

# Potentielle negative Folgen von (lernenden) Algorithmen als Entscheidungshilfen für Diagnostik und Intervention im Gesundheitswesen

## Potential negative consequences of (machine learning) algorithms on quality of diagnostics and intervention

### Kurztitel

Nutzung von Algorithmen

### Autoren

Sebastian Völker, Heike Köckler,  
mit Bezug zu Diskussionen mit Gerd Antes und Felix Tretter

Im Gesundheitswesen werden zunehmend Algorithmen für Diagnostik und Therapie eingesetzt. Insbesondere (Deep Learning) Algorithmen werden die Qualität von Diagnose und Therapie verändern. Algorithmen verbessern und erleichtern medizinische Versorgung bei gleichzeitiger Kostenersparnis, jedoch entstehen auch Risiken in verschiedenen Kontexten und Anwendungen. Risiken entstehen durch i) primäre quantitative Evidenzgewinnung mittels großer Fallzahlen ohne individuelle Beurteilung des Patienten und seiner Lebensumstände, ii) eine fehlende Möglichkeit der qualitativen Bewertung von Algorithmen durch medizinisches Personal, iii) schwer oder nicht-nachvollziehbare automatisch produzierte Ergebnisse aufgrund der Funktionsweise und Transparenz des Algorithmus.

Zur Vermeidung dieser Risiken ist es wichtig, ad i) individuelle, durch quantitativ generierte Evidenz beurteilte Gesundheitszustände in Wissenschaft und Praxis mit qualitativen Bewertungen zu ergänzen, ad ii) die Qualität der Algorithmen, deren Nutzen, Risiken und Kosten, in einem Health Technology Assessment transparent und unabhängig zu bewerten, ad iii). ein kritisch-reflektiertes,

in der Praxis nutzbares, informationstheoretisches und technisches Grundverständnis in der Ausbildung der Gesundheitsberufe curricular zu verankern und zu fördern, ad iv) die Humanwissenschaften in Studien- und Ausbildungsgängen mit Gesundheits-, Daten- und Informatikbezug im Sinne einer "human-centered digitalization" zu stärken.

## *Beschreibung des Unseens Nutzung von Algorithmen*

Wir diskutieren, wie durch Algorithmen erzeugte Daten und Entscheidungsunterstützungen die Qualität von Diagnosen und Intervention im Gesundheitswesen beeinflussen. Die Nutzung digitaler Daten und Technologien verbessert in der Regel die medizinische Leistung. In der Diagnostik erlaubt diese einen differenzierten Einblick, z.B. durch bildgebende Verfahren<sup>1</sup>. Im Rahmen des Projekts DiDat diskutieren wir insbesondere Risiken und potentielle negative (Neben-) Folgen (Unseens) bei der Nutzung von (lernenden) Algorithmen.

Der Fokus liegt im Folgenden auf Programmen bzw. Apps, die lernende Algorithmen für medizinisches Fachpersonal einsetzen. Dabei richten wir unsere Aufmerksamkeit auf den Einsatz von algorithmischen Anwendungen für Diagnostik und Therapieentscheidungen, bei denen das medizinische Personal keinen tiefergehenden Ein-

blick in zugrundeliegende Entscheidungsstrukturen und Qualität der verwendeten Algorithmen und ihrer Grenzen sowie ihrer Zuverlässigkeit erhält.

Als spezifische Risiken sind zu benennen:

- (I) Algorithmen lernen häufig aufgrund großer Fallzahlen für die oftmals unklar ist, ob diese für die jeweils untersuchten Individuen eine passende Grundgesamtheit darstellen. Für welche Individuen eine Anwendung geprüft ist und für welche nur eingeschränkt oder nicht, ist eine Anforderung, die nicht immer erfüllt ist.

Schlechte Anwendungen resultieren, wenn den in Gesundheitsberufen Tätigen unklar ist, wie die Daten und Empfehlungen der digitalen Technologie mit den qualitativen Informationen und Befunden in Beziehung zu setzen sind.

