Definition nur dann von *Big Data* sprechen, wenn das gesamte Genom einer Person sequenziert wurde und nicht nur das Gen in dem eine bestimmte Mutation vermutet wird.
- *Big Data* als **Vermögenswert**: Das multinationale IT-Unternehmen Gartner definiert *Big Data* als „*high volume, high velocity, and /or high variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation*“ (Gartner 2019). Durch die Charakterisierung von *Big Data* als Vermögenswert behandelt die Firma Daten als etwas, das Eigentum darstellen und veräußert werden kann (dazu kritisch etwa Birch 2017).
- Für den **österreichischen Kontext** besonders relevant ist die Definition des Forschungsorganisationsgesetzes (FOG): *Big Data* sind demnach „die Verarbeitung großer Mengen von wenig oder nicht strukturierten Daten“ (§ 2b Z 3 FOG). In den Erläuterungen wurde klargestellt, dass zur Begriffsfindung die Definition des National Institute of Standards and Technology, U.S. Department of Commerce (NIST) vom September 2015 (in den Erläuterungen unrichtig als September 2017 bezeichnet) herangezogen wurde. Mittlerweile besteht eine revidierte Fassung der NIST-Definition vom Juni 2018.

Im Einklang mit Wissenschaftlerinnen und Wissenschaftern und wie Gernot Rieder und Judith Simon (2016) hält die österreichische Bioethikkommission ein Verständnis von *Big Data*, das sich auf die bloß technische Dimension datenintensiver Praktiken in der Wissenschaft und Medizin bezieht, für wenig fruchtbar. In den Worten Rieders und Simons:

„[E]s scheint viel zielführender, Big Data als komplexes sozio-technisches Phänomen zu sehen das auf einem Zusammenspiel aus technischen, wissenschaftlichen, und kulturellen Faktoren beruht. Während die technologische Dimension [von Big Data] auf Fortschritte bei Hardware, Software, sowie bei der Infrastruktur und die wissenschaftliche Dimension auf neue Formen der Datenanalyse verweist, so bezieht sich die kulturelle Dimension (a) auf die allgegenwärtige Verwendung von Informations- und Kommunikationstechnologien in unserer Gesellschaft und (b) auf die wachsende Bedeutung und Autorität quantifizierter Information in unterschiedlichen Lebensbereichen“ (Rieder/Simon 2016, 2; Übersetzung ins Deutsche durch die Bioethikkommission).

Wir verstehen *Big Data* also als soziotechnische Praxis, die immer in Verbindung mit den politischen, ökonomischen, und gesellschaftlichen Faktoren gesehen werden muss, die diese Praxis ermöglichen.

### **3.5.2 Woher kommen die Daten für Big Data-Analysen in der Medizin?**

Aufgrund der fortschreitenden „Datafizierung“ unserer Gesellschaft stehen der Forschung heute Daten über verschiedenste Aspekte unseres Lebens zur Verfügung, die in der Vergangenheit nicht datenmäßig erfasst wurden. Früher mussten Forscherinnen und Forscher, die wissen wollten, wie sich eine Person ernährt, wieviel sie sich täglich bewegt, oder wie sich die Einnahme eines neuen Medikaments auf ihre Darmflora auswirkt, darauf beschränken, die Menschen direkt zu befragen. Heute können diese Daten voll- oder halbautomatisiert über Smartphone-Apps oder andere tragbare Sensoren erfasst werden. Dies bedeutet, dass Information nicht mehr durch das Gespräch mit der Patientin bzw. dem Patienten, sondern über unmittelbare Interrogation ihres Körpers gewonnen wird. Gleichzeitig hat die Forschung heute Zugang zu Datensätzen über die Gesundheitsversorgung großer Teile der Bevölkerung (Hoeyer 2016; Kautzky-Willer et al. 2017). Dies bedeutet, dass Daten, die heute für medizinisch relevante Fragestellungen ausgewertet werden können, nicht mehr – wie früher – fast ausschließlich aus der Klinik oder aus der medizinischen Forschung kommen. Heute kommen diese Daten aus vielen verschiedenen Quellen, wie auch aus Smartphones (Apps, Bewegungsartung, etc.), aus öffentlichen Archiven, von Wetter- und Umweltsensoren, von privaten Unternehmen, die ihre Daten entweder für Geld oder als philanthropische „Spende“ der medizinischen Forschung zur Verfügung stellen, etc. (Vayena et al. 2018). Aufgrund der heute zur Verfügung stehenden digitalen und computationalen Hilfsmittel können Daten aus unterschiedlichen Quellen zusammengeführt und so ausgewertet werden, dass sich generelle Trends auf der Bevölkerungsebene ableiten lassen (z. B. dass die Einnahme von lipidreduzierenden

Medikamenten das Krebsrisiko bei Diabetespatientinnen und Diabetespatienten senkt (Kautzky-Willer et al. 2017). Darüber hinaus lassen sich auch probabilistische Aussagen über Einzelpersonen tätigen (z. B. ob eine Person ein höheres oder niedrigeres Risiko hat, an einer chronischen Erkrankung zu leiden, als die Durchschnittsbevölkerung).

Erwähnenswert ist in diesem Kontext auch das Verschwimmen der Trennlinie zwischen medizinischer Forschung und medizinischer Praxis in der Nutzung von Daten. Daten, die für klinische Zwecke gesammelt wurden, wie z. B. Informationen über Diagnose, verschriebene Medikamente, Labortests, und andere Informationen die in Patientenakten festgehalten werden, können heute systematisch nach versteckten Mustern untersucht werden – wie z. B. nach bisher unbekanntem Zusammenhängen zwischen Patientencharakteristika und prognostischen Parametern (z. B. Patientinnen und Patienten, die mit einem bestimmten Medikament behandelt werden genesen nach einer Operation besonders schnell; oder Patientinnen und Patienten mit einer bestimmten Komorbidität haben nach einem bestimmten Eingriff besonders häufig Komplikationen). Diese Einsichten, die aus dem hypothesenfreien Durchsuchen großer Datensätze gewonnen werden, können zur Behandlung individueller Patientinnen und Patienten – also in der klinischen Praxis – hinzugezogen werden und gleichzeitig auch die medizinische Forschung bereichern (Wooden et al. 2017). Der Respekt vor dem Recht informationaler Selbstbestimmung individueller Patientinnen und Patienten, von denen diese Daten stammen, sowie andere datenschutzrechtliche Aspekte sind vor dem Hintergrund solcher neuen Möglichkeiten besonders wichtig.

### 3.5.3 Quantität und Qualität

Eine andere bedeutsame Erwägung in diesem Kontext ist die Qualität der erhobenen Daten. In einer Zeit, in der alle für die medizinische Forschung und Praxis verwendeten Daten von Klinikerinnen bzw. Klinikern oder professionellen Forscherinnen bzw. Forschern gesammelt wurden, wusste man in den meisten Fällen, wie gut oder schlecht die Qualität der Daten war, die man verwendete. In einer Situation, in der Daten aus vielen verschiedenen Quellen kommen, wird es zunehmend schwieriger, die Qualität der Datensätze zu beurteilen. Oft ist eine solche Beurteilung praktisch und technisch unmöglich, weil die Firma, die den Sensor, die App, oder den Apparat zur Datenerhebung zur Verfügung stellt, Informationen über die Datensammlung als Betriebsgeheimnis behandelt. Eine weitere Herausforderung im Kontext der Datenqualität ist die Frage der Repräsentativität und damit der Vergleichswerte. In der Vergangenheit waren z. B. in den Datensätzen einer Gebietskrankenkasse weite Teile der Bevölkerung vertreten – und damit auch in den Forschungsvorhaben die diese Datensätze verwendeten. Wenn heute Telekommunikationsunternehmen oder Fitness-App-Betreiber Daten über die Bewegungsströme ihrer Kundinnen und Kunden der Forschung zur Verfügung stellen (z. B. um die Verbreitung eines Virus zu rekonstruieren), dann „fehlen“ in diesen Datensätzen all jene Menschen, die kein Mobiltelefon-Abo haben oder keine Fitness-Apps benutzen. Da es sich hierbei oft um marginalisierte Personengruppen handelt, laufen *Big Data*-Ansätze in der medizinischen Forschung Gefahr, auf die Gruppe der „*wealthy and healthy*“ – also jene

Menschen, die wohlhabend und vergleichsweise gesund sind – zu fokussieren und jene Menschen zu übersehen, die Unterstützung am meisten brauchen würden.

Ein zusätzliches Problem für die Interpretation von digitalen Daten ist das Vorhandensein guter Meta-Daten, also von Information, die der Datenanalystin oder dem Datenanalysten sagt, wen oder was diese Daten repräsentieren, welches die Stärken und Schwächen der Datensätze sind, etc. Die Aussage, dass Daten für sich selbst sprechen, ist im Kontext digitaler *Big Data* nicht zutreffend (Leonelli 2016). Um Daten sinnvoll nutzen zu können braucht es gute und verständliche Information über den Kontext der Datensammlung, z. B. wie genau wurden die Personen ausgewählt, deren Daten in der Datenbank gespeichert wurden? Wie wurden wesentliche Kategorien wie z. B. diagnostische, therapeutische, demographische, etc. Parameter definiert? Wurden diese Kategorien nachprüfbar standardisiert angewandt? etc. Wenn solche kontextuelle Information nicht zur Verfügung steht, besteht das Risiko, dass die Daten nicht sinnvoll interpretiert werden können (oder falsche Interpretationen suggerieren).

Allgemein lässt sich zur Frage der Datenqualität sagen, dass Kriterien zur Messung der Qualität der Daten nicht abstrakt formuliert werden können, sondern immer praxis-, kontext- und zielspezifisch festgeschrieben werden müssen. Wenn Daten von Mobilfunkbetreibern zur Forschung in der digitalen Epidemiologie herangezogen werden, sind andere Kriterien von Bedeutung als in einem Kontext, in dem auf der Basis der Analyse elektronischer Gesundheitsakte herausgefunden werden soll, welche Menschen von einer spezifischen medizinischen Intervention besonders profitieren würden.

### **3.5.4 Datenqualität und Künstliche Intelligenz (KI)**

In der medizinischen Forschung ist bereits lange bekannt, dass eine medizinische Applikation zur Muster-/Regelerkennung auf Grundlage des Datenmaterials falsche Schlüsse ziehen kann. So wurde im Jahr 1997 eine Studie durchgeführt, um Methoden des maschinellen Lernens für den medizinischen Einsatz nutzbar zu machen. Ziel der Studie war, eine Überlebensprognose für Patientinnen und Patienten mit Pneumonie zu errechnen, sodass Patientinnen und Patienten mit einem errechneten hohen Risiko stationär, Patientinnen und Patienten mit einem geringen Risiko hingegen ambulant behandelt werden (Caruna et al. 2015; Cooper et al. 1997). Auf Basis des Trainingsdatensatzes entwickelte das regelbasierte System die (unrichtige!) Regel, dass Patientinnen und Patienten mit Asthmavergangenheit nur ein geringes Risiko aufweisen und demnach ambulant behandelt werden können. Dies konnte damit erklärt werden, dass Patientinnen und Patienten mit Pneumonie, die zusätzlich unter Asthma litten, nach deren Aufnahme unverzüglich in die Notfallambulanz eingewiesen wurden, wo sie intensiv betreut wurden; in Folge besserte sich ihre Überlebensprognose gegenüber Patientinnen und Patienten (ohne Asthma), denen diese intensive Pflege nicht zukam. Zwar erkannte die Künstliche Intelligenz sohin richtig, dass Asthmapatientinnen und Asthmapatienten (im Ergebnis) ein geringeres Risiko aufwiesen, konnte dies aber nicht in den Kontext zur intensiveren medizinischen Betreuung setzen. Ein wesentliches Problem im Bereich der Datenqualität liegt folglich darin, dass Daten, genauso wie die eingesetzten Algorithmen „voreinge-

nommen“ oder „*biased*“ sein können, was auf unterschiedliche Ursachen zurückzuführen ist (z. B. Erhebung und anschließende Auswahl der Daten, initiales Design und autonome Weiterentwicklung des Algorithmus, fehlerhafte Bedienung etc., vgl. auch 3.7.3).

Das Problem der bzw. die Forderung nach Datenqualität im Bereich der medizinischen Forschung ist folglich nicht nur auf *Big Data* begrenzt, sondern betrifft sämtliche (auch unverknüpfte) Datensammlungen (Patientenakten, Daten aus medizinischen Studien etc.). Damit zeigt sich aber wiederum die enge Verzahnung mit den eingesetzten technischen Möglichkeiten und der damit verbundenen Nachvollziehbarkeit des Entscheidungsprozesses (Black-Box-Problem).

Was die Erhebung von Daten an der Patientin bzw. am Patienten mittels Sensoren betrifft, gibt es bereits Regelwerke, die die Qualität der eingesetzten Sensoren sicherstellen (z. B. MPG, PHG, PSG). So werden im MPG (bzw. in der MPVO als Nachfolgebestimmung) Herstellern strikte Vorgaben auferlegt, um die Sicherheit und Zuverlässigkeit von Medizinprodukten zu gewährleisten. Damit ein Medizinprodukt zugelassen wird, muss dessen Wirksamkeit und Sicherheit im Rahmen einer klinischen Prüfung vorab untersucht werden. Als Medizinprodukt wird kraft gesetzlicher Vorgabe auch Software verstanden. Diese öffentlich-rechtliche Schutznorm wird durch zivilrechtliche Haftungsbestimmungen, wie etwa die des PHG, flankiert, um Patientinnen und Patienten vor fehlerhaften Produkten, und damit eine Gefährdung ihrer Gesundheit, zu bewahren. Im Bereich der Produkthaftung ergeben sich in diesem Zusammenhang schwierige Abgrenzungsprobleme, wann ein Produkt dem Regime des PHG zuordenbar ist. Das subsidiär anzuwendende PSG ergänzt den qualitativen Schutz von Sensoren.

Zuzustimmen ist, dass weiterführende Auswertungen dieser Daten durch private Unternehmen (z. B. bei Fitness-Tracker) nur dann für Zwecke der medizinischen Forschung bzw. im medizinischen Bereich verwendet bzw. verknüpft werden sollen, wenn die Entscheidungsfindung nachvollziehbar ist, das von einem privaten Unternehmen eingesetzte KI-System daher „*explainable*“ ist, was bei datenbasierten Systemen unter Einsatz neuronaler Netze derzeit nur beschränkt möglich und Gegenstand intensiver Forschung ist. Jedoch können z. T. auch mit den isolierten Rohdaten der Sensoren beeindruckende Ergebnisse erzielt werden – so ist die Anwendung *DeepHeart* nur unter Zuhilfenahme neuronaler Netze und den Sensordaten von Smart Watches (z. B. *Apple Watch*) fähig, bestimmte kardiovaskuläre Risiken (Bluthochdruck, Diabetes, Schlafapnoe, hoher Cholesterinspiegel) mit hoher Sensitivität [98%] und Spezifität [90%] aufzuzeigen.

## 3.6 Zentrale Begriffe: Assoziation, Korrelation und Kausalität

Assoziation, Korrelation und Kausalität sind Begriffe, die je nach Kontext mehrere Bedeutungen haben. Im alltäglichen Sprachgebrauch meint Assoziation eine gedankliche Verbindung, Korrelation die mathematisch beschreibbare Wechselbeziehung zwischen zwei oder mehreren Ereignissen und Kausalität bezieht sich auf den Grund für ein Ereignis, also auf Erkennen einer Ursache und ihrer Wirkung. Ganz generell spielt Kausalität, auch wenn sie schwer fassbar ist, gerade im Medizin- und Gesundheitsbereich eine besondere Rolle. Bereits die einfache Frage, ob es die Tablette war, die den Kopfschmerz aufgelöst hat oder ob dieser auch ohne Tablette aufgehört hätte, verdeutlicht, dass Kausalität komplizierter ist, als gemeinhin angenommen, da sich nicht ganz einfach klären lässt, ob das Medikament, die verstrichene Zeit, beides oder etwas ganz Anderes dafür sorgte, dass der Kopfschmerz verging.

### 3.6.1 Begriffliche Neubestimmungen

Mit der Entwicklung datenintensiver Biomedizin, die letztlich Basis für die heute in vielen klinischen Kontexten wichtige *genetic/genomic medicine* ist (Horton/Lucassen 2019), erfuhren auch die Begriffe Assoziation, Korrelation und Kausalität einen Bedeutungszuwachs, der zum Teil den wissenschaftlichen Erkenntnisgewinn in Bezug auf Kausalität in komplexen lebenden Systemen und die Entwicklung von Methodik und Technologie in Statistik, Epidemiologie und Computerwissenschaften widerspiegelt, die auch die Grundlage für maschinelles Lernen darstellt (s. 3.7). Dies führt auch dazu, dass wissenschaftliche und philosophische Konzepte bezüglich dieser Begriffswelten zum Teil neu vermessen werden, sodass gerade Kausalität ein sehr facettenreicher Begriff ist. Geht es beispielsweise darum, die Wirksamkeit einer Behandlung oder unerwünschte Nebenwirkungen zu erkennen, so kann eine randomisierte klinische Studie hier zwar Aufschluss geben, die Ergebnisse erlauben jedoch oft nicht, direkte Kausalzusammenhänge zu belegen, weswegen deren Auffinden eine wesentliche Herausforderung für die biomedizinische Informatik darstellt (Kleinberg/Hripcsak 2011).

Kausalzusammenhänge haben in der Medizin drei wesentliche Anwendungsgebiete, nämlich Prädiktion, Erklärung (Diagnostik und dadurch begründbare Therapie) sowie Politik (im Sinn von Behandlungs-Entscheidungen, aber auch von gesundheitspolitischen Maßnahmen). Dabei betreffen Erklärungen vor allem die Beziehung von zwei Phänomenen und warum sie miteinander assoziiert sind, sowie spezielle Ereignisse (warum sie überhaupt eintraten oder warum sie so eintraten, wie es geschehen ist; Kleinberg/Hripcsak 2011). Da lernende Algorithmen Kausalitätsmodelle verwenden müssen, um entscheiden zu können (s. 3.7), kommt diesen zugrundeliegenden Modellen somit eine immense Bedeutung zu.