- (II) Ein fehlendes, einheitliches Health-Technology Assessment verhindert die qualitative Einordnung digitaler Gesundheitsanwendungen und deren Datengrundlagen, die nicht als Medizinprodukt definiert werden.
- (III) Da das medizinische Personal Teile seiner Tätigkeit an digitale Technologien überträgt, braucht es ein hinreichendes Grundverständnis über die Algorithmen, ihre Eigenschaften und Fehlercharakteristik, um Vertrauen zu

gewinnen und bei Fehlfunktionen fehlerhafte Resultate zu erkennen und zu verwerfen.

Es ist genauer zu untersuchen, inwieweit die Schwierigkeiten der in Gesundheitsberufen Tätigen bei der Anwendung und Vermittlung der Ergebnisse an die Individuen (bei bedeutsamen Diagnosen und Interventionen) Gegenstand einer standardisierten Zulassung werden sollten. Wir sehen hier ein potentiell Risiko, welches wissenschaftlich genauer bewertet werden sollte.

## Ursachen und Erklärung zur Entstehung dieses Unseens

Die maschinelle Analyse von (Gesundheits-) Daten eröffnet mehrere Vorteile. Der Rückgriff auf große Datenmengen ermöglicht eine präzisere Diagnostik, individualisierte Therapie und neue Optionen bei Prävention und Gesundheitsversorgung und -förderung. Das Monitoring von Erkrankungen und die potentielle Fehlerreduktion entlastet das Fachpersonal in Gesundheitsberufen und führt, unterstützt durch eine Prozessautomatisierung, zu einer Kostenreduktion (Jannes et al., 2018). Der Einsatz von Algorithmen birgt jedoch auch Risiken.

Ein Algorithmus trifft auf Basis zweier Grundlagen gesundheitsbezogene Entscheidungen. Entweder werden Parameter, wie bspw. Vitalparameter, Risikofaktoren oder Biomarker genutzt, um auf Basis von fest vom Menschen programmierten Regeln, also dem Algorithmus, Zusammenhänge aufzudecken (Beispiel: subkutan verankerte Insulinpumpen<sup>2</sup>). Die durch den Algorithmus getroffenen Diagnosen und Therapieoptionen sind durch die verwendete Codesprache schwer nachvollziehbar. Die Konstruktion eines regelbasierten Algorithmus auf Basis von a priori festgelegten

Annahmen bestimmt, wie Ergebnisse produziert werden und kann zu falschen Analysen führen (Flores et al., 2016). Die Gefahr besteht darin, dass in Gesundheitsberufen Tätige eine falsche Vorentscheidung eines Algorithmus ganz oder zu großen Teilen übernehmen.

Lernende Algorithmen hingegen suchen auf Basis von Entscheidungsbäumen, Regressionsanalysen oder neuronalen Bayesischen Netzwerken im ständigen Optimierungsprozess noch unbekannte Zusammenhänge in Datenbeständen (Beispiel: Erkennung von Infektionskrankheiten in bildgebender Diagnostik<sup>3</sup>). Neuronale Netze simulieren nach dem Vorbild des Gehirns ein Netzwerk aus miteinander verbundenen Neuronen. Je mehr Neuronen und Schichten existieren, desto komplexere Sachverhalte lassen sich abbilden. Bei diesem sog. „Deep Learning“ werden aus vorhandenen Daten Muster extrahiert und klassifiziert. Dies versetzt einen Algorithmus bzw. KI in die Lage selbstständig und ohne menschliches Zutun seine Fähigkeiten zu verbessern. Anders als beim maschinellen Lernen, wo der Mensch in die Analyse der Daten und in den eigentlichen Entscheidungsprozess eingreift, sorgt beim Deep Learning der Mensch lediglich dafür,