Auch der Begriff der Komplexität, den wir zumeist als eine nicht näher definierte Umschreibung von „schwierig“ verwenden, bedarf eines vertieften Verständnisses von biologischen Systemen, die sich, anders als mechanische Systeme, in Lernprozessen

weiterentwickeln, wodurch Komplexität im Zusammenhang von Medizin und Gesundheit mehr als nur „schwierig“ ist (Sturmberg 2018). Interessant ist in diesem Kontext die konvergente Entwicklung verschiedener Wissenschaftsbereiche, die letztlich zu einer neuen Sichtweise auf Lebensprozesse führen, die im idealen Fall frei von Wertungen und Stereotypen ist und insofern „objektiver“ als bisher wird. Dabei ergeben sich jedoch zahlreiche Unwägbarkeiten und ethische Fragen, die gerade auch die biomedizinische Grundlagenforschung betreffen, die selbst zunehmend auf die Methoden der Bioinformatik und Datenwissenschaft angewiesen sind (Überall/Werner-Felmayer 2019).

### 3.6.2 Das Beispiel der genetischen Assoziationsanalyse

Folgendes Beispiel illustriert die Bedeutung von Assoziation, Korrelation und Kausalität bei der Interpretation genetischer/genomischer Daten für die Ausprägung komplexer Merkmale (Phänotypen), wie etwa Körpergröße oder beispielsweise neurodegenerativer Erkrankungen (Alzheimer, Parkinson), Diabetes (Typ 2), Erkrankungen des Herz-Kreislauf-Systems und Krebs. Hier liefern genetische Assoziationsanalysen wichtige Einsichten. Bei komplexen Merkmalen ist das Konzept von „*one-gene, one-mutation, one-outcome*“ (Gallagher/Chen-Plotkin 2018, 717) Modellen zum Verständnis der Krankheit, wie wir es von „klassischen“ genetischen Erkrankungen, etwa Chorea Huntington kennen, nicht ausreichend, da in der Regel zahlreiche Gene und nicht-genetische Faktoren an ihrer Ausprägung beteiligt sind und ihre Vererbungsmuster nicht den Mendel'schen Regeln folgen (Gallagher/Chen-Plotkin 2018). Die Assoziationsanalyse war zunächst eine Möglichkeit des *data mining* in großen, relativ hochwertigen, weil strukturierten genomischen Datensätzen. Sie sucht Muster, die eine Beziehung von gemeinsam auftretenden genetischen Variablen statistisch nahelegen. Im medizinischen Kontext werden Assoziationsanalysen z.B. mit Genexpressions-Daten oder mit Daten zu genetischer Variation wie *single-nucleotide polymorphisms* (SNPs) durchgeführt, können aber auch zur Prädiktion von biologisch relevanten Interaktionsnetzwerken beispielsweise von Proteinen herangezogen werden (Atluri et al. 2009). Genomweite Assoziationsstudien (*genome wide association studies, GWAS*) förderten während der letzten Jahre tausende Genvarianten zutage, die mit einer bestimmten Wahrscheinlichkeit mit der Ausprägung komplexer Merkmale oder mit Krankheitsrisiken assoziiert sind, d. h. die mit einer bestimmten statistisch relevanten Häufigkeit bei Personen mit bestimmten Erkrankungen gefunden werden.

Da bisher allerdings nur in wenigen Fällen ein kausaler Zusammenhang zwischen dem Auftreten einer Variante und der Ausprägung des Merkmals gefunden werden konnte, bleibt die biologische Bedeutung derartiger Varianten vielfach unklar (Gallagher/Chen-Plotkin 2018). Zumeist liegen die statistisch mit einer bestimmten Erkrankung oder einem komplexen Merkmal assoziierten SNPs außerhalb Protein-codierender Gene, sodass häufig auch unklar ist, welche Gene betroffen sind, welche molekularen Mechanismen beteiligt sind und wie sich Veränderungen der betreffenden Genfunktion oder Genregulation auf das Krankheitsrisiko auswirken (Gallagher/Chen-Plotkin 2018). Hinzu kommt, dass jede mittels GWAS aufgefundene Variante mit hunderten bis tausenden weiteren Varianten assoziiert ist, die ebenfalls statistisch signifikant mit dem Auftreten



eines Merkmals assoziiert sind, ohne dass deshalb etwas über ihre biologische Funktion, also ihre Kausalität ausgesagt werden könnte.

Dieses Phänomen wurde auch als *the missing heritability problem* bezeichnet und zeigt sich beispielsweise in einem vermeintlich einfachen Zusammenhang, nämlich der Ausprägung von Körpergröße. Körpergröße ist ein Merkmal, für dessen Ausprägung – wie wir im Zeitalter von GWAS gelernt haben – eine sehr große Anzahl von Genen mit jeweils relativ kleinen Effekten, aber auch Umwelteinflüsse von Bedeutung sind, wobei trotz Analysen der Genomdaten sehr vieler Menschen (Größenordnung 1 Million) die meisten der aufgefundenen Genvarianten keine statistisch signifikante Korrelation aufweisen, also *unmappable* sind (Barton et al. 2019). Um diese Genvarianten trotzdem zu berücksichtigen, benützt man in neuerer Zeit sogenannte *polygenic risk scores*. Diese Scores sind das Ergebnis von Risikoanalysen, bei denen alle Varianten als Risikofaktoren summiert und nach ihrer Effektstärke gewichtet werden, sodass aus komplexen, nicht im Detail bzw. im Sinn von Kausalität verstandenen Zusammenhängen ein Maß für die Wahrscheinlichkeit bzw. das Risiko der Ausprägung eines bestimmten Merkmals, wie Körpergröße oder Krankheit errechnet werden kann. GWAS-Studien sind zudem immer auch Populationsstudien, was bedeutet, dass neben Vererbung und unterschiedlichen Umweltfaktoren auch ein *confounding factor*, also eine sogenannte Drittvariable oder Störvariable eine Rolle bei der Prädiktion von *risk scores* spielt, wie z. B. die Struktur der Population. Dadurch sind *polygenic risk scores*, die auf der Auswertung vieler kleiner Effekte beruhen, jedoch fehleranfällig, wie dies in Zusammenhang mit GWAS-Studien zu Körpergröße deutlich wird (Barton et al. 2019). Durch die zusätzliche Annahme von *polygenic adaptation*, also dem Konzept, dass Anpassung im Lauf der Evolution durch natürliche Selektion auf von vielen Genen beeinflusste Merkmale entsteht, fördern *polygenic risk scores* zudem eine Überbewertung von Unterschieden einzelner Populationen bezüglich ihrer Krankheitsrisiken und anderer komplexer Merkmale (Rosenberg et al. 2019; s. 3.7.3 für Näheres).

Durch *polygenic risk scores* ergibt sich eine Verschiebung des ursprünglichen Ziels genomischer Analysen, Gene zu identifizieren, die ursächlich an der Ausprägung eines Merkmals beteiligt sind hin zur Prädiktion eines Phänotyps an Hand von Prädispositionen, die mit bestimmten Merkmalen assoziiert sind. Dies ist ein fundamentaler Unterschied zu bisherigen Konzepten von Genetik und Vererbung, die diagnostisch ausgerichtet waren. Es ist also extrem wichtig, sich dieser Verschiebung von einem diagnostischen zu einem prädiktiven Verständnis von einer Vielzahl genetischer und anderer Ursachen für bestimmte Merkmale bewusst zu sein, um nicht in die Falle eines landläufigen Reduktionismus zu tappen, der dann letztlich zur Verletzung ethischer Standards in Bezug auf die Würde des Einzelnen in seiner jeweils eigenen Ausprägung führt. Hinzu kommt, dass Prädiktion impliziert, dass Ereignisse voraussagbar sind. Im Kontext genetischer Varianten ist jedoch häufig unklar, wie stark sie die Ausprägung eines Merkmals beeinflussen. *Polygenic risk scores* bieten zwar eine Zahl, die statistische Zusammenhänge mit einem möglichen Risiko herstellen, stellen jedoch keine Vereinfachung komplexer funktioneller Zusammenhänge zwischen Genen, Umwelt und Evolution dar. Trotzdem

fördern sie, wie die aktuelle wissenschaftliche Literatur zum Thema zeigt, offenbar das Missverständnis eines ursächlichen Zusammenhangs von genetischen Variationen und komplexen Merkmalen (Rosenberg et al. 2019). (S. auch 3.7.3). Im klinischen Kontext können *polygenic risk scores* allerdings zunehmend nutzbringend eingesetzt werden, vor allem im Bereich von Früherkennung und Prävention von häufigen Erkrankungen des Erwachsenenalters, wie beispielsweise koronaren Herzerkrankungen (Torkamani et al. 2018). Dabei wird deutlich, dass sich aus den Assoziationsstudien zur Mustererkennung in genomischen Daten innerhalb von knapp 20 Jahren Prädiktionsstudien mit hohem Relevanzpotential entwickelt haben, die ein vertieftes und mathematisch fassbares Verständnis von Risiko und Kausalität benötigen und in genau definierten Kontexten durchgeführt werden müssen.

### 3.7 Maschinelles Lernen – Algorithmen – technische Grundlagen

Um aus *Big Data* „Sinn“ zu generieren, sind mathematische Modelle und statistische Verfahren zur Mustererkennung, Klassifizierung, Prädiktion und, als wichtigstes Merkmal, von Kausalzusammenhängen (Kausalschlüsse, *causal inference*) notwendig, wobei eine Vielzahl an Variablen und ihre Beziehung zueinander berücksichtigt werden müssen. Die Entwicklungen auf diesem Gebiet haben zuletzt auch die Erwartungen befeuert, dass autonome Systeme mit menschenähnlicher Intelligenz demnächst Realität sein werden. Judea Pearl, Pionier der Kausalitätsforschung, sieht jedoch derzeit noch eine Reihe fundamentaler Hindernisse, wie etwa die Unfähigkeit dieser Systeme, auf Neues, nicht Vorprogrammiertes zu reagieren, die fehlende Erklärbarkeit („*black boxes*“, s. 3.7.4) für getroffene Vorhersagen und Empfehlungen sowie das fehlende Verständnis für Ursache-Wirkungsbeziehungen. Diese könnten jedoch durch rechnerische Werkzeuge zur Ursachenmodellierung gelöst werden (Pearl 2019).

In der Medizin werden Methoden der Künstlichen Intelligenz seit der Entwicklung computerunterstützter klinischer Entscheidungssysteme in den 1970er Jahren eingesetzt. Seit dieser Zeit ist es Forscherinnen und Forschern zunehmend gelungen, große Mengen unstrukturierter Daten, wie etwa natürlichsprachige Texte und Bilder, im Rahmen unterschiedlicher Formen des sogenannten maschinellen Lernens zu nutzen. Maschinelles Lernen hat sich in den letzten Jahren zu einer der erfolgreichsten Formen der Künstlichen Intelligenz entwickelt. Man versteht darunter die Möglichkeit, Computer mit der Fähigkeit, aufgrund der eigenen Tätigkeit – wie etwa dem Analysieren von Daten aus Überwachungskameras – zu lernen, zu versehen. Dies bedeutet, dass sich Computer selbst „re-programmieren“ und dass daher nicht alle ihre Tätigkeiten vorprogrammiert wurden. Der mit Daten gefütterte Computer generiert komplexe analytische Modelle und verfeinert und optimiert diese auf der Basis eines Lernmodells, um die Genauigkeit einer angestrebten Problemlösung zu verbessern.

Während vor diesen neuen Möglichkeiten des maschinellen Lernens in der klinischen Praxis relativ simple, lineare Modelle mit nur einigen Variablen verwendet wurden und es mühsamer und langwieriger Arbeit bedurfte um Daten aus verschiedenen Quellen zu gewinnen und zu standardisieren, verwenden manche Lernmodelle heute bereits Millionen von Daten und Zehntausende Prädiktoren um zu genaueren Prognosen zu gelangen (Rajkomar et al. 2018). Die Forschungsliteratur beschreibt zahlreiche Beispiele des maschinellen Lernens, die bessere Diagnosen, etwa auch bei seltenen Erkrankungen, das rechtzeitige Identifizieren von Risikopatientinnen und Risikopatienten, das Verhindern unerwünschter Wirkungen von Medikamenten sowie eine effizientere Allokation von Ressourcen versprechen.

### 3.7.1 Was ist maschinelles Lernen?

Der „lernende“ Anteil des maschinellen Lernens beruht auf Algorithmen, die auf Optimierung hin angelegt sind. Algorithmen sind nicht neu, und keineswegs eine Erfindung des Computerzeitalters. Algorithmen sind regelbasierte Prozeduren, die es – Menschen oder Computern – erlauben, ein Problem systematisch zu lösen. Sie bestehen aus einer Sequenz eindeutiger Instruktionen, z. B. einer „wenn–dann“-Entscheidungskette mit mehreren Verzweigungen. Algorithmen, die im Kontext des maschinellen Lernens verwendet werden, dienen im Allgemeinen dazu Fehler zu minimieren oder die Genauigkeit eines Outputs, wie etwa einer Prognose, zu maximieren. Algorithmen haben typischerweise die Form einer Formel, eines Diagramms, wie etwa ein Entscheidungsbaum oder einer Skala. Einfache Beispiele aus der Medizin sind der Body-Mass-Index oder der *Framingham risk score* für kardiovaskuläre Erkrankungen, die auf der Grundlage eines Algorithmus berechnet werden.

Die Ergebnisse eines maschinellen Lernverfahrens hängen von der Qualität des Input (den Daten) und des Algorithmus ab. Üblicherweise sind die ersten Vorhersagen nicht besonders genau oder sogar falsch und es ist möglich, ihre Abweichung vom korrekten Ergebnis zu messen und den Fehler zur Verbesserung des Algorithmus zu verwenden. Dies ist allerdings nur möglich, sofern es zur Validierung der Ergebnisse geeignete Datensätze gibt. Diese Fähigkeit haben **neuronale Netzwerke**. Sie messen die Abweichung (oder den Fehler) und modifizieren die Parameter bis eine höhere Genauigkeit erzielt ist. Sie stellen einen **Optimierungsalgorithmus** dar.

Von „**tiefem Lernen**“, einer Unterkategorie des maschinellen Lernens, spricht man im Fall von Algorithmen, die eine besondere Genauigkeit bei der Lösung zahlreicher Probleme erzielt haben. Sie basieren auf künstlichen neuronalen Netzen. Dies sind im Wesentlichen mathematische Modelle, die dem Vorgang nachgebildet sind, wie das menschliche Gehirn Informationen interpretiert und aus diesen Schlüsse zieht. Ein neuronales Netz besteht aus vielen einzelnen Neuronen, die meistens in mehreren miteinander verbundenen Schichten (Layern) angeordnet sind. Die Zahl der Layer bestimmt unter anderem den Grad der Komplexität, den ein künstliches neuronales Netz abbilden kann. Viele Layer machen ein neuronales Netz „tief“.

Für das Verständnis maschinellen Lernens ist außerdem die Unterscheidung verschiedener Typen wichtig (Jannes et al. 2018; Deo 2015):

- Von Menschen überwachte Maschinenlern-Algorithmen wenden Erlerntes auf neue Daten an, um zukünftige Ereignisse vorherzusagen. Der Lernalgorithmus wird in diesem Fall mit vorgegebenen Beispielen trainiert, zu denen das intendierte Ergebnis bereits bekannt ist und er ist, nach ausreichendem Training, imstande, selbständig Ergebnisse mit neuem Datenmaterial zu generieren. Es vergleicht außerdem die erzielten Ergebnisse mit dem intendierten Output um Fehler zu identifizieren und das Modell zu modifizieren.
- Im Gegensatz dazu werden unüberwachte Maschinenlern-Algorithmen für Probleme verwendet, bei denen die Daten noch nicht klassifiziert oder beschrieben wurden. Der Lernalgorithmus ist in der Lage, neue Muster und Korrelationen selbstständig zu erkennen, ohne dabei auf vorgegebene Prototypen zurückzugreifen. Es wird spontan nach Merkmalen klassifiziert.
- Zwischen diesen beiden beschriebenen Lernverfahren liegt das bestärkende Lernen. Hierbei wird zwar wie beim supervisierten Lernen ein Ziel vorgegeben, doch der Algorithmus muss die Realisation wie beim nicht überwachten Lernen selbstständig finden. Der Algorithmus geht in einer Art „*Trial and Error*“-Verfahren vor, um das Ergebnis zu optimieren.

Ein einfaches Beispiel supervisierten Lernens ist etwa die automatisierte Interpretation eines EKGs oder von Röntgenbildern auf der Grundlage eines Mustererkennungsverfahrens. Hier wird dem Algorithmus eine beschränkte Auswahl von Diagnosen vorgegeben und die Aufgabe des Algorithmus besteht darin, die Daten korrekt zu klassifizieren. In diesem Fall – wenn man die richtige Diagnose kennt – ist es auch möglich, die Qualität des Algorithmus zu bewerten, was in anderen Kontexten oft nicht möglich ist. Der Gewinn gegenüber der Diagnose durch eine erfahrene Ärztin bzw. einen erfahrenen Arzt ist die durch die Analyse einer großen Menge von Daten möglicherweise erhöhte Genauigkeit der Diagnose.

Beim nicht supervisierten Lernen geht es hingegen darum, neue Muster in bislang unstrukturierten und unkategorisierten Daten zu erkennen. Angesichts der Heterogenität und Multifaktorialität vieler Erkrankungen besteht ein großes Interesse daran, deren Varianten (im Sinne einer Präzisionsmedizin; s. *National Research Council* 2011; Prainsack 2015) zu identifizieren und zu definieren und somit neue, besser angepasste Therapiemöglichkeiten entwickeln zu können.

Trotz der enormen Fortschritte in den Analyseverfahren ist maschinelles Lernen mit bislang nur ansatzweise gelösten Herausforderungen konfrontiert. Diese betreffen

- die hohe Komplexität der zu verarbeitenden Daten;
- das Risiko von Fehlerhaftigkeit und Bias der verwendeten Algorithmen;
- der tendenzielle Verzicht auf (kausale) Begründungen zugunsten von u. U. nicht interpretierbaren Korrelationen.

### 3.7.2 Datenkomplexität

Maschinelles Lernen in Verbindung mit „*Big Data*“ ist nicht nur mit einer großen Anzahl von Daten, sondern auch mit sehr unterschiedlichen Datenquellen und der Heterogenität dieser Daten konfrontiert. In Bezug auf die Gesundheitsversorgung sind dies etwa Forschungs- und Entwicklungsdaten, wie Labor- oder Pharmadaten, klinische Verwaltungsdaten, elektronische Gesundheitsdaten sowie patientengenerierte Gesundheitsdaten. Diese stammen aus sozialen Medien oder anderen Onlineressourcen, *Smartphone-Apps* und „*wearables*“ (mit Sensoren ausgestatteten, in die Kleidung integrierte oder implantierte Geräte). Dementsprechend heterogen ist die Modalität dieser Daten: bereits codierte Informationen, freier Text, sowie Bildmaterial, gemischt mit anderen Typen von Signalen oder Graphen. Diese Datenkomplexität steht hinter der Entwicklung von der klassischen multivariaten Datenanalyse zu einer neuen „*Data Science*“. Dies ist eine Disziplin, die sich damit beschäftigt, Daten zu „reinigen“, für die Analyse zu präparieren und zu analysieren.

Eng verbunden mit dem großen Volumen und der Heterogenität der zu verarbeitenden Daten ist ein Phänomen, das mit „Störgeräusch-Akkumulation“ (*noise accumulation*) bezeichnet wird. Prädiktive Analysen mithilfe statistischer Methoden beinhalten häufig simultane Schätzungen von mehreren Parametern. Beispielsweise sind Herzgefäßerkrankungen mit zahlreichen Faktoren assoziiert, wie Rauchen, Übergewicht, Diabetes, erhöhter Cholesterinspiegel, Bluthochdruck und anderen Risikofaktoren. Wenn der akkumulierte Schätzfehler (oder das so genannte „Rauschen“ (*noise*), also Signale, die keine Bedeutung für die zu beantwortende Frage haben) dieser Parameter hoch ist, kann dies zur Folge haben, dass der Lernalgorithmus Variablen mit signifikantem Erklärungswert übersieht. Das ist ein mathematisches Problem, dessen Bewältigung jedoch notwendig ist, um das Vertrauen in automatisierte medizinische Entscheidungssysteme zu stärken (Gandomi/Haider 2015).

Ein damit verbundenes weiteres Problem ist, was „zufällige Korrelation“ (*spurious correlation*) genannt wird. Damit wird das Phänomen bezeichnet, dass eine statistische Analyse zu falschen Ergebnissen kommen kann, wenn allein aufgrund des massiven Datenvolumens Korrelationen zwischen Daten ausgewiesen werden, die keinen sinnhaften Zusammenhang miteinander haben (z. B. die Anzahl der Störche in einer Region mit der Anzahl der Geburten; beide können im selben Zeitraum ansteigen, ohne sinnhaft miteinander verbunden zu sein). Je mehr Variablen, desto mehr Korrelationen können statistisch signifikant sein. Das hängt damit zusammen, dass in großen Datensets große Abweichungen eher auf Varianz (oder *noise*/Rauschen) als auf Information (oder „Signal“) zurückzuführen sind (Calude/Longo 2017). In einem Aufsatz mit dem Titel „The End of Theory: The Data Deluge Makes the Scientific Method Obsolete“, argumentierte Anderson (2008): „... with enough data, the numbers speak for themselves“.

### 3.7.3 Bias

Ein viel diskutiertes Problem ist das, was als potentielle systematische Verzerrung (Bias) von Algorithmen bezeichnet wird. Citron and Pasquale (2014) beschreiben dies als ein generelles Problem zunehmender Digitalisierung.

“By scored society, they mean the current state in which unregulated, opaque, and sometimes hidden algorithms produce authoritative scores of individual reputation that mediate access to opportunity. These scores include credit, criminal, and employability scores. Citron and Pasquale particularly focused on how such systems violate reasonable expectations of due process, especially expectations of fairness, accuracy, and the existence of avenues for redress” (Osoba/Welser 2017, 11).

Maschinelle Lernalgorithmen seien die Sklaven der Daten von denen sie lernen, wird argumentiert. Bekannt ist auch der Slogan „*garbage in garbage out*“, also „Müll rein und Müll raus“. Es werden verschiedene Quellen einer möglichen systematischen Verzerrung der Anwendung von maschinellen Lernalgorithmen unterschieden (Gianfrancesco et al. 2019):

- **Bias des Algorithmus:** Dieser Bias hat nichts mit dem Datenset zu tun, sondern ist eine mathematische Eigenschaft des verwendeten Algorithmus. Sein Gegenstück ist die Varianz: Algorithmen mit hoher Varianz (Empfindlichkeit auf kleinere Schwankungen in den Trainingsdaten) sind zwar mit einer höheren Datenkomplexität kompatibel, aber auch sensibler gegenüber Rauschen und daher weniger geeignet, mit Daten außerhalb des Trainingssets umzugehen.
- **Bias des Sample:** Ein solcher liegt vor, wenn das Datenset, welches zum Training des Algorithmus verwendet wird, den Problemraum für das Modell nicht adäquat repräsentiert. Es kommen verschiedene Techniken zur Vermeidung eines solchen Bias zum Einsatz, wie die Validierung von Samples im Hinblick auf ihre Repräsentativität oder die Identifizierung von Charakteristika der Population, die das Sample repräsentieren sollte. In Bezug auf die USA argumentieren beispielsweise Cohen und Grave (2017), dass ethnische Minoritäten, Frauen, Menschen mit niedrigem sozial-ökonomischem Status und Migrantinnen bzw. Migranten nur unzureichend in Datensätzen (elektronische Patientendateien, Genomdatenbanken, Präsenz im Internet und den sozialen Medien) repräsentiert seien und damit auch deren spezifischen Gesundheitsprobleme. So waren laut Need und Goldstein (2009) 96% der Teilnehmer an Genome-Mapping-Studien europäischer Herkunft.
- **Auf Vorurteilen beruhender (*prejudicial*) Bias:** Damit ist der Einfluss kultureller oder anderer Stereotypen auf Patientendaten gemeint. Zum Beispiel weist Hamerlund (2018) ein ethnisches Bias in den Prädiktoren für invasive chirurgische Eingriffe im Fall akuter myokardialer Infarkte nach. Dieser, die Gesundheitsdaten von Patientinnen und Patienten afro-amerikanischer Herkunft beeinflussende, Bias hängt damit zusammen, dass diese mit größerer Wahrscheinlichkeit von Chirurgen und Chirurgen mit höheren (risiko-bereinigten) Mortalitätsraten bei

Eingriffen behandelt werden, aber auch damit, dass es dieser Patientengruppe schwerer fällt, die Symptome eines Infarkts zu identifizieren und sie später einen Arzt aufsuchen. Wichtig ist es, zu betonen, dass auf Vorurteilen beruhender Bias nicht intendiert oder gar bewusst sein muss; so genannter impliziter Bias (*implicit bias*; Greenwald/Krieger 2006) tritt z. B. dann auf, wenn Menschen, die sich selbst keiner Vorurteile bewusst sind, die dominanten Vorstellungen von Hierarchien oder Gesellschaftsordnungen in ihr Handeln übersetzen, ohne sich dessen bewusst zu sein (FitzGerald/Hurst 2017). Ein Beispiel dafür stellt die Praxis dar, Beschwerden von Frauen, die auf Herz-Kreislaufkrankungen hindeuten könnten, weniger ernst zu nehmen, weil letztere als „männliche“ Erkrankungen gelten. Umgekehrt galt Osteoporose lange Zeit als ein Problem von Frauen nach der Menopause, obwohl ein Drittel der mit dieser verbundenen Brüche des Hüftgelenks bei Männern auftreten und diese die zweifache Mortalitätsrate von Frauen aufwiesen (Looker et al. 1997). Vor solchen systematischen Verzerrungen sind weder Menschen noch Maschinen (die ja dann bereits mit verzerrten Daten gefüttert wurden) gefeit.

- **Mess-Bias:** Fehlklassifikation von Erkrankungen und Messfehler sind häufige Quellen von Bias in Beobachtungsstudien und Analysen von elektronischen Patientendateien (Pot et al. 2019). Auch Gianfrancesco et al. (2019) bringen das Beispiel der ungleichen Diagnose und Behandlung von Frauen und Männern, die an einer koronaren Herzerkrankung leiden. Ein auf solchen verzerrten Daten basierender Lernalgorithmus wird den Bias der medizinischen Praxis reflektieren und Patientinnen bzw. Patienten fehlassifizieren.

Das Thema Gender Bias in der Medizin findet derzeit viel Beachtung. Es illustriert die verschiedenen Formen des Bias: aufgrund der unzureichenden Repräsentation von Frauen in klinischen Studien sowie aufgrund von Vorurteilen und Messfehlern. Auch die Bioethikkommission hat seit 2007 das Thema Gender in alle Diskussionen einbezogen. Eine Studie über die Art und Weise, wie Frauen und Männer somatische Symptome beschreiben, zeigt beträchtliche Unterschiede (Barsky et al. 2011). In „Pain and Prejudice“ argumentiert Jackson: „*Centuries of female exclusion has meant women’s diseases are often missed, misdiagnosed or remain a total mystery*“ (Jackson 2019). Trotz vielfältiger Bemühungen dieser Vernachlässigung der Frauen in der Forschung und dem resultierenden Gender Bias entgegenzuwirken, sind diese immer noch auf ein begrenztes Set von Erkrankungen beschränkt, mit beträchtlichen nationalen Unterschieden. Beispiele sind Alzheimer, wo der Vorteil der durchschnittlich besseren Gedächtnisleistungen von Frauen dazu führt, dass diese oft (zu) spät diagnostiziert werden (Sundermann et al. 2017). Im Bereich psychischer Erkrankungen ist der Gender Bias besonders ausgeprägt. Beispiele sind der Ausdruck von Schmerzen durch Frauen und Männer und ihre Interpretation durch die behandelnden Ärztinnen und Ärzte (Samulowitz et al. 2018); oder das Faktum, dass Männer und Buben wesentlich häufiger (im Verhältnis 4:1) die Diagnose ADS erhalten als Frauen und Mädchen, da die Diagnoseinstrumente hauptsächlich an männlichen Probanden entwickelt und validiert wurden (z. B. Beggiato et al. 2017).

Diesen und zahlreichen weiteren Studien ist zu entnehmen, dass der „Bias des Sample“ (und dieser gleichzeitig auch ein auf Vorurteilen beruhender [„prejudicial“] Bias) potentiell viele Erkrankungen und ihre Repräsentation in *Big Data* Datenbanken betrifft. Chin Yee und Upshur (2018) argumentieren, dass diese verschiedenen Formen des Bias den der medizinischen Praxis inhärenten Quellen der Unsicherheit bei Diagnose und Therapieplanung entsprechen. Es gäbe wenig gute Gründe anzunehmen, dass diese Unsicherheiten und Fehler durch die schiere Menge von Daten ausgeglichen werden könnten.

### 3.7.4 Begründbarkeit

Maschinelles Lernen divergiert nicht grundsätzlich von der konventionellen Statistik. Doch während diese auf das Testen von Hypothesen bezüglich kausaler Zusammenhänge fokussiert, geht es beim maschinellen Lernen weniger um die Interpretierbarkeit von Modellen. Der Fokus liegt vielmehr auf der prädiktiven Performanz und Generalisierbarkeit der Modelle. Der Kern ärztlichen Handelns, so argumentieren Chin-Yee und Upshur (2018), liege jedoch gerade im Begründen – *„clinical judgement refers to the range of complex reasoning tasks and actions performed by clinicians in the context of offering diagnosis, therapeutic options, and prognosis to patients regarding their health and illness“* (638).

Maschinelle Lernalgorithmen werden häufig als uneinsehbare „*Black Boxes*“ bezeichnet, denn die Methoden der Generierung von Daten-Modellen lassen sich nur schwer (falls überhaupt) interpretieren, da die Funktionen, die die Daten (Input) mit einem bestimmten Ergebnis (Output) verbinden zu komplex sind. So mag ein voll automatisierter, auf Algorithmen basierter Prozess das Bemühen, eine kausale Erklärung zu finden unterlaufen. Das heißt, dass Ärztinnen und Ärzte sich mit dem Problem konfrontiert sehen, dem System vertrauen zu müssen, ohne dass dessen Schlüsse für sie einsichtig sind und ohne dass sie diese ihren Patientinnen und Patienten erläutern können (Vellido 2018).

Andererseits wird argumentiert, das Beharren auf einsichtigen kausalen Zusammenhängen könnte die Vorteile algorithmen-generierter Prozesse kompromittieren, weil die dafür notwendigen Simplifikationen die Effizienz der Modelle untergraben. Modelle könnten sich in der Praxis als genau und effizient erweisen, obwohl sie nicht transparent sind. Außerdem werden Bedenken gegenüber dem Ausschließen von Korrelationen geäußert, die sich aufgrund des gegenwärtigen Wissensstandes der Medizin nicht erklären lassen, aber dennoch hohen Prädiktionswert haben (Zarsky 2018). Diese Korrelationen können jedoch nur in ihren spezifischen Kontexten von Bedeutung sein, während sie keine Generalisierung erlauben. Der Verzicht auf das Auffinden von, mit der Korrelation verbundener, Kausalität bzw. von einem zu Grunde liegenden Mechanismus kann dazu führen, mögliche Nebeneffekte zu übersehen. Auch vergibt man sich die Möglichkeit, grundlegende Erkenntnisse zu gewinnen, beispielsweise über Krankheitsmechanismen.

Ein weiteres Argument lautet, dass der Fokus auf nicht weiter erklärbare Korrelationen möglicherweise Muster zu Tage fördere, die im Sinne der Anwendbarkeit interessant sein können; doch werden auch zusätzlich Fehlerquellen eingeführt. Beispielsweise können mit Augenmerk auf Korrelationen Individuen unterschiedlich gruppiert und in



der Folge unfair behandelt bzw. stigmatisiert werden (Zarsky 2018). Ein klassisches und weit verbreitetes Beispiel dafür ist die Verwendung der Kategorie „Race“ (meist selbst identifiziert) als ungenau definiertes Proxy für die Zugehörigkeit zu einer in erster Linie soziokulturell definierten Population und die Vorhersage von daraus resultierenden Krankheitsrisiken für einzelne Individuen. Dies hat beispielsweise zu Fehldiagnosen von Hämoglobinopathien (Sichelzellen-Anämie,  $\beta$ -Thalassämie) oder einer Unterdiagnose von Cystischer Fibrose bei Afroamerikanerinnen und Afroamerikanern geführt (Yudell et al. 2016). Diese Praxis, die gerade in den letzten Jahren (der sogenannten post-genomischen Ära) wieder verstärkt in medizinische Fachpublikationen Eingang gefunden hat führt dazu, dass u. a. das Verständnis für die genetische Vielfalt von Menschen dem völlig veralteten Konzept kontinentaler „Rassen“ neuerlich unterworfen wird, anstatt dieses endlich aufzulösen und valide, d. h. auch durch einen kausalen Zusammenhang erklärbare Korrelationen von genetischen Markern zu identifizieren (Bonham et al. 2018; Cooper et al. 2018). Da dies ein ausgesprochen komplexes Gebiet ist, ist in diesem Kontext nicht mit einer baldigen Lösung zu rechnen, obwohl der Ruf nach neuen Lösungen und einer Beschäftigung mit dem Thema lauter wird (Yudell et al. 2016; Bonham et al. 2018; Cooper et al. 2018). Darüber hinaus sind simplifizierte gen-deterministische Erklärungen für Intelligenz oder menschliches Verhalten auch in Zeiten von „Big Data“ modern und werden in Medienberichten auch bevorzugt aufgegriffen (Barton et al. 2019; Rosenberg et al. 2019; Torkamani et al. 2018; Comfort 2018).

Es gibt allerdings auch Bemühungen, dem „Black Box“-Effekt, der durch die Schwierigkeiten Ergebnisse nachzuvollziehen bzw. zu interpretieren entsteht entgegenzuwirken. Beispiele sind etwa regelbasierte Repräsentationen, die mit medizinischen Begründungen kompatibel sind oder Nomogramme, wie sie von Medizinerinnen und Medizinern zur Visualisierung der relativen Gewichte der einer Diagnose zugrundeliegenden Symptome verwendet werden. Wachter et al. (2018) diskutieren die Möglichkeit, Transparenz über das zur Verfügung-Stellen von kontrafaktischen Erklärungen zu erhöhen. Es sei nicht unbedingt notwendig, den Algorithmus selbst verständlich zu machen, was aufgrund dessen Komplexität häufig auch gar nicht möglich wäre. Kontrafaktische Erklärungen bestehen aus einfachen „wenn–dann“-Aussagen, beispielsweise: „*If your 2-Hour serum insulin level was 154.3, you would have a score of 0.51; If your Plasma glucose concentration was 158.3 and your 2-Hour serum insulin level was 160.5, you would have a score of 0.51*“ (Wachter et al. 2018, 21). Wachter et al. betrachten kontrafaktische Erklärungen als eine zumutbare, „*lightweight form of explanation*“; diese ließen sich zudem leicht automatisch berechnen:

“Unlike existing approaches that try to provide insight into the internal logic of black box algorithms, counterfactual explanations do not attempt to clarify how decisions are made internally. Instead, they provide insight into which external facts could be different in order to arrive at a desired outcome” (43).

Kontrafaktische Deduktion (*counterfactual inference*) kann auch eingesetzt werden, um von Maschinen getroffene Vorhersagen „*approximately fair*“ zu machen, indem *fairness classifier* eingeführt werden, die unabhängig vom Kausalitätsmodell, das der Algorithmus verwendet, funktionieren (Russell et al. 2017).

## 4 Fallbeispiele

Wie schon in den vorangegangenen Kapiteln ausgeführt, sind mit *Big Data*-Analysen und Künstlicher Intelligenz große Erwartungen in Hinblick auf eine verbesserte Gesundheitsversorgung verbunden.