dass Informationen für das Lernen bereitstehen, wie z.B. eine Methodik. Bei lernenden Algorithmen besteht die Gefahr aus verschiedenen Gründen (z.B. der Fehlerfortpflanzung etwa aus fehlerhaften Daten) fehlerhafte Ergebnisse (Artefakte) zu erzeugen (Castelvecchi, 2016). Diese Fehlfunktionen können häufig weder durch den Programmierer noch durch den Anwender nachvollzogen werden. Die Informationen, die durch diese Algorithmen bereitgestellt werden, können demnach gerade von Nicht-Technikern kaum noch erklärt und verstanden werden. Damit erhöht sich die Wahrscheinlichkeit, dass Ergebnisse ohne eingehende Prüfung übernommen werden (Klingel, 2019)<sup>4</sup>.

Bei allen Algorithmen ist die Datenbasis entscheidend für die Qualität des Ergebnisses (Jannes et al., 2018). Bei lernenden Algorithmen wird ein möglicher Bias in den Daten, auf denen der Algorithmus trainiert wurde, in das Ergebnis übertragen (Rebitschek & Gigerenzer, 2020). Wenn beispielsweise die Datenbasis fast ausschließlich aus einer Population erwachsener, männlicher, Probanden aus der Mittelschicht besteht, können fehlerhafte Resultate für Personen produziert werden, die

nicht in ausreichendem Maße dieser Population entsprechen; im genannten Beispiel wären dies Kinder, Frauen, Einkommensschwache etc.

Programme, die gesundheitsbezogene Daten mittels Algorithmen verarbeiten, werden zumeist von Drittanbietern aus dem Technologiebereich entwickelt und vertrieben. Die Anwender können die Funktionsweise und Transparenz eines Algorithmus mit seinen zugrundeliegenden Prozessen und Rechenmodellen nur selten durchdringen. Dies liegt u.a. aufgrund des Betriebsgeheimnisses der Hersteller.

Zur Bewertung der Qualität von Gesundheits-Apps existieren in Deutschland unterschiedliche Gütesiegel auf Basis verschiedener Kriterien<sup>5</sup> (Albrecht, 2016). Die Validität bei den Meisten dieser Siegel wurde jedoch nicht überprüft (Rebitschek & Gigenzer, 2020). Um Nutzen, Risiken und Kosten von Medizinprodukten bewerten zu können, wird in Deutschland ein Health Technology Assessment durchgeführt. In diesem müssen auch Auswirkungen auf PatientInnen und Organisationen dargelegt werden. Derzeit existiert für Algorithmen, die nicht den Kriterien eines Medizinproduktes entsprechen, jedoch als solches von Anwendern wahrgenommen werden (z.B.

„Google Fit“-App, eine App zum Gesundheits- und Aktivitätstracking), kein bundesweit einheitliches Health Technology Assessment auf Basis eines evaluierten Sets von Gütekriterien<sup>6</sup>. Anwendern dieser Algorithmen steht daher kein einheitlicher Bewertungsmaßstab zur qualitativen Einordnung zur Verfügung.

Bei fehlender Reflektion der Informationen durch den Anwender erhöht sich das Risiko, dass Fehlleistungen des Algorithmus gar nicht oder nur spät aufgedeckt werden (Manrai et al., 2018). Die Ursache für die Fehlleistung bleibt durch fehlendes, grundlegendes Funktionsverständnis und fehlende Transparenz dem Anwender teilweise oder vollständig verborgen. Wenn sich Ergebnisse eines Algorithmus einer qualitativen und kritischen Bewertung entziehen, besteht die Gefahr, dass das Fachpersonal im Gesundheitswesen Kompetenzen über gesundheitsrelevante Entscheidungen verliert.

Das Vertrauen in die/den im Gesundheitsberuf Tätige/n ist ein wichtiger Faktor für das Gelingen einer Therapie oder Intervention. Es besteht ein Konflikt zwischen der strukturierten Logik eines Algorithmus und qualitativen (PatientInnen-) Informationen.