Beispiele zeigen, dass sich durch den Einsatz von maschinellem Lernen und Künstlicher Intelligenz Vorteile hinsichtlich einer besseren standardisierten, optimierten Auswertung von in erster Linie Bilddaten in der Diagnostik ergeben. Für die endgültige Diagnosefindung sind allerdings zusätzliche klinische Aspekte relevant und daher ist für die endgültigen Entscheidungen in jedem Fall die Expertin bzw. der Experte notwendig. Ein großer Vorteil von maschinellem Lernen und Künstlicher Intelligenz ist darin zu sehen, dass es damit möglich ist, mehr Menschen weltweit eine optimale Diagnostik zugänglich zu machen. Unabdingbare Voraussetzung für derartige Entwicklungen ist allerdings eine gute standardisierte Datengrundlage, wie es auch im Rahmen einer Studie der Universitätsklinik für Dermatologie der Medizinischen Universität Wien gezeigt werden konnte (s. u.). In der Folge sollen einzelne Beispiele für den Einsatz von KI im klinischen Betrieb vorgestellt werden, wiewohl diese derzeit vor allem im Rahmen wissenschaftlicher Projekte bearbeitet und nur in sehr eingeschränktem Ausmaß bereits in der medizinischen Routine eingesetzt werden.

### 4.1 Public Health – Gesundheitsplanung

Infolge von Datenanalysen kann auf Veränderungen im Gesundheitszustand der Bevölkerung früher reagiert und können entsprechende Präventionsmodelle entwickelt, sowie zeitgerecht die zur Versorgung notwendigen Einrichtungen geschaffen werden. So gibt es ein Projekt der Sozialversicherung gemeinsam mit Stefan Thurner vom Institut für die Wissenschaft komplexer Systeme der Medizinischen Universität Wien im Rahmen dessen an Hand anonymisierter Datensätze Vorhersagen zu gesundheitlichen Entwicklungen getroffen werden können. Zu diesem Zweck wurde jede Erkrankung, jede ärztliche Behandlung von acht Millionen Österreicherinnen und Österreichern, egal ob diese im niedergelassenen Bereich oder im Spital erfolgte, in den Jahren 2006–2007 analysiert. Damit kann eine Vorhersage des Risikos für jede einzelne Erkrankung (insgesamt 1642)

in unterschiedlichen Teilen der Bevölkerung gemacht werden (Sauter et al. 2014; Klimek et al. 2016), abhängig von Alter und Geschlecht. Die Analyse erfolgte anonymisiert und ermöglicht eine Erstellung einer „Krankheitsdemographie“ für Österreich.

Weiters untersuchten Stefan Thurner und Peter Klimek in Zusammenarbeit mit Alexandra Kautzky-Willer, Universitätsklinik für Innere Medizin III der Medizinischen Universität Wien die personalisierten Erkrankungsrisiken für Diabetespatientinnen und Diabetespatienten. Es konnten mehr als 100 „Erkrankungspaare“ identifiziert werden (Klimek et al. 2015). Damit haben behandelnde Ärztinnen und Ärzte die Möglichkeit mittels Risikoscores präventiv zu therapieren bzw. die Patientinnen und Patienten gezielt in Richtung möglicher Folgeerkrankungen zu befragen um frühzeitig reagieren zu können. Derartige Analysen ermöglichen auch die Abschätzung zukünftig entstehender Kosten im Gesundheitssystem.

## 4.2 Bildgebende Verfahren

Im Vorfeld finden sich vor allem Methoden des maschinellen Lernens, wo auf Basis von Algorithmen die Diagnostik in der Onkologie (z.B. betreffend Brust-Tumore, Lungentumore) unterstützt wird. So ermöglichen neue sensitive Techniken (PET/MRI) und multiparametrische Auswertungen eine bessere Differenzierung und Klassifizierung von bösartigen und gutartigen Brustveränderungen (Vogl et al 2019). Erster Schritt ist hier eine Verbesserung bei der „Erkennung“ einer Erkrankung, in einem zweiten Schritt geht es aber in Richtung Prognose bzw. Verlauf um eine möglichst individuelle zielgerichtete Therapie zu gewährleisten (*European Society of Radiology [ESR] 2019*).

Eine weitere Anwendung beschäftigt sich mit der Erkennung von „Phänotypen“ in großen klinischen Routinepopulationen. Hier werden Bilddaten genutzt um Gruppen von Patientinnen und Patienten mit ähnlichen Charakteristika zu identifizieren und so diagnostische Kategorien zu hinterfragen, um sie als Werkzeuge für Entscheidungsfindungen in Modellen des maschinellen Lernens zu etablieren.

Zukünftige Anwendungen fokussieren in erster Linie auf den Einsatz neuer Medikamente und Therapien, da hier Methoden, die die Vorhersage von Behandlungseffekten z. B. im Rahmen klinischer Studien ermöglichen, zu einer besseren Charakterisierung von Patientengruppen beitragen können.

## 4.3 Neurologie

In der Neurologie sind bildgebende Verfahren ein wesentliches Werkzeug für die klinische Diagnostik. Ein wichtiges Anwendungsgebiet ist die strukturelle Magnetresonanztomographie (MRT), auch *Magnetic Resonance Imaging* (MRI), oder Kernspintomographie zur Differenzialdiagnose neurodegenerativer Erkrankungen, zum Beispiel der Unterscheidung atypischer Parkinson-Syndrome von Morbus Parkinson (Scherfler et al. 2016). Da die

Diagnostik derzeit vor allem auf das Frühstadium der Erkrankungen abzielt, sind in neuerer Zeit entwickelte Verfahren zur automatisierten Erkennung von krankheitsrelevanten Hirnarealen und die 3D-Volumetrie dieser Areale von großer Bedeutung. Wie eine Studie der Medizinischen Universität Innsbruck kürzlich zeigte, konnte mittels KI-gestützter Bildauswertung eine wesentliche Verbesserung bei der Erkennung und Zuordnung atypischer Parkinson-Syndrome etabliert werden, was sowohl für die Behandlung als auch für die Durchführung klinischer Studien von besonderer Relevanz ist (Krismer et al 2019).

Trotz der durch eine Form Künstlicher Intelligenz möglichen Steigerung von Effizienz und Sensitivität bleibt die Validierung durch in der Hirnanatomie erfahrenes Personal allerdings essentiell (Scherfler et al. 2016).

Derartige Systeme entlasten somit einerseits und liefern einen genaueren diagnostischen Einblick. Sie generieren jedoch auch Arbeit. Derzeit rechnet man damit, dass sich die Diagnostik in diesem Bereich durch computergestützte Bild-Auswertungen um etwa 20–30 % verbessern wird.

## 4.4 Dermatologie

Im Rahmen einer Studie der Universitätsklinik für Dermatologie der Medizinischen Universität Wien konnte gezeigt werden, dass Künstliche Intelligenz dem Menschen bei der Diagnose von pigmentierten Hautveränderungen wie Muttermalen und Melanomen überlegen ist. Ziel war die Unterscheidung zwischen gutartigen und bösartigen pigmentierten Hautveränderungen. Lerngrundlage war die Bilddatenbank HAT10.000, die mehr als 10.000 digitalisierte auflichtmikroskopisch angefertigte Aufnahmen von sieben verschiedenen Typen pigmentierter Hautveränderungen umfasst (Tschandl et al. 2018). 511 Dermatologinnen und Dermatologen unterschiedlichsten Ausbildungsgrades aus 63 Ländern, von Anfängerinnen und Anfängern bis zu Expertinnen und Experten mit langjähriger Expertise, wurden hier mit den Befund-Algorithmus von 77 mit automatisierten Auswertungen arbeitenden Labors verglichen.

Die besten menschlichen Expertinnen und Experten erkannten 18,8 Bilder von 30 richtig, während die besten Maschinen 25,4 richtige Klassifizierungen schafften. Es konnte hier sehr eindrücklich gezeigt werden, dass die wichtigste Grundlage solcher Ergebnisse eine umfassende, gut standardisierte und qualitativ hochwertige Datenbank ist (Tschandl et al. 2018).

Die Autorinnen und Autoren weisen allerdings darauf hin, dass die Maschine bei diesem Experiment zwar klar überlegen ist, aber den Menschen keinesfalls ersetzt, da der Computer nur eine optische Momentaufnahme analysiert, zur Diagnose eines Patienten aber auch die Verlaufsbeobachtung gehört sowie die Einschätzung, ob der Betroffene aufgrund diverser Begleitumstände ein Risikopatient ist, wie sich eine Veränderung der Haut durch Ertasten anfühlt und der Vergleich mit anderen Muttermalen am Körper dazu gehört. Die Interpretation dieser Ergebnisse sei weiterhin dem Menschen überlassen.

## 4.5 Ophthalmologie

Ein international viel beachtetes Projekt ist ein von der Universitätsklinik für Augenheilkunde und Optometrie in Zusammenarbeit mit dem Zentrum für medizinische Physik und biomedizinische Technik der Medizinischen Universität Wien entwickeltes bildgebendes Verfahren. Die hochauflösende optische Kohärenztomographie (OCT) ermöglicht die frühzeitige Diagnostik von Netzhauterkrankungen und damit deren gezielte Behandlung.

Dabei handelt es sich um ein bildgebendes Verfahren, das berührungsfrei genaue Schichtbilder der Netzhaut herstellt und mittels automatisierter Algorithmen auswertet. Mit dieser Methode können mit einer hohen Auflösung innerhalb weniger Sekunden Erkrankungen der Netzhaut diagnostiziert werden, um die therapeutisch notwendigen Schritte im Sinne einer personalisierten Medizin zu setzen. Für die Entwicklung der OCT als Methode wurden Christoph Hitzenberger und Adolf Fercher vom Zentrum für medizinische Physik und Biomedizinische Technik 2017 mit dem Dolores H. Russ Prize, dem „Nobelpreis für Ingenieurwissenschaften“ geehrt.

Es ist geplant, dass in naher Zukunft jede Augenärztin und jeder Augenarzt auf diese Technologie zugreifen kann und damit die Diagnostik und Therapie für die ca. 170 Millionen Menschen weltweit, die an einer Makuladegeneration leiden, optimiert wird.

Die in 4.4 und 4.5 beschriebenen Projekte sind Beispiele dafür, dass mit Hilfe von Künstlicher Intelligenz und neuen Technologien eine bessere standardisierte Versorgung vieler Patientinnen und Patienten niederschwellig ermöglicht werden kann (Gerendas et al. 2018).

## 4.6 Onkologie

Das Potential maschinellen Lernens für die Onkologie lässt sich an einem Projekt der Grundlagenforschung des Christian Doppler Labors für *Applied Metabolomics* (CDL-AM) der Medizinischen Universität Wien illustrieren. Dieses beschäftigt sich mit der Entwicklung eines nicht-invasiven diagnostischen Verfahrens zur Optimierung von Therapie und Verlaufskontrollen in der Onkologie. Letztendlich soll ein Algorithmus zur Verfügung stehen, der aus individuellen Patientendaten mit hoher Wahrscheinlichkeit den weiteren Verlauf der Erkrankung prognostizieren kann.

Im Rahmen des Projekts werden dafür Algorithmen entwickelt, die funktionelle Bilddaten aus der Positronen-Emissions-Tomographie (PET) mit histopathologischen Bildsowie auch genetischen Daten aus Gewebeproben derselben Patientinnen und Patienten kombinieren. Diese Daten werden verwendet, um mit maschinellen Lernprogrammen die PET generierte Krebsdiagnose so zu optimieren, dass eine invasive Diagnostik (Biopsie) nur mehr in Ausnahmefällen notwendig ist. Neben der schwierigen Entnahme kann so auch die eingeschränkte Aussagekraft einer Biopsie in einem immer größer werdenden, heterogenen Tumor vermieden werden. Mutationen in der Erbsubstanz und auch epigenetische Veränderungen des Chromatins beeinflussen den Stoffwechsel und damit

auch die Architektur und das Aussehen des Tumors. Mittels PET können Textur und Entwicklungsphasen, aber auch der Stoffwechsel eines Tumors funktionell bestimmt und Metastasen nachgewiesen werden. Dies ermöglicht die strukturelle und funktionelle Vermessung der gesamten Tumormasse in Echtzeit.

Die Bilddaten sind daher eine Reflexion des Status des Tumor-Stoffwechsels, aus denen sich prognostische Aussagen über den weiteren Verlauf der Erkrankung ableiten lassen. Das Verfahren wird gemeinsam mit dem Unternehmenspartner Siemens Medical Solutions entwickelt. Zunächst werden häufig vorkommende Krebsarten wie Kolorektalkarzinome, Prostatakarzinome und aggressive Tumore aus dem HNO-Bereich analysiert. Im ersten Schritt werden lernende Algorithmen mit PET-Daten von Patientinnen und Patienten gefüttert, die am AKH Wien behandelt wurden und von denen auch die auf Gewebeproben basierte histo- und molekularpathologische Diagnose sowie deren vollständige Krankengeschichte bekannt sind. Auf diese Weise wird eine retrospektive Korrelation zwischen bildgebenden Parametern, der gewebsbasierten morphologischen Diagnose sowie (epi)genetischen Parametern und den entsprechenden Krankheitsverläufen hergestellt.

Parallel dazu werden zudem Tiermodelle entsprechend den untersuchten Krebserkrankungen generiert, die eine „*proof of principle*“ Überprüfung der Korrelation spezifischer Gendefekte und PET-Daten anhand einer speziell für Mäuse hergestellten  $\mu$ PET-Maschine ermöglichen.

## 4.7 Intelligente Aggregation und Visualisierung von Gesundheitsdaten

Der Einsatz intelligenter Anwendungen ist nicht auf die unmittelbare Diagnostik beschränkt, sondern ist darüber hinaus geeignet, die Informationsgrundlagen für das medizinische, therapeutische und pflegerische Handeln in Krankenanstalten besser aufzubereiten.

So wird in Krankenanstalten heute jede diagnostische, medizinische und therapeutische Handlung dokumentiert. Die Dokumentation wird jeweils in einem Krankenhaus-Informationssystem (KIS) gespeichert und verwaltet. Diese umfangreiche und detaillierte Dokumentation dient als Grundlage für nachfolgende Diagnose- und Therapieentscheidungen sowie das therapeutische und pflegerische Handeln. Das manuelle, nicht-automatisierte und nicht-gefilterte Abrufen von zahlreichen und mannigfaltigen Dokumenten im Original (z. B. Röntgenbilder, Laborbefunde, sonstige Texte und Scans) sowie das wiederholte Abfragen derselben Informationen von Patientinnen und Patienten wird von den Angehörigen des Gesundheitspersonals jedoch – nicht zuletzt aufgrund der zunehmenden Fülle an verfügbaren Informationen – als zusätzliche Arbeitsbelastung erlebt. Das Informationsproblem ist an den Schnittstellen besonders groß. Dies betrifft den innerprofessionellen (z. B. zwischen Ärztinnen und Ärzten verschiedener stationärer Abteilungen) als auch den multidisziplinären (z. B. zwischen Ärztinnen bzw. Ärzten

und Therapeutinnen bzw. Therapeuten) Informationsaustausch der unterschiedlichen Angehörigen des Gesundheitspersonals. Obwohl moderne KIS die Daten zunehmend strukturiert erfassen, abspeichern, ausgeben und verwalten, erfordert das Abrufen relevanter Information nach wie vor individuelle, manuelle und nicht-automatisierte Handlungen, das zu nicht-gefilterten Ergebnissen führt. Der mit der Informationsgewinnung verbundene Zeitaufwand ist unter anderem von den Zugriffsrechten und der Systemkenntnis der suchenden Person und von der Systemgeschwindigkeit abhängig. Er reduziert die faktische Behandlungszeit.

Es gibt daher verschiedene Bemühungen diese Schnittstelle durch die Anwendung von intelligenten (technologiebasierten) Lösungen zu verbessern. Ein FFG-Forschungsprojekt mit dem Titel „*Smart Aggregation and Visualisation of Health Data*“ (SMARAGD) zielt beispielsweise darauf ab, technische Komponenten für die intelligente Aggregation und Visualisierung von Informationen aus elektronischen Gesundheitsdaten zu entwickeln, die auf berufsgruppenspezifische Anforderungen in Krankenanstalten umfassend Rücksicht nimmt (exemplarisch bezieht sich das Projekt auf die für Ergotherapeutinnen und Ergotherapeuten sowie Physiotherapeutinnen und Physiotherapeuten relevante Informationen). Dieses Forschungsvorhaben wird seinem Gegenstand und Ziel entsprechend von unterschiedlichsten Disziplinen (Informatik, Medizinische und Gesundheitswissenschaften, Sozialwissenschaften, Rechtswissenschaften) getragen. Daraus erklärt sich die große Zahl an Forschungspartnern. Das Projekt wird von der IMC Fachhochschule Krets geleitet, weitere beteiligte Forschungseinrichtungen sind die SYNYO GmbH, die Karl-Franzens-Universität Graz, das Linz Institute of Technology der Johannes Kepler Universität Linz, die Medizinische Universität Graz, die Know-Center GmbH und die Universität Wien. Es wird von der Österreichischen Forschungsförderungsgesellschaft (FFG) teilfinanziert.

# 5 Ethische Aspekte der Künstlichen Intelligenz

Das Verhältnis von Ethik und Medizin ist uralte und hat im Lauf der Geschichte verschiedene Wandlungen erfahren. Wenn man „Ethik“ von „Moral“ abgrenzt und erstere als Theorie von Moralität und Sittlichkeit versteht, lässt sich „Moral“ als die Gesamtheit von sittlich bezogenen Prinzipien, Regeln und Normen auffassen, die für das Zusammenleben von verschiedenen Gesellschaften und Individuen Bedeutung erlangen. Die Handlungsziele von „Moral“ und „Ethik“ sind eng mit den Handlungszielen der Medizin verbunden. Vereinfacht lässt sich sagen, dass die Handlungsziele der „Ethik“ und „Moral“ in Grundlagen für ein gutes und gelingendes (Zusammen)Leben, das Handlungsziel der Medizin im Bewahren oder Wiederherstellen von Gesundheit liegt. Allen aber geht es nicht allein um Wissen, sondern um die Anwendung dieses Wissens in Hinblick auf ihre jeweiligen Handlungsziele.

„Medizinethik“ wird gegenwärtig in erster Linie als ein Bereich der sogenannten Angewandten Ethik verstanden, der es vorrangig um Orientierungshilfen und Anleitungen zu „gutem und richtigem medizinischem Handeln“ im Sinne der wesentlichen medizinethischen Prinzipien (Respekt vor Autonomie, Nicht-Schaden, Wohltun und Gerechtigkeit) geht.

Der Wandel in der Medizinethik seit der Antike (Hippokrates) hat vielerlei Ursachen. Zunächst ist es der durch Technik und Technologie verursachte unendlich vergrößerte Handlungsspielraum, zugleich aber auch die Pluralität der Lebensstile und Moralauffassungen im historischen und kulturellen Wandel, die einen solchen verursacht haben.

Als Kehrseite der nicht zuletzt durch *Big Data* und Künstlicher Intelligenz entstandenen Diagnose-, Therapieansätze und -möglichkeiten lässt sich ein gewachsenes Misstrauen gegenüber dem Gesundheitssystem konstatieren, das in erster Linie die Anonymität und den Verlust an Humanität sowohl am Krankenbett als auch in der Arztpraxis beklagt. Dies wurzelt nicht zuletzt in der wachsenden Macht der Bürokratien unserer Gesundheitssysteme sowie in der mangelnden Transparenz ihrer ökonomisch bedingten Entscheidungen.

Besondere Herausforderungen für die medizinische Ethik sind Grenz- und Extremsituationen wie etwa künstliche Befruchtung, technologiebasierte künstliche Lebensverlängerung, Organtransplantationen sowie intensivmedizinische Maßnahmen, bei denen es nicht nur um das Überleben sondern auch um die Lebensqualität der Patientin bzw. des Patienten geht. Derartige Entscheidungen können nicht nur auf Basis von Algorithmen gelöst werden, da sie nicht allein aufgrund quantitativer Kriterien getroffen werden können.

Aufgabe der Medizinethik muss es sein, dem Überwuchern von technologischen, ökonomischen und administrativen Faktoren Grenzen zu setzen und den humanen



Aspekt, die sogenannte „Würde“ des Menschen, als gleichberechtigt bei allen medizinischen Überlegungen einzufordern. Dass die Gefahr besteht, die Ärztin bzw. den Arzt als „Gesundheitsingenieure“ und die Patientin bzw. den Patienten als „zu reparierendes Objekt“ zu verstehen, bildet seit langem einen Fokus in der medizinethischen Diskussion.

## 5.1 Veränderungen der Beziehung von Ärztin bzw. Arzt und Patientin bzw. Patient

Der vermehrte Einsatz von Entscheidungssystemen, die auf maschinellem Lernen basieren, führt zu einer Veränderung der Arzt-Patienten-Beziehung. Das ist nicht unbedingt ein völlig neues Phänomen, da sich diese Beziehung historisch aufgrund der unterschiedlichsten Entwicklungen, wie die Vervielfältigung und teilweise Technisierung medizinischer Diagnose- und Therapieverfahren, die zunehmende Spezialisierung des Arztberufs und die organisatorische Komplexität der Gesundheitsversorgung, laufend verändert. Die elektronische Gesundheitsakte war ursprünglich dazu konzipiert, Abrechnungen zu erleichtern, bzw. zu erstellen. Sie war nicht für die Verwendung durch Ärztinnen bzw. Ärzte während des Gesprächs mit Patientinnen bzw. Patienten gedacht. Sehr oft stören sie das Wohlbefinden der Ärztinnen bzw. Ärzte und tragen zur Zermürbung bis hin zu Burnout-Symptomen bei. Es gibt zu viel zu dokumentieren, gerade auch bei der auch zwischenmenschlich so wichtigen Erstkonsultation (z. B. Ventres et al. 2006). Beispielsweise schlussfolgerten Lown und Rodriguez (2012) aus einer empirischen Studie: *„Screen-driven communication inhibits patients' narratives and diminishes clinicians' responses to patients' cues about psychosocial issues and emotional concerns“* (392). Das fundamentale Bedürfnis von (kranken) Menschen ist hingegen, von der Ärztin bzw. vom Arzt Zuwendung zu erhalten.

Die Beziehung von Ärztin bzw. Arzt und Patientin bzw. Patient beruht auf einem Deliberationsprozess, in dem Befunde, Beobachtungen und Entscheidungen kommuniziert, mit Begründungen versehen und nicht nur im Hinblick auf „Fakten“, sondern auch auf Werte, Präferenzen und Einstellungen erörtert werden (e.g. Wirtz et al. 2005). Die Befürchtung besteht, dass der Einsatz von maschinellem Lernen in der Medizin diese auf Vertrauen und Dialog beruhende Beziehung zugunsten statistischer „Argumente“ verdrängen könnte; vor allem dann, wenn die Ergebnisse einer prädiktiven Analyse nicht kausal interpretiert und einsichtig gemacht werden können (s. 2.3.4). Doch selbst, wenn es gelingt, alternative Erklärungsmodelle, wie etwa kontrafaktische Erklärungen, in der medizinischen Praxis zu etablieren, bleiben Bedenken gegenüber quantitativ-mathematischen Ansätzen der medizinischen Begründung bestehen, wie dies ein Beispiel von Cohen et al. (2017) illustriert:

“Imagine a woman undergoing treatment for breast cancer and trying to decide whether to opt for partial or the much more invasive radical mastectomy. The doctor recommends the radical mastectomy. When asked why, he says ‘for patients like you we know from the

data that it tends to be the best option.’ When she asks ‘what is it about my case that makes you think that’, shall he respond ‘the algorithm has examined 10,000 variables from your EHR and, based on its validated model, determines this is what is appropriate in your case?’ To be sure that may be much better than answering as the current physician might that ‘given the limited number of patients I have seen in my practice, what I learned in medical school, and what I have read in the literature, I think this will be better for you’ — but will the patient accept the former answer as better?” (452).

Ein weiteres ethisches Problem ist damit verknüpft, dass prädiktive Analyseverfahren tendenziell, wie bereits erwähnt, den Blickwinkel von individuellen Patientinnen und Patienten zu einer Patientenpopulation lenken. Regeln, die bereits jetzt Gesundheitsorganisationen zu implementieren versuchen, erhalten durch diese ein noch größeres Gewicht. Chin-Yee und Upshur (2017) argumentieren, dass dadurch der, ärztliches Handeln kennzeichnende, ethische, patienten-zentrierte Fokus auf die Fragen „*What is the significance for this patient?*“ und „*What should be done for this individual?*“ in den Hintergrund gedrängt würde:

“Such questions can be addressed, but require us to move beyond the quantitative evidence emphasized by data-driven approaches, contextualizing these approaches with the qualitative, personal evidence which emerges through a process of dialogue between physician and patient, and then applying the totality of the evidence through the exercise of phronesis” (644).

Letztlich bleiben Menschen Wesen, die Bindung, Aufmerksamkeit und Fürsorge benötigen. Krankheiten, insbesondere schwere Krankheiten, bewirken oft regressive Zustände in Menschen, da sie oftmals große Angst induzieren und Kindlichkeit, Abhängigkeit und Sicherheitsbedürfnis wecken. Gleichzeitig sollen aber auch die fortschrittlichsten technischen Möglichkeiten zur Verfügung stehen, wissenschaftliche Präzision, die beste Therapie – alles auch mit Hilfe von Künstlicher Intelligenz. Die Ärztin bzw. der Arzt soll die Patientin bzw. den Patienten kennen und sein Wissen fürsorglich und fachkundig anwenden. Die größte Hoffnung bei Künstlicher Intelligenz ist die Möglichkeit, die wertvolle Zeit, die gespart werden kann, dafür zu nützen, das Vertrauen zwischen Ärztinnen bzw. Ärzten und Patientinnen bzw. Patienten zu vertiefen. Es gäbe genügend Zeit bei den Konsultationen, sodass tiefere und mitfühlendere Verbindungen aufgebaut werden könnten. Es wäre auch überlegenswert, wie das Training von Ärztinnen und Ärzten derart umgestaltet werden kann, um dieses bessere Verständnis für Patientinnen und Patienten zu lehren. In seinem Buch „*Deep Medicine*“ zeigt Eric Topol, wie die Vermehrung von Maschinen das diagnostische Pouvoir erhöhen und sich somit der Fundus an medizinischem Wissen für alle Ärztinnen und Ärzte vergrößern wird.

## 5.2 Wo liegt die Verantwortung bei maschinenbasierten Entscheidungen?

In der Debatte um die Implementierung von Verfahren des Maschinenlernens wird immer wieder auf das häufige Auftreten von ärztlichen Fehlern verwiesen, verbunden mit der Annahme, solche Fehler ließen sich durch den Einsatz von Künstlicher Intelligenz vermindern. In einer mittlerweile klassischen Studie haben Fagerhaugh et al. (1987) die Quellen von Fehlern in klinischen Settings (Personen, Organisation der Arbeitsabläufe, Prozeduren, Geräte) beschrieben, aber auch die Schwierigkeiten von Gesundheitsorganisationen, diese Probleme zu adressieren und mit ihnen umzugehen. Außerdem verweist die Literatur auf zahlreiche Fehlerquellen der unterschiedlichsten medizinischen Technologien, einschließlich jener, die mit *Big Data*-Analysen verbunden sind: systematische Fehler (wie Bias der Daten), nicht-überprüfbare oder falsche Daten, usw. Ärztinnen und Ärzte, die ja weiterhin die Verantwortung für die Entscheidungsgrundlagen ihres Handelns tragen, können diese nur durch eine kritische Sichtung ihrer Informationsquellen einlösen, soweit dies im Arbeitsalltag möglich ist.

Mit dem Einsatz von *Big Data*, Künstlicher Intelligenz und maschinellem Lernen eröffnen sich mehrere Spannungsfelder, von deren Bewältigung abhängt, ob und wie Ärztinnen und Ärzte dieser Verantwortung nachkommen können.

### 5.2.1 Ärztliche Handlungsspielräume

Das Wahrnehmen von Verantwortung ist umso schwieriger zu gewährleisten, je weniger Handlungsspielräume Ärztinnen und Ärzten angesichts von (teilweise) automatisierten Entscheidungsprozessen offenstehen. Automationsunterstützte Entscheidungssysteme werden häufig mit unterschiedlichen Optionen für Interventionen verbunden. Um Ärztinnen und Ärzten ausgewählte Optionen bei Vorliegen eines spezifischen Befunds nahelegen, können in das zugrundeliegende Modell „Default“-Regeln oder „Framing“-Effekte eingebaut werden, die den Präferenzen der Gesundheitsorganisation entsprechen. Diese Überlegung geht auf die von Thaler und Sunstein (2008) vertretene, teils stark kritisierte, Idee des „*nudging*“ zurück – die Entscheidungsmöglichkeiten in einer spezifischen Situation so zu präsentieren, dass Personen die „vernünftigste“ wählen. Auf die Situation von Ärztinnen und Ärzten angewandt bedeutet dies, dass diese etwa beim Vorliegen des Risikos einer Sepsis durch das System auf die aus der Sicht der Gesundheitsorganisation ökonomischste Variante „hingelenkt“ werden und andere Varianten einer speziellen Begründung, das heißt auch eines zusätzlichen Aufwandes, bedürfen (Cohen et al. 2014).

Die Möglichkeit, die ihrer Rolle entsprechende Verantwortung wahrzunehmen, hängt demnach von den Entscheidungsspielräumen ab, die die Implementierung von prädiktiven Analysesystemen offenhält und davon, wie stark Ärztinnen und Ärzte in der Wahrnehmung dieser Spielräume unterstützt bzw. behindert werden. Cohen et al. (2014) diskutieren die Varianten einer „*Opting out*“- und „*Opting in*“-Option:

„The model could trigger a patient care intervention that would occur unless the physician or nurse overrode it (opting out) or could suggest an option that the physician or nurse would have to order or carry out (opting in). The choice of default options becomes even more challenging if the recommended action carries substantial risks as well as benefits“ (1145).

### 5.2.2 Das Kriterium der Systemkritikalität

In diesem Zusammenhang wird auch das Konzept der Systemkritikalität verwendet. Diese setzt sich zusammen aus dem Risiko, das mit einer Entscheidung verbunden ist und der strukturellen Entscheidungskritikalität. Dabei bedeutet „Risiko“ die Kombination aus Wahrscheinlichkeit eines Schadenseintritts (z. B. durch Fehlentscheidung) und Schwere des zu befürchtenden Schadens. Dagegen bedeutet strukturelle Entscheidungskritikalität die Kombination aus Komplexität der Entscheidung (vom schlichten Abbild der Realität über eine wertende Einschätzung der Realität bis hin zur multifaktoriellen Prognose einer künftigen Realität), Wirkungen der Entscheidung (von einem bloß abstrakt denkbaren Handlungskontext über einen konkreten Handlungskontext bis hin zur unmittelbaren Implementierung) und Reversibilität der Wirkungen (von voller Reversibilität bis hin zur Irreversibilität). Je höher die Systemkritikalität, desto höher sind die Systemanforderungen in Bezug auf die Rolle der Ärztin bzw. des Arztes im Entscheidungsprozess, Transparenz und Erklärbarkeit der Ergebnisse sowie Kontrolle des Systems.

Ein weiteres wichtiges Unterscheidungskriterium ist der Automatisierungsgrad. Eine von Pasumaran et al. (2000) eingeführte Taxonomie differenziert vier Stufen der Automatisierung: (1) das System filtert die für eine Analyse relevant eingestufte Information; (2) es verwendet diese Information für eine Diagnose; (3) es schlägt eine Handlungsabfolge vor; (4) es führt diese Handlungen aus. Erst auf dieser letzten Stufe ist die Entscheidung über Diagnose und Therapie tatsächlich voll automatisiert. Die kritische Frage bezüglich der Einführung eines auf maschinellem Lernen basierenden Entscheidungssystems ist demnach jene nach dem *„tradeoff in which more automation yields better human-system performance when all is well, but induces increased dependence so that it will produce more problematic performance when things fail“* (Wickens et al. 2010, 389). Das heißt, ein System mag, wie in den beschriebenen Fallstudien beschrieben, durchaus bessere Resultate erzielen als die behandelnde Ärztin bzw. der behandelnde Arzt, allerdings nur dann *„when all is well“* – alle Voraussetzungen für eine optimale Handlungsentscheidung erfüllt sind.

### 5.2.3 Der Umgang mit Bias und fehlender Transparenz

Ein für die Wahrnehmung von Verantwortung zentraler Aspekt sind die Möglichkeiten des Umgangs mit Bias und fehlender Transparenz, die Ärztinnen und Ärzten offenstehen. Maschinelle Lernalgorithmen werden häufig als nicht einsehbare *„Black Boxes“* bezeichnet. Damit ist gemeint, dass Ärztinnen und Ärzte sich unter Umständen mit dem Problem konfrontiert sehen, einem prädiktiven Analysesystem vertrauen zu müssen, ohne dass dessen Schlüsse für sie einsichtig sind und ohne dass sie diese ihren Patientinnen und Patienten erläutern können (Vellido 2018). Ein weiteres Problem stellen die verschiedenen

Formen des Bias dar – Bias des Algorithmus, Bias des Samples, auf Vorurteilen beruhender Bias, Mess-Bias (s. 3.7.3) – die die Aussagekraft der Ergebnisse von prädiktiven Analysen einschränken bzw. die zu falschen Ergebnissen führen können.

Aus diesen Beschränkungen maschinellen Lernens und der ihnen zugrundeliegenden Datensätze und Algorithmen folgt,

“that it is important to have ‘humans in the loop’ in AI systems, even as they become more and more advanced and accurate. Humans are needed to check the quality of the data as well as interpret and contextualize the results” (Ferryman 2019, 1).

Diese Forderung kann keinesfalls nur an Ärztinnen und Ärzte gerichtet sein. Sie betrifft auch Datenspezialistinnen bzw. Datenspezialisten und Systementwicklerinnen bzw. Systementwickler. Zu den vorgeschlagenen Ansätzen die Forderung „*to have humans in the loop*“ zu realisieren, zählen: das Beharren auf Transparenz (Einsicht darin zu geben, wie ein Modell zu Entscheidungen gelangt) oder, wo dies aufgrund der Komplexität der Modelle nicht möglich ist, alternative Formen der Begründbarkeit zur Verfügung zu stellen (z. B. kontrafaktische Erklärungen); die Stärkung von Methoden, „algorithmische Fairness“ zu gewährleisten; Forscherinnen und Forscher zu verpflichten, Genderanalysen durchzuführen, usw.

Insgesamt gilt es abzuwägen zwischen den Möglichkeiten neue Diagnosemöglichkeiten zu erhalten, die die Künstliche Intelligenz bietet und der für ärztliches Handeln fundamentalen dialogischen Komponente, die dem Begründen und Einsichtigmachen – „*telling a story*“ – großen Wert beimisst. Zarsky (2018) spricht von einer „Obsession mit Kausalität“ und kontrastiert diese beiden Ansätze folgendermaßen:

“Insisting on the development of mechanisms brings the risk of rejecting correlations that current science cannot prove or even explain and yet nonetheless have predictive value. Such a policy choice thus will tilt the balance of expected outcomes in favor of existing knowledge

versus

[...] data must be explained through the telling of a ‘story’ [...] to prevent the analyst from mistaking noise for a signal, an error that would delude others as well” (Zarsky 2018, 50)

Aus ethischer Sicht lässt sich zusammenfassen, dass mit dem Einsatz von *Big Data*, Künstlicher Intelligenz und maschinellem Lernen vor allem das Problem der Verantwortung vor eine neue Dimension gestellt wird. Verantwortung kann aus ethischer Sicht entweder nur dreistrahlig (WER, WAS, und WOVOR?) oder aber sechsstrahlig (WER, WAS, WOVOR, WESWEGEN, WOFÜR und WANN?) interpretiert werden (Ropohl 1987, 155).