Sofern diese unterschiedlichen Wissensquellen durch in Gesundheitsberufen Tätige nicht verbunden werden können, so kann ein Individuum das Vertrauen in den Experten verlieren und führt zu einem Kompetenzverlust der in Gesundheitsberufen Tätigen.

Ein Algorithmus erzeugt Ergebnisse, aus denen Befunde und Evidenz einer besonderen mittelbaren Art auf Basis der Konstruktion des Deep Learning-Algorithmus, eingegebenen Daten und/oder vorgegebenen Kriterien (z.B. Trainingsdatensätze,

Grenzwertfestlegungen) resultieren. Gerade bei Deep-Learning-Algorithmen hat der Mensch keinen Einfluss auf die Ergebnisse des Lernprozesses des Algorithmus. Kontextuelles sowie individuelles Wissen zu Personen und Persönlichkeiten (bspw. persönliche Präferenzen, Ängste und Überzeugungen) kann unter- bzw. unberücksichtigt bleiben, wenn unerfahrene Anwender ohne ausreichendes informationstheoretisches Verständnis den Empfehlungen des Systems Folge leisten, was die Wahrscheinlichkeit diagnostischer, therapeutischer, interventioneller oder handlungsbezogener Fehler erhöhen kann<sup>7</sup>.

## An welchen Zielen orientiert sich ein Umgang mit den Unseens

Gesundheit als Menschenrecht wurde bereits 1946 auf der International Health Conference in New York in der Konstitution der Weltgesundheitsorganisation (WHO), festgeschrieben<sup>8</sup>. Dieses beinhaltet das Recht gesund zu sein und allen Menschen Zugang zu Gesundheitsversorgung zu ermöglichen. Im 1966 auf der Generalversammlung der Vereinten Nationen verabschiedeten Internationalen Pakt über wirtschaftliche, soziale und kulturelle Rechte wurde

das Recht jedes Menschen auf das für ihn erreichbare Höchstmaß an körperlicher und geistiger Gesundheit anerkannt<sup>9</sup>.

Ziele der Gesundheitspolitik sollten demnach die Prävention von Krankheiten und Unfällen sowie die Gesundheitsförderung sein, so dass das Gesundheitssystem so effizient und kostengünstig wie möglich arbeiten kann, um die Gesundheit und das Wohlbefinden sowie die Zufriedenheit der Bevölkerung sicherzustellen.

Durch die Digitalisierung wurden neue Diagnostik- und Behandlungsmöglichkeiten geschaffen. Zum Beispiel erleichtert die personalisierte Medizin die Kommunikation zwischen den einzelnen Akteuren des Gesundheitswesens und ermöglicht es dem einzelnen Patienten, seine Gesundheit durch Apps und Informationen im Internet stärker zu steuern. Im Bereich Diagnostik und Intervention ist es wichtig, dass der Einsatz von zeitgemäßen digitalen Verfahren bei Bildgebung, Selbstüberwachung durch den Patienten und Entscheidungsfindung (im medizinischen und Public Health Bereich) geprüft wird.

Jedoch sollten auch unbeabsichtigte Nebeneffekte („Unseens“) beim Einsatz neuer digitaler Methodiken wie Algorithmen betrachtet werden. Beispielsweise beinhaltet die Magnetresonanztomographie eine Vielzahl von Algorithmen, deren Konstruktion der Anwender nicht vollständig verstehen muss. Um eine Anwendbarkeit von Algorithmen in der gesundheitlichen/medizinischen Praxis zu ermöglichen, müssen jedoch die automatisch generierten Ergebnisse eine Bewertung durch das anwendende Fachpersonal im Gesundheitswesen ermöglichen. Die Datenbasis, auf die

die Algorithmen zurückgreifen, soll so aufgebaut sein, dass sie keinen Bias gegenüber sozialen Gruppen beinhaltet und reproduziert.

Im Gesundheitswesen ist eine informierte, kognitive und emotionale Bewertung von algorithmisch generierten Informationen von großer Bedeutung (Jannes et al., 2018). Maschinell generierte