## 5.3 Soziale Gerechtigkeit – die Unvermeidbarkeit distributiver Effekte?

*Big Data* läuft Gefahr, Daten über spezifische gesellschaftliche Randgruppen – Migrantinnen und Migranten, Personen mit niedrigem ökonomischem Status, usw. – nicht systematisch zu erfassen. Die nicht ausreichende Repräsentanz dieser Personengruppen in den Datensätzen hängt damit zusammen, dass diese weniger oft die Ärztin bzw. den Arzt besuchen, nicht oder weniger häufig im Internet nach Gesundheitsinformation suchen, keine Kreditkarte oder auch nicht jene „*wearables*“ besitzen, die kontinuierlich Daten über Lebensgewohnheiten und Gesundheitszustand registrieren. Sehr häufig sind dies auch diejenigen Personen, deren Situation vermehrter Forschung und Intervention bedürfte (z. B. Cohen et al. 2017). Darüber hinaus bedingt ein historisch tief verankerter Gender Bias in der Biomedizin – trotz aller Bemühungen, diesem entgegenwirken – dass genderspezifische Aspekte zahlreicher Erkrankungen in den Datensätzen nicht ausreichend repräsentiert sind.

Zu diesem Bias in den Datensätzen kommt hinzu, dass ein prädiktives, auf maschinellem Lernen basierendes Modell nicht neutral ist. Die Entwicklung eines solchen Modells beinhaltet Entscheidungen darüber, welche Probleme priorisiert, welche Algorithmen verwendet und in welcher Weise das Modell eingesetzt werden soll.

“A machine learning model is architected from the programmers that create it, the algorithm and metrics used, and the data it takes as input. When a development team programs a machine learning model they must choose carefully: what type of algorithm is used, how the algorithm is set up, what metrics and parameters are used, and on what data the algorithm is trained and tested. Creators’ influence can show up in unexpected ways” (Shadowen 2017, 9).

Solche Entscheidungen haben angesichts der Ressourcenknappheit im Gesundheitswesen unvermeidliche distributive Effekte. Diese kommen besonders in jenen Fällen zum Tragen, in denen das Ergebnis einer Analyse Nachteile für individuelle Patientinnen und Patienten mit sich bringen könnte, während es andere bevorteilt. Eine solche Entscheidung wäre dadurch gerechtfertigt, dass sie sich insgesamt positiv auf die Gesundheit einer spezifischen Patientenpopulation auswirkt. Ein in der Literatur zitiertes Beispiel ist das einer Ärztin bzw. eines Arztes, die bzw. der vor der Entscheidung steht eine Patientin bzw. einen Patienten mit moderaten organischen Dysfunktionen intensivmedizinisch betreuen zu lassen, während die prädiktive Analyse andere Patientinnen und Patienten als vorrangig für eine solche Betreuung identifiziert. Angesichts einer Knappheit von Intensivbetten ist dies eine Situation, in der einer Patientin bzw. einem Patienten, die bzw. der davon profitieren könnte, eine medizinische Leistung verweigert wird, weil es andere, von einem Algorithmus identifizierte Personen gibt, deren Zustand als risikoreicher bewertet wird:

“A different set of patients are admitted in a world where the analytics help guide the decision-making and they too may benefit, neither benefit nor be harmed, or be harmed. If the predictive analytics approach is to be used and be justified it must be because overall more patients benefit than under the status quo” (Cohen et al. 2014, 463).

Zwar treffen auch gegenwärtig Ärztinnen und Ärzte angesichts der Ressourcenknappheit immer wieder Entscheidungen mit möglichen distributiven Effekten. Dies wird jedoch möglicherweise durch die Möglichkeiten der prädiktiven Analyse sowie der in den Therapievorschlägen eingebauten „*choice architecture*“ systematisiert und verstärkt.

Da Modelle nicht neutral sind, kann auch die fehlende Diversität jener, die Künstliche Intelligenz Applikationen entwickeln ein Problem darstellen. Als Beispiel wird angeführt, dass Apple ein umfassendes Tool zur Erfassung des Gesundheitszustandes entwickelte, in das bis 2015 die Menstruationszyklen nicht erfasst wurden – „*which may not be surprising considering the engineering team is predominately male*“ (Ferryman 2019, 1).

## 5.4 Neue Anforderungen an Ärztinnen und Ärzte

Die Adoption von *Information Infrastructures in Healthcare* (IIH), verstärkt durch die Verfügbarkeit von „*Big Data*“, führt dazu, dass das Gesundheitspersonal (teilweise auch die Patientinnen und Patienten) zunehmend Aufgaben, die sich unter dem Begriff „*Data Work*“ zusammenfassen lassen, übernehmen. Immer mehr Personen verwenden mobile Apps, die gesundheitsrelevante Daten für die persönliche Verwendung, aber auch für professionelle Zwecke erfassen. Ärztinnen und Ärzte sind dazu angehalten, unterschiedliche Systeme mit Daten zu bedienen, die nicht nur die Interaktionen mit individuellen Patientinnen und Patienten dokumentieren, sondern für unterschiedliche sekundäre Zwecke, also auch „*Big Data*“, bereitgestellt werden müssen (Pine et al. 2018).

Außerdem sollten Ärztinnen und Ärzte imstande sein, die Ergebnisse von prädiktiven Analysen für ihre Entscheidungen, in Kooperation mit Kolleginnen bzw. Kollegen und Angehörigen anderer Gesundheitsberufe, sowie in Gesprächen mit Patientinnen und Patienten zu interpretieren, gegenüber anderen Informationsquellen abzuwägen und umzusetzen.

Die positiven Seiten des Zur-Verfügung-Stehens prädiktiver Analysen können sich nur entfalten, wenn Gesundheitsorganisationen Ärztinnen und Ärzten Unterstützung und Zeit für Reflektion und kritische Analyse bieten. Zusätzlich wurde vorgeschlagen, neue Berufsgruppen zu schaffen, deren primäre Aufgabe es ist, Ärztinnen und Ärzte in der Analyse und Interpretation von Gesundheitsdaten zu unterstützen. Fiske et al. (2019) definieren das Anforderungsprofil solcher „*Health Information Counselors*“ (HIC) folgendermaßen:

“HICs would have broad knowledge of various kinds of health data and data quality evaluation techniques, as well as analytic skills in statistics and data interpretation. Trained also in interpersonal communication, health management, insurance systems, and medico-legal aspects of data privacy, HICs would know enough about clinical medicine to advise on the relevance of any kind of data for prevention, diagnosis, and treatment.” (37)

Darüber hinaus sollten die neuen Anforderungen, die sich mit dem Einsatz von prädiktiven Analysesystemen stellen auch in die medizinische Aus- und Weiterbildung aufgenommen werden. Beispielsweise fragen Cohen et al. (2017):

“Is the current state of medical education adequate to make physicians (as well as nurses, hospital administrators, etc.) wise users of predictive analytics? Medical education is already densely packed with a myriad of kinds of learning, but data science has traditionally not been a focus. Would widespread adoption of predictive analytics be met with widespread improvements in data science education in medical school, or would (and should?) it become a specialized set of learning for a subset of physicians with others just told to ‘trust the algorithm’?” (451).

## 6 Schlussfolgerungen

Durch die Einbettung digitaler Technologien in den unmittelbaren Kontakt zwischen Arzt und Patienten verändern sich die wechselseitigen Rollen. Auf der einen Seite machen digitale Technologien die Patientin bzw. den Patienten viel stärker zur aktiven und selbstbestimmten Partnerin bzw. Partner im Prozess der Prävention, Diagnose und Therapie. Auf der anderen Seite gerät die Ärztin bzw. der Arzt immer stärker in die Rolle einer Vermittlerin bzw. eines Vermittlers zwischen der Technik und der Patientin bzw. dem Patienten. Beide Entwicklungen haben unmittelbare Auswirkungen auf die ärztliche Aus- und Fortbildung ebenso wie auf die Ausstattung medizinischer Einrichtungen und die Struktur des Gesundheitssystems, jeweils einschließlich der rechtlichen Rahmenbedingungen. Die Eigenart des Arzt-Patienten-Verhältnisses stellt aber auch erhöhte Anforderungen an die ethische Gestaltung digitaler Technologien („*Ethics by Design*“).

Auf der Grundlage der unter 1 bis 5 dargelegten Erwägungen unterbreitet die Bioethikkommission beim Bundeskanzleramt der Bundesregierung die folgenden zusammenfassenden Feststellungen und Empfehlungen. Dabei gelten die Ausführungen, die im Folgenden zum ärztlichen Handeln gemacht werden, analog auch für den Bereich der Pflege.



## 6.1 Verbesserung der medizinischen Versorgung durch Einsatz digitaler Technologien

Moderne digitale Technologien bieten im Sinne des Wohltunsprinzips die Chance für eine Verbesserung der medizinischen Versorgungsqualität. Allerdings ist bei ihrem Einsatz zu beachten, dass sich ärztliches Handeln im Gesundheitswesen nicht nur rein auf Daten beziehen darf, sondern wissens- und erfahrungsbasiert sein muss. Außerdem ist Digitalisierung kein Selbstzweck, sondern sollte von der Ambition geleitet sein, die ärztlichen Handlungsmöglichkeiten entscheidend zu verbessern. Ihr ethisch vertretbarer Einsatz setzt grundsätzlich eine solche Verbesserung voraus, d. h. die Richtigkeit der Befunderhebung, die Treffsicherheit der Diagnose, die Erfolgswahrscheinlichkeit der empfohlenen Therapie oder die Erfolgsquote bei einem medizinischen Eingriff müssen beim Einsatz der betreffenden Technologie grundsätzlich mindestens gleich gut und idealerweise besser sein als beim Einsatz herkömmlicher Technologien und rein menschlicher Akteure. Daraus folgt, dass ihr Einsatz bei erreichbarer Verbesserung des Therapieerfolgs nicht nur zulässig, sondern ethisch geboten ist. Voraussetzung ist der eindeutige Nachweis einer solchen Verbesserung des Therapieerfolgs durch entsprechende medizinische Evidenz. Ein weiteres, ethisch relevantes, Argument für den Einsatz und für die Schaffung der Grundlagen digitaler Technologien ist ein erzielbarer allgemeiner medizinischer Fortschritt zum Wohle künftiger Patientengenerationen, selbst wenn in einem solchen Fall keine Verbesserung des Therapieerfolgs für die betroffene individuelle Patientin bzw. den individuellen Patienten erreicht werden kann. Der Einsatz moderner digitaler Technologien ist grundsätzlich ethisch unzulässig, wenn der medizinische Qualitätsstandard herabgesetzt wird, selbst dann, wenn ein Gewinn an Effizienz entstünde. Digitalisierung darf keinesfalls in Konkurrenz zu den beiden Handlungsmaximen – dem Nichtschadensprinzip („*Do not harm*“) und dem Wohltunsprinzip („*Do good*“) treten, noch mit diesen abzuwägen sein.

Die Qualität moderner digitaler Technologien wird durch eine Vielzahl von Faktoren bestimmt. Im Falle von Tools, die auf Künstlicher Intelligenz und maschinellem Lernen beruhen, sind dies etwa die Qualität der Analyse- und Trainingsdaten, die adäquate Schulung des medizinischen Personals, sowie die Berücksichtigung der Rolle von ärztlichem Erfahrungswissen und des Dialogs mit der Patientin bzw. dem Patienten. Dabei ist bei der Bewertung der medizinischen Qualität stets das emotionale Wohlbefinden des Patienten und der Patientin sowie des Arztes und der Ärztin zu berücksichtigen. Dieses Wohlbefinden hängt wiederum von einer Reihe von Faktoren ab, wie dem Aufbau eines Vertrauensverhältnisses und einer gelungenen Kommunikation zwischen der Ärztin bzw. dem Arzt und der Patientin bzw. dem Patienten, aber auch von der Zeit, die für Gespräche und Nachfragen zur Verfügung steht. Diese Voraussetzungen sind beim Einsatz digitaler Technologien durch eine entsprechende Organisation zu berücksichtigen. Mögliche Zeitersparnisse sollten dafür genutzt werden, Vertrauen und emotionales Wohlbefinden zu stärken und keinesfalls „wegrationalisiert“ werden.

## 6.2 Distributive Effekte und Bias

Künstliche Intelligenz und maschinelles Lernen führen nicht „automatisch“ zu mehr sozialer Gerechtigkeit im Gesundheitssystem. Es zeigt sich zunehmend, dass algorithmische Systeme zu einer Verstärkung von Bias und Diskriminierung, wie etwa aufgrund von Gender, Alter, ethnischen Merkmalen oder sozioökonomischem Status, beitragen können. Deren Quellen zu erkennen, zu reflektieren und ihnen durch eine bias-sensitive Systemgestaltung und -nutzung entgegenzuwirken, stellt eine Herausforderung aber auch eine gesellschaftliche Chance dar.

## 6.3 Veränderungen der Arzt-Patienten-Beziehung

Durch den Einsatz von Maschinen hat sich das Zusammenwirken von Arzt und Maschine verändert. Während viele Entscheidungsprozesse in Richtung Maschine verlagert wurden, kamen ganz neue Aufgaben für die Ärztin bzw. den Arzt hinzu. Ihr Einfluss ist am höchsten, wenn die Maschine zwar Entscheidungsunterstützung bietet und die tatsächliche Entscheidung über Diagnose und / oder Therapie jedoch bei der Ärztin bzw. dem Arzt liegt. Ihr Einfluss ist am geringsten, wenn Ärztin bzw. Arzt allenfalls noch für den Fall von Auffälligkeiten einsatzbereit im Hintergrund stehen und die Maschine tendenziell autonom „handelt“. Dazwischen liegt ein breites Spektrum verschiedener Gestaltungen. Beim Einsatz von algorithmischen Systemen ist anzustreben, dass eine von der Maschine errechnete Diagnose oder ein Therapievorschlag alleine nicht zu einer faktischen Entscheidung führen darf. Ärztin bzw. Arzt sollten weiterhin dazu verpflichtet sein zu begründen, weshalb sie dem Vorschlag eines algorithmischen Systems folgen bzw. von diesem abweichen. Dies ist insbesondere im Falle einer hohen Systemkritikalität, also einem hohen Risiko einer möglichen Entscheidung, organisatorisch abzusichern.

Durch die Auslagerung von Funktionen auf Maschinen entstehen Freiräume für die Ärztin bzw. den Arzt. Es wäre ethisch bedenklich, diese Freiräume ausschließlich zur Effizienzsteigerung im Wege des Abbaus von medizinischem Personal zu nutzen. Vielmehr sollten diese Freiräume auch gezielt dazu eingesetzt werden, der einzelnen Patientin bzw. dem einzelnen Patienten mehr menschliche Zuwendung in der persönlichen Arzt-Patienten-Beziehung zukommen zu lassen. Neue digitale Technologien führen daher im Idealfall nicht zu einer Reduzierung des menschlichen Faktors in der Arzt-Patienten-Beziehung, sondern im Gegenteil zu dessen Verstärkung und der Konzentration auf das Wesentliche.

Eine Voraussetzung dafür, dass diese Potentiale für bessere persönliche Zuwendung und eine Verstärkung des menschlichen Faktors tatsächlich genutzt werden können, ist eine entsprechende Gestaltung des Gesundheitssystems, einschließlich der Honorar- und Vergütungsstrukturen. Gewonnene Zeit kann nur dann wirklich in die Beziehung zwischen Ärztin bzw. Arzt und Patientin bzw. Patient investiert werden, wenn eine Aufwertung der „sprechenden Medizin“ erfolgt und auch entsprechende finanzielle Anreize gesetzt werden.

Ärztinnen und Ärzte müssen in ihrer universitären Ausbildung und ihrer daran anschließenden praktischen Tätigkeit die nötigen Kenntnisse und Fertigkeiten erlernen um Entscheidungen betreffend die Einbindung algorithmischer Systeme in ihre täglichen patientenbezogenen Tätigkeiten umzusetzen. Dabei kommt die Ärztin bzw. der Arzt mit dem Einsatz moderner digitaler Technologien immer häufiger in eine neue Vermittlerrolle zwischen Patientin bzw. Patient und Technik, wobei es darum geht, den Einsatz von Maschinen offen zu legen und der Patientin bzw. dem Patienten die Kriterien, die zu einem Entscheidungsvorschlag der Maschine geführt haben, zu erläutern. Dabei muss die Ärztin bzw. der Arzt über Kenntnisse betreffend Datenqualität, Leistungsfähigkeit algorithmischer Berechnungen und Aussagekraft von Wahrscheinlichkeiten und anderen Ergebnissen verfügen, um erstens Berechnungsergebnisse selbst zutreffend einschätzen zu können und sie zweitens der Patientin bzw. dem Patienten auch vermitteln zu können.

Die Verfügbarkeit von qualitativ hochwertigen Daten ist eine grundlegende Bedingung für die medizinische Forschung und für eine Versorgung der Patientinnen und Patienten auf einem dem Stand der medizinischen Wissenschaften entsprechendem Niveau. Es sind alle rechtlichen, organisatorischen und technischen Voraussetzungen für die rechtssichere Sammlung, Kuratierung und Nutzung von Daten im intra- und extramuralen Gesundheitswesen zu schaffen. Entscheidend ist, dass Daten national und international gemeinsam genutzt werden können und der Forschung sowie der medizinischen Praxis zur Verfügung stehen. Bei diesen Zielen ist die Einhaltung des grundrechtlich gebotenen Datenschutzniveaus, insbesondere auch durch die weitest mögliche Synthetisierung oder zumindest Anonymisierung von Daten, sowie die Wahrung ethischer Standards sicherzustellen. Ein hinreichender Datenschutz ist speziell auch an den Schnittstellen zwischen Krankenanstalten, Forschungseinrichtungen, niedergelassenen Ärztinnen und Ärzten sowie Patientinnen und Patienten erforderlich. Dadurch soll auch die Akzeptanz und das Vertrauen der Patientinnen und Patienten in datengestützte Anwendungen gestärkt und dauerhaft gewährleistet werden.

Durch den Einsatz moderner digitaler Technologie ergeben sich eine Reihe von sonstigen ethischen Fragen, die als solche nicht neu sind, quantitativ oder qualitativ aber eine neue Dimension erlangen. Das betrifft etwa den Umgang mit Zufallsbefunden („*incidental findings*“), die quantitativ häufiger anfallen können, weil Maschinen eine viel größere Bandbreite an Informationen verarbeiten und auswerten können. Ärztinnen und Ärzte sind verpflichtet, sich mit dem Umgang mit Zufallsbefunden auseinanderzusetzen. Patientinnen und Patienten sind über die Auswirkungen von möglichen Zufallsbefunden zu informieren. Das Recht auf Nichtwissen muss jedenfalls sichergestellt werden.

## 6.4 Konsequenzen für Verantwortung und Systemdesign

Der ethisch vertretbare Einsatz von modernen digitalen Technologien setzt Erklärbarkeit und Transparenz der letztlich getroffenen medizinischen Entscheidungen voraus, wobei mit wachsender Systemkritikalität auch gesteigerte Anforderungen zu stellen

sind. Die Forderung nach Transparenz betrifft Datenquellen und Datenqualität; die für eine Entscheidung maßgeblichen Hauptparameter, wie etwa Alter, Geschlecht, Krankheitsgeschichte, Bildbefund, sowie den Umfang, in welchem bestimmte Annahmen auf Korrelationen oder Kausalitäten beruhen, gegebenenfalls kombiniert mit kontrafaktischen Erklärungen. Erklärbarkeit und Transparenz zu gewährleisten liegt vor allem in der Verantwortung der Systemgestalter und muss bei den Zulassungsverfahren algorithmischer Systeme Berücksichtigung finden und in den Anwenderinformationen enthalten sein.

Eine für die Sicherung der Datenqualität zentrale Maßnahme besteht darin, dem Datenset, das einem algorithmenbasierten System zugrunde liegt, Metadaten über dessen Konstruktion hinzuzufügen. Die Nachvollziehbarkeit der Kontextualität von Daten, die für einen bestimmten Zweck erhoben wurden, zu gewährleisten ist vor allem dann zu empfehlen, wenn diese zu Analyse Zwecken in anderen Kontexten eingesetzt werden sollen. Damit soll es möglich werden, einen eventuellen Bias des Datensets zu erkennen und zu beheben.

Der Einsatz von Künstlicher Intelligenz in der Medizin muss für alle Beteiligten – das sind insbesondere Patientinnen und Patienten, Ärztinnen und Ärzte sowie die Träger von Krankenanstalten – rechtssicher möglich sein. Rechtsvorschriften sollen – etwa in Gestalt eines beweglichen Systems – eine der jeweiligen Technik und ihrem Einsatzfeld adäquate Kontrolle vorsehen, die verschiedene Formen annehmen kann, von reiner ex post-Kontrolle über begleitende Audits bis hin zu Zulassungs- bzw. Zertifizierungserfordernissen.

Der Spitalsträger, die eine Ordination leitende Ärztin bzw. der eine Ordination leitende Arzt sollte für den – nicht bloß die Entscheidung eines Menschen unterstützenden – Einsatz eines „autonomen“ Systems in gleicher Weise Verantwortung übernehmen und zivilrechtlich haften wie wenn menschliches medizinisches Personal eingesetzt worden wäre. Der Einsatz moderner digitaler Technologien im unmittelbaren Kontakt mit Patientinnen und Patienten setzt aus ethischer Sicht voraus, dass kein Verantwortungsvakuum entsteht. Daher müssen die geltenden zivil- und strafrechtlichen Vorschriften daraufhin überprüft werden, ob sie auch bei Einsatz „autonomer“ und vernetzter Systeme noch eine angemessene Verteilung von Verantwortung und Haftung gewährleisten. Dazu gehört auch das Entstehen-Müssen für technische Hilfsmittel, welche menschliche Abwägungs- und Entscheidungsprozesse ersetzen, nach ähnlichen Grundsätzen, wie sie für menschliche Erfüllungsgehilfen gelten (vgl. § 1313a ABGB). Die Einführung einer „elektronischen Person“, also die Anerkennung von Rechtspersönlichkeit für Roboter und „Künstliche Intelligenz“, ist demgegenüber abzulehnen.

## 6.5 Konsequenzen für die ärztliche Ausbildung

Die Befähigung von Ärztinnen und Ärzten algorithmische Systeme in die täglichen patientenbezogenen Tätigkeiten einzubinden, soll auch in den Curricula der medizinischen Studien schwerpunktmäßig berücksichtigt werden. Bei der Gestaltung der Ausbildung ist zu berücksichtigen, dass der generelle Kompetenzerhalt der ärztlichen Profession durch

moderne digitale Technologien insofern infrage gestellt wird, als oft jahrzehntelange klinische Tätigkeit erforderlich ist, um zu einer erfahrenen Ärztin bzw. einem erfahrenen Arzt zu werden. Es ist in erster Linie eine Frage der ärztlichen Ausbildung, für entsprechende Kompetenzen auch künftiger Generationen von Ärztinnen und Ärzten zu sorgen.

Bei der Entscheidung über den Einsatz und die Gestaltung moderner digitaler Technologien sind Aspekte des speziellen und generellen ärztlichen Kompetenzerhalts im Sinne eines kompetenzsensitiven Designs stets zu berücksichtigen, die sich unmittelbar auf verbleibende menschliche Kontrollfunktionen (z. B. „*human in the loop*“-Prinzip) auswirken, aber ebenso auf die medizinische Versorgungsqualität in Ausnahmesituationen (z. B. im Katastrophenfall oder bei Cyberangriffen) sowie auf die Innovationskraft künftiger Generationen (z. B. Entwicklung neuer medizinischer Technologien). Der spezielle Kompetenzerhalt muss durch entsprechende Systemgestaltung sichergestellt werden. Dabei soll systemseitig ein Rollenwechsel eingeführt werden, d. h. der Ärztin bzw. dem Arzt immer wieder auch die Erstentscheidung ohne Kenntnis des algorithmischen Entscheidungsvorschlags zugewiesen werden. Es bedarf der Kompensation durch kompetenzsensitives Design unter bewusstem Einsatz zwingend zu absolvierender Trainingsmodule und unter Ausnutzung von Chancen für die medizinische Fortbildung, welche durch die weltweite Vernetzung und Verfügbarkeit von Fallstudien und Bildmaterial entstehen.

Grundlegende Kompetenzen im Datenrecht (v. a. Datenschutzrecht) und in der Datenethik müssen künftig fester Bestandteil der Ausbildung aller medizinischer Berufe sein. Zudem aber entsteht ein Bedarf an der Vermittlung ganz neuer Kompetenzen, die bislang nicht oder kaum Teil der ärztlichen (und pflegerischen) Ausbildung gewesen sind. Diese Kompetenzen betreffen ein grundlegendes Verständnis der Funktionsweise moderner digitaler Technologien, insbesondere der Bedeutung von Datenqualität, der Gefahren von Bias und Diskriminierung, der Aussagekraft algorithmischer Berechnungsergebnisse und allgemein der Grenzen algorithmischer Systeme.

## 6.6 Empfehlungen

1. Moderne digitale Technologien bieten im Sinne des Wohltunsprinzips die Chance für eine Verbesserung der medizinischen Versorgungsqualität. Ihr Einsatz ist bei erreichbarer Verbesserung des Therapieerfolges nicht nur zulässig, sondern ethisch geboten.
2. Das emotionale Wohlbefinden der Patientin und des Patienten sowie der Ärztin und des Arztes ist bei der Bewertung der medizinischen Qualität stets zu berücksichtigen.
3. Auch der allgemeine medizinische Fortschritt zum Wohle künftiger Patientengenerationen stellt ein ethisch relevantes Argument für den Einsatz digitaler Technologien dar.
4. Künstliche Intelligenz und maschinelles Lernen führen nicht automatisch zu mehr sozialer Gerechtigkeit im Gesundheitssystem. Mögliche Quellen von Bias und Diskriminierung zu erkennen, zu reflektieren und ihnen durch eine bias-sensitive

Systemgestaltung und -nutzung entgegenzuwirken, stellt eine Herausforderung aber auch eine gesellschaftliche Chance dar.

5. Durch den Einsatz von Maschinen hat sich das Zusammenwirken von Arzt und Maschine verändert. Beim Einsatz von algorithmischen Systemen ist anzustreben, dass eine von der Maschine errechnete Diagnose oder ein Therapievorschlag alleine nicht zu einer faktischen Entscheidung führen darf.
6. Durch die Auslagerung von Funktionen auf Maschinen entstehen Freiräume für die Ärztin bzw. den Arzt. Diese Freiräume sollten auch gezielt dazu eingesetzt werden, der einzelnen Patientin bzw. dem einzelnen Patienten mehr menschliche Zuwendung in der persönlichen Arzt-Patienten-Beziehung zukommen zu lassen.
7. Gewonnene Zeit kann nur dann wirklich in die Beziehung zwischen Ärztin bzw. Arzt und Patientin bzw. Patient investiert werden, wenn eine Aufwertung der „sprechenden Medizin“ erfolgt und auch entsprechende finanzielle Anreize gesetzt werden.
8. Ärztinnen und Ärzte müssen in ihrer universitären Ausbildung und ihrer daran anschließenden praktischen Tätigkeit die nötigen Kenntnisse und Fertigkeiten erlernen um Entscheidungen betreffend die Einbindung algorithmischer Systeme in ihre täglichen patientenbezogenen Tätigkeiten umzusetzen.
9. Es sind alle rechtlichen, organisatorischen und technischen Voraussetzungen für die rechtssichere Sammlung, Kuratierung und Nutzung von Daten im intra- und extramuralen Gesundheitswesen zu schaffen.
10. Der ethisch vertretbare Einsatz von modernen digitalen Technologien setzt Erklärbarkeit und Transparenz der letztlich getroffenen medizinischen Entscheidungen voraus.
11. Der Einsatz von Künstlicher Intelligenz in der Medizin muss für alle Beteiligten – das sind insbesondere Patientinnen und Patienten, Ärztinnen und Ärzte sowie die Träger von Krankenanstalten – rechtssicher möglich sein.
12. Der Spitalsträger oder der eine Ordination leitende Arzt sollte für den – nicht bloß die Entscheidung eines Menschen unterstützenden – Einsatz eines „autonomen“ Systems in gleicher Weise Verantwortung übernehmen und zivilrechtlich haften wie wenn menschliches medizinisches Personal eingesetzt worden wäre. Dies soll garantieren, dass kein Verantwortungsvakuum entsteht.
13. Die Befähigung von Ärztinnen und Ärzten algorithmische Systeme in die täglichen patientenbezogenen Tätigkeiten einzubinden, soll auch in den Curricula der medizinischen Studien schwerpunktmäßig berücksichtigt werden.
14. Bei der Entscheidung über den Einsatz und die Gestaltung moderner digitaler Technologien sind Aspekte des speziellen und generellen ärztlichen Kompetenzerhalts im Sinne eines kompetenzsensitiven Designs stets zu berücksichtigen.
15. Grundlegende Kompetenzen im Datenrecht (v. a. Datenschutzrecht) und in der Datenethik müssen künftig fester Bestandteil der Ausbildung aller medizinischer Berufe sein.

# 7 Literatur

## Monographien und Artikel

Aichberger S, Thurner S (2016). Disentangling genetic and environmental risk factors for individual diseases from multiplex comorbidity networks. *Scientific reports* 6, 39658.

Anderson C (2008). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired Magazine* 16.07, 2.

Antes G (2016). "Is the age of causality over?". *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen* 112, 16–22.

Atluri G et al. (2009). Association Analysis Techniques for Bioinformatics Problems. In: Rajasekaran Sanguthevar (Ed.): *Bioinformatics and Computational Biology, lecture Notes in Bioinformatics* 5462, Springer Berlin Heidelberg, 1–13.

Balint M (1964). *Der Arzt, sein Patient und die Krankheit*; Stuttgart, Klett.

Barsky AJ, Peekna HM, Borus JF (2001). Somatic symptom reporting in women and men. *Journal of general internal medicine* 16, 266–275.

Bartholomew K, Horowitz LM (1991). Attachment styles among young adults: Test of a four category model. *Journal of Personality and Social Psychology* 61, 226–244.

Barton N, Hermisson J, Nordberg M (2019). Why structure matters. *eLIFE* 8, e45380.

Beckman HB, Frankel RM (1984). The Effect of Physician Behavior on the Collection of Data. *Annals of Internal Medicine* 101 (05), 692–696, <https://doi.org/10.7326/0003-4819-101-5-692>.

Beggiato A et al. (2017). Gender differences in autism spectrum disorders: divergence among specific core symptoms. *Autism Research* 10, 680–689.

Berendas BS, Waldstein SM, Schmid-Erfurth U (2018). Screening und Management retinaler Erkrankungen mittels digitaler Medizin. *Der Ophthalmologe* 115 (9), 728–736.

Bioethikkommission (2008). Empfehlungen mit Genderbezug für Ethikkommissionen und klinische Studien.

Birch K (2017). Rethinking value in the bio-economy: Finance, assetization, and the management of value. *Science, Technology, & Human Values* 42/3, 460–490.

Birklbauer A (2019). Die Bedeutung von (medizinischen) Leitlinien im Strafrecht. *Journal für Medizin und Gesundheitsrecht* 1/2019, 16 ff.

Bonham VL, Green ED, Pérez-Stable EJ (2018). Examining how race, ethnicity, and ancestry data are used in biomedical research. *JAMA* 320 (15), 1533–1534.

Borck C (2016). *Medizinphilosophie*. Hamburg, Junius Verlag.

Bowlby J (2003). *Bindung und menschliche Entwicklung: John Bowlby, Mary Ainsworth und die Grundlagen der Bindungstheorie*. Stuttgart, Klett-Cotta.

Calude CS, Longo G (2017). The deluge of spurious correlations in big data. *Foundations of science* 22 (3), 595–612.

Caruana R et al. (2015). Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-day Readmission. *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1721–1730.

Chin Yee B, Upshur RE (2018). Clinical judgement in the era of big data and predictive analytics. *Journal of Evaluation in Clinical Practice*, 635–637, <https://doi.org/10.1111/jep.12852>.

Citron DK, Pasquale FA (2014). The Scored Society: Due Process for Automated Predictions. *Washington Law Review* 89, 1–32, <https://doi.org/10.1042/CS20180436>.

Cohen IG et al. (2014). The legal and ethical concerns that arise from using complex predictive analytics in health care. *Health affairs* 33/7, 1139–1147.

Cohen IG, Graver HS (2017). Cops, docs, and code: A dialogue between big data in health care and predictive policing. *UC Davis Law Review* 51, 437–474.

Comfort N (2018). Genetic determinism redux. *Nature* 561, 461–463.

Cooper RS, Nadkarni GN, Ogedegbe G (2018). Race, ancestry, and reporting in medical journals. *JAMA* 320 (15), 1531–1532.

Dang BN et al. (2017). Building trust and rapport early in the new doctor-patient relationship: a longitudinal qualitative study. *BMC Med Educ* 32, <https://doi.org/10.1186/s12909-017-0868-5>.



Datenethikkommission der deutschen Bundesregierung (2019). Gutachten der Datenethikkommission. Berlin, o.V.

Deo RC (2015). Machine learning in medicine. *Circulation* 132, 1920–1930.

Dörn S (2016). Programmieren für Ingenieure und Naturwissenschaftler: Grundlagen. Berlin, Vieweg.

Editorial (2017). Clinical decision making: more than just an algorithm. *The Lancet Oncology* 18/12, 1553.

Europäisches Parlament (2017). Entschließung mit Empfehlungen an die Kommission zu zivilrechtlichen Regelungen im Bereich Robotik. Empfehlung Nr. 59f.

European Parliament (2020). Draft Report with recommendations to the Commission on a Civil liability regime for artificial intelligence. 2020/2014(INL), [https://www.europarl.europa.eu/doceo/document/JURI-PR-650556\\_EN.pdf](https://www.europarl.europa.eu/doceo/document/JURI-PR-650556_EN.pdf).

Epstein D, ProPublica (2017). When evidence says no, but doctors say yes. *Atlantic* (1).

Esteva A et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542 (7639), 115–118.

European Society of Radiology (ESR) (2019). What the radiologist should know about artificial intelligence – an ESR white paper. *European Society of Radiology Insights into Imaging*, <https://doi.org/10.1186/s13244-019-0738-2>.

Expert Group on Liability and New Technologies (2019). Report on liability for Artificial Intelligence and other Emerging Digital Technologies. European Union.

Fagerhaugh SY et al. (1987). Hazards in hospital care: ensuring patient safety. San Francisco, Jossey-Bass Publishers.

Ferryman K, Pitcan M (2018). Fairness in precision medicine. *Data & Society*.

Fischer F (2017). Ethische Aspekte von E-Health aus der Perspektive von Public Health. In: Müller-Mielitz S, Lux T (eds) *E-Health-Ökonomie*, 141–151, [https://doi.org/10.1007/978-3-658-10788-8\\_9](https://doi.org/10.1007/978-3-658-10788-8_9).

Fiske A, Buyx A, Prainsack B (2019). Health Information Counselors: A New Profession for the Age of Big Data. *Academic Medicine* 94/1, 37–41.

FitzGerald C, Hurst S (2017). Implicit bias in healthcare professionals: a systematic review. *BMC medical ethics* 18 (1), 19.

Forbes Insights Team (2019). *AI And Healthcare: A Giant Opportunity*. 11 Feb 19.

Foucault M (1973). *Die Geburt der Klinik*. München, Carl Hanser Verlag.

Gallagher MD, Chen-Plotkin AS (2018). The Post-GWAS Era: From Association to Function. *The American Journal of Human Genetics* 102, 717–730.

Gandomi A, Haider M (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management* 35.2, 137–144.

Gerendas BS, Waldstein SM, Schmidt-Erfurth U (2018). Screening und Management retinaler Erkrankungen mittels digitaler Medizin. *Der Ophthalmologe* 115(9), 728–736.

Gianfrancesco MA et al. (2019). Potential biases in machine learning algorithms using electronic health record data. *JAMA internal medicine* 178 (11), 1544–1547.

Greenwald AG, Hamilton Krieger L (2006). Implicit bias: Scientific foundations. *California Law Review* 94/4, 945–967.

Gründinger W et al. (2019). *Mensch Moral Maschine – Digitale Ethik, Algorithmen und künstliche Intelligenz*. Berlin, Bundesverband Digitale Wirtschaft (BVDW) e.v., 12.

Gundling F et al. (2019). Defizite in der Gesundheitskompetenz stationär behandelter Patienten – eine Querschnittstudie. *Deutsche Medizinische Wochenschrift* 144 (04), 21–29.

Hammarlund N (2018). Racial Treatment Disparities after Machine Learning Surgical-Appropriateness Adjustment. *SSRN Electronic Journal* 3057607.

Hecker D et al. (2017). Künstliche Intelligenz und die Potentiale des maschinellen Lernens für die Industrie. *Wirtschaftsinformatik & Management* 9 (5), 26–35.

Hehner S, Biesdorf S, Möller M (2018). *Digitalisierung Gesundheitswesen: die Chancen für Deutschland*. Digital McKinsey.

Hehner S, Liese K et al. (2018). Digitalisierung in deutschen Krankenhäusern – Eine Chance mit Milliardenpotenzial für das Gesundheitssystem. *McKinsey & Company*.

Hipp R, Schlude C, Göller N (2017). Healthcare of the Future – The digital revolution of the healthcare sector. Porsche Consulting.

Hipp R et al. (2018). Patient im Fokus – Investitionschancen in das digitale Gesundheitswesen. Porsche Consulting.

Hoeyer K (2016). Denmark at a crossroad? Intensified data sourcing in a research radical country. The ethics of Biomedical Big Data, Springer, Cham, 73–93.

Horton RH, Lucassen AM (2019). Recent developments in genetic/genomic medicine. Clinical Science 133 (5), 697–708.

Jackson G (2019). The female problem: how male bias in medical trials ruined women's health. The Guardian 13 Nov 2019, <https://www.theguardian.com/lifeandstyle/2019/nov/13/the-female-problem-male-bias-in-medical-trials>.

Jannes M et al. (2018). Algorithmen in der digitalen Gesundheitsversorgung - Eine interdisziplinäre Analyse; Gütersloh, im Auftrag der Bertelsmann Stiftung.

Kaltenbach T, Erharter M et al. (2016). Digital & Disrupted - All change for healthcare. Roland Berger GmbH.

Kautzky-Willer A, Thurner S, Klimek P (2017). Use of statins offsets insulin-related cancer risk. Journal of internal medicine 281 (2), 206–216.

Kleinberg S, Hripcsak G (2011). A review of causal inference for biomedical informatics. Journal of Biomedical Informatics 44(6), 1102–1112.

Klimek P et al. (2015). Quantification of diabetes comorbidity risks across life using nation-wide big claims data. PLOS Computational Biology 11 (4), e1004125.

Klimek P, Aichberger S, Thurner S (2016). Disentangling genetic and environmental risk factors for individual diseases from multiplex comorbidity networks. Scientific reports 6, 39658.

Kopetzki C (2008). „Off-label-use“ von Arzneimitteln. In: Ennöckl D, Raschauer N, Schulev-Steindl E, Wessely W (Hrsg.), Über Struktur und Vielfalt im Öffentlichen Recht. Wien/ New York, Springer, 73 ff.

Krimer F et al. (2019). Morphometric MRI profiles of multiple system atrophy variants and implications for differential diagnosis. Movement Disorders 34(7), 1041–1048, <https://doi.org/10.1002/mds.27669>.

Laney D (2001). 3D data management: Controlling data volume, velocity and variety. META Group.

Leonelli S (2016). Data-centric biology: A philosophical study. University of Chicago Press.

Looker AC et al. (1997). Prevalence of low femoral bone density in older U.S. adults from NHANES III. *Journal of Bone and Mineral Research* 12 (11), 1761–1768, <https://doi.org/10.1359/jbmr.1997.12.11.1761>.

Lown B, Rodriguez D (2012). Commentary: Lost in translation? How electronic health records structure communication, relationships, and meaning. *Academic Medicine* 87 (4), 392–394.

Mayrhofer (2014). Off-label-use von Analgetika in der perioperativen Kinderschmerztherapie aus rechtlicher Sicht. *Der Schmerz*, 65–66.

Manyika J et al. (2011). Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute.

Mitchell A, Vaze A, Rao S (2009). Clinical diagnosis of depression in primary care: A meta analysis. *The Lancet* 374 (9690), 609–619.

Müller-Mielitz S (2017). E-Health-Ökonomie – Begriff und Abgrenzung. in: Müller-Mielitz S, Lux T, E-Health-Ökonomie, Wiesbaden, Springer Gabler, 35–49.

Nanji KC et al. (2018). Medication-related clinical decision support alert overrides in inpatients. *Journal of the American Medical Informatics Association* 25 (5), 476–481.

National Research Council (2011). Toward precision medicine: Building a knowledge network for biomedical research and a new taxonomy of disease. Washington (DC), National Academies Press.

Need AC, Goldstein DB (2009). Next generation disparities in human genomics: concerns and remedies. *Trends in Genetics* 25 (11), 489–494.

Obermeyer Z, Emanuel EJ (2016). Predicting the future – big data, machine learning, and clinical medicine. *The New England journal of medicine* 375 (13), 1216.

Obermeyer Z, Lee TH (2017). Lost in Thought – The Limits of the Human Mind and the Future of Medicine. *New England Journal of Medicine* 377 (13), 1209–1211, <https://doi.org/10.1056/NEJMp1705348>.

Open Letter to the European Commission, Artificial Intelligence And Robotics (2018), <http://www.robotics-openletter.eu/>.

Osoba O, William W IV (2017). An intelligence in our image: The risks of bias and errors in artificial intelligence. Santa Monica (CA), RAND Corporation.

Parasuraman R, Sheridan TB, Wickens CD (2000). A model of types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics – Part A* 30 (3), 286–297.

Parikh RB, Obermeyer Z, Navathe AS (2019). Regulation of predictive analytics in medicine. *Science* 363 (6429), 810–812.

Pearl J (2019). The Seven Tools of Causal Inference, with Reflections on Machine Learning. *Communications of the ACM* 62 (3), 54–60.

Pine K et al. (2018). Data Work in Healthcare: Challenges for Patients, Clinicians and Administrators. Conference Paper, 433–439, <https://doi.org/10.1145/3272973.3273017>.

Pot M, Spahl W, Prainsack B (2019). The Gender of Biomedical Data: Challenges for Personalised and Precision Medicine. *Soma* 9:2-3, 170–187.

Prainsack B (2015). Is personalized medicine different? (Reinscription: the sequel) A response to Troy Duster. *The British journal of sociology* 66 (1), 28–35.

Prainsack B (2019). Precision Medicine Needs a Cure for Inequality. *Current History* 118 (804), 11–16.

Rajkomar A et al. (2018). Scalable and accurate deep learning with electronic health records. *NPJ Digital Medicine* 1, 18–28.

Rieder G, Simon J (2016). Datatrust: Or, the political quest for numerical evidence and the epistemologies of Big Data. *Big Data & Society* Vol 3, issue 1, <https://doi.org/10.1177/2053951716649398>.

Rimscha M (2014). *Algorithmen kompakt und verständlich: Lösungsstrategien am Computer*. Wiesbaden, Springer Vieweg.

Ropohl G (1987). Neue Wege, die Technik zu verantworten. In: Lenk H, Ropohl G: *Technik und Ethik*, Stuttgart, Reclam, 149–176.

Rosenberg NA et al. (2019). Interpreting polygenic scores, polygenic adaptation, and human phenotypic differences. *Evolution, Medicine, and Public Health* (1), 26–34, <https://doi.org/10.1093/emph/eoy036>.

Russell C et al (2017). When Worlds Collide: Integrating Different Counterfactual Assumptions in Fairness. *Advances in Neural Information Processing Systems* 30 (NIPS).

Samulowitz A et al. (2018). "Brave men" and "emotional women": a theory-guided literature review on gender bias in health care and gendered norms towards patients with chronic pain. *Pain Research and Management*, 14.

Sauter SK et al. (2014). Analyzing Healthcare provider Centric networks through secondary use of health chains data. *IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, 522–525.

Scherfler C (2018). Automatisierte Magnetresonanztomographie Analyse bei Parkinson-Syndromen. *Newsletter der Österreichischen Parkinson Gesellschaft* (1).

Scherfler C et al. (2016). Diagnostic potential of automated subcortical volume segmentation in atypical parkinsonism. *Neurology* 86 (13), 1242–1249.

Sennaar K (2019). Machine Learning for Medical Diagnostics – 4 Current Applications. *Emerj Artificial Intelligence Research*.

Shadowen AN (2017). Ethics and Bias in Machine Learning: A Technical Study of What Makes Us "Good". New York, John Jay College of Criminal Justice, Student Thesis.

Strauss A et al. (1985). *Social Organization of Medical Work*. Chicago, The University of Chicago Press.

Street RL Jr, Gordon H, Haidet P (2007). Physicians' communication and perceptions of patients: Is it how they look, how they talk, or is it just the doctor?. *Soc Sci Med* 65, 586–98.

Street RL Jr et al. (2009). How does communication heal? Pathways linking clinician–patient communication to health outcomes. *Patient education and counseling* 74 (3), 295–301.

Street RL Jr, Haidet P (2011). How well do doctors know their patients? Factors affecting physician understanding of patients' health beliefs. *J Gen Intern Med* (1), 21–27.

Sturmberg JP (2018). Embracing complexity in health and health care – Translating a way of thinking into a way of acting. *Journal of Evaluation in Clinical Practice* 24 (3), 598–599.

Sundermann EE et al. (2017). Does the female advantage in verbal memory contribute to underestimating Alzheimer's disease pathology in women versus men?. *Journal of Alzheimer's Disease* 56 (3), 947–957.

Taylor K (2015). *Connected Health: How digital technology is transforming health and social care*. Deloitte Centre for Health Solutions.

Thaler RH, Sunstein CR (2008). *Nudge: improving decisions about health, wealth, and happiness*. New Haven & London, Yale University Press.

Topol E (2019), *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*. New York, Basic Books.

Torkamani A, Weininger NE, Topol EJ (2018). The personal and clinical utility of polygenic risk scores. *Nature Reviews Genetics* 19 (9), 581–590.

Tschandl P, Rosendahl C, Kittler H (2018). The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific data* (5), 180161.

Tschandl P et al. (2019). Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: an open, web-based, international, diagnostic study. *The Lancet Oncology* 20 (7), 938–947.

Tversky A, Kahneman D (1974). Judgment under uncertainty: Heuristics and biases. *Science* (185) 4157, 1124–1131.

Ubel PA, Scherr KA, Fagerlin A (2017). Empowerment failure: How shortcomings in physician communication unwittingly undermine patient autonomy. *The American Journal of Bioethics* 17 (11), 31–39.

Überall M, Werner-Felmayer G (2019). Integrative Biology and Big-Data-Centrism: Mapping out a Bioscience Ethics Perspective with a S.W.O.T. Matrix. *OMICS: A Journal of Integrative Biology* 23(8), 371–379.

Vayena E et al. (2018). Policy implications of big data in the health sector. *Bulletin of the World Health Organization* 96 (1), 66–68.

Vellido A (2019). Societal issues concerning the application of artificial intelligence in medicine. *Kidney Diseases* 5 (1), 27–33.

Vellido A et al. (2018). Machine learning in critical care: state-of-the-art and a sepsis case study. *BioMed Eng OnLine* 17(Suppl 1), 135, <https://doi.org/10.1186/s12938-018-0569-2>.

Ventres W et al. (2006). Physicians, patients, and the electronic health record: an ethnographic analysis. *The Annals of Family Medicine* 4 (2), 124–131.

Vogl WD et al. (2019). Automatic segmentation and classification of breast lesions through identification of informative multiparametric PET/MRI features. *European Radiology Experimental* 3 (1), 18.

Wachter S, Mittelstadt B, Russell C (2018). Counterfactual explanations without opening the black box: automated decisions and the GDPR. *Harvard Journal of Law & Technology* 31 (2), 841–887.

Wickens CD et al. (2010). Stages and levels of automation: An integrated meta-analysis. *Human Factors and Ergonomics Society Annual Meeting Proceedings* 54 (4), 389–393.

Wilm S et al. (2004). At Which Point Does the General Practitioner Interrupt his Patients at the Beginning of a Consultation?. *Z Allg Med* 80, 53–57. <http://doi.org/10.1055/s-2004-44933>.

Wirtz V, Cribb A, Barber N (2006). Patient-doctor decision-making about treatment within the consultation - A critical analysis of models. *Social science & medicine* 62.1, 116–124.

Wooden B et al. (2017). Using big data to discover diagnostics and therapeutics for gastrointestinal and liver diseases. *Gastroenterology* 152 (1), 53–67.

Yudell M et al. (2016). Taking race out of human genetics. *Science* 351, 564–565.

Zarsky T (2018). Correlation versus Causation in Health-Related Big Data Analysis. In Cohen I, Lynch H, Vayena E, Gasser U (Eds.), *Big Data, Health Law, and Bioethics*, Cambridge, Cambridge University Press 42–55.

Zentrale Ethikkommission bei der Bundesärztekammer (2013). Stellungnahme: Ärztliches Handeln zwischen Berufsethos und Ökonomisierung. Das Beispiel der Verträge mit leitenden Klinikärzten und -ärztinnen. *Deutsches Ärzteblatt* 110 (38).



Zimmermann-Rittereiser M, Schaper H (2016). Big Data – An Efficiency Boost in the Healthcare Sector. In: Big Data in Medical Science and Healthcare Management - Diagnosis, Therapy, Side Effects (11), 131–138.

## Websites

Gartner IT-Glossary: Big Data – [www.gartner.com/en/information-technology/glossary/big-data](http://www.gartner.com/en/information-technology/glossary/big-data)

Price L. The Digital Health Hype Cycle – [www.healthcare.digital/single-post/2019/01/12/The-Digital-Health-Hype-Cycle-2019](http://www.healthcare.digital/single-post/2019/01/12/The-Digital-Health-Hype-Cycle-2019)

# Abkürzungen

AAL	Ambient Assisted Living
ABGB	Allgemeines bürgerliches Gesetzbuch
ADS	Aufmerksamkeitsdefizitsyndrom
Art.	Artikel
ÄrzteG	Ärztegesetz
bzw.	beziehungsweise
CDL-AM	Christian Doppler Laboratory for Applied Metabolomics
d. h.	das heißt
DSG	Datenschutzgesetz
DSGVO	Datenschutzgrundverordnung
E-	Electronic
E-Health	Electronic Health
EG	Europäische Gemeinschaft
EGE	European Group on Ethics in Science and New Technologies
EHR	Electronic health record
EKG	Elektrokardiogramm
ELGA	Elektronische Gesundheitsakte
ESR	European Society of Radiology
et al.	und andere
EU	Europäische Union
EWG	Europäische Wirtschaftsgemeinschaft
FOG	Forschungsorganisationsgesetz
GTelG	Gesundheitstelematikgesetz
GWAS	Genome-wide association studies
HIC	Health Information Counselors
HNO	Hals-Nasen-Ohren-Heilkunde
i. d. S.	in diesem Sinne
IIH	Information Infrastructures in Healthcare
IKT	Informations- und Kommunikationstechnologien
IT	Informationstechnologie
KI / AI	Künstliche Intelligenz / Artificial Intelligence
KIS	Krankenhaus-Informations-System
MDR	Medical Device Regulation
MPG	Medizinproduktegesetz
MPVO	Medizinprodukteverordnung
MUW	Medizinischen Universität Wien
NEC	National Ethics Councils
NIST	National Institute of Standards and Technology

Nr.	Nummer
OCT	optische Kohärenztomographie
PET / MRI	Positronen-Emissions-Tomografie / Magnetic Resonance Imaging
PHG	Produkthaftungsgesetz
PSG	Produktsicherheitsgesetz
RFID	Radio Frequency Identification
S.	Seite
s.	siehe
s. u.	siehe unten
SNPs	Single Nucleotide Polymorphisms
sog.	sogenannte
UNESCO	United Nations Educational, Scientific and Cultural Organization
USA	United States of America
v. a.	vor allem
VbVG	Verbandsverantwortlichkeitsgesetz
vgl.	vergleiche
VO	Verordnung
Z	Ziffer
z. B. / e. g.	zum Beispiel
z. T.	zum Teil

## Mitglieder der Bioethikkommission für das Mandat 2017 bis 2020

### **Vorsitzende**

Dr. Christiane Druml

### **Stv. Vorsitzender**

Univ.-Prof. Mag. Dr. Markus Hengstschläger

### **Stv. Vorsitzender**

Univ.-Prof. Dr. h.c. Dr. Peter Kampits

Univ.-Prof. DDr. Matthias Beck

Univ.-Prof. Dr. Alois Birklbauer

Dr. Andrea Bronner

Univ.-Prof. Dr. Christian Egarter

Dr. Thomas Frühwald

Dr. Ludwig Kaspar

Univ.-Prof. Dr. Lukas Kenner

Dr. Maria Kletecka-Pulker

Univ.-Prof. Dr. Ursula Köller MPH

Univ.-Prof. Mag. Dr. Michael Mayrhofer

Univ.-Prof. Dr. Johannes Gobertus Meran MA

Dr. Stephanie Merckens

Univ.-Prof. Dr. Siegfried Meryn

Univ.-Prof. Dr. Christina Peters

Univ.-Prof. Mag. Dr. Barbara Prainsack

Univ.-Prof. DDr. Walter Schaupp

Univ.-Prof. Dr. Andreas Valentin MBA

Dr. Klaus Voget

Univ.-Prof. Dr. Ina Wagner

Priv.-Doz. Dr. Jürgen Wallner MBA

Univ.-Prof. Dr. Christiane Wendehorst LL.M

Univ.-Prof. Dr. Gabriele Werner-Felmayer



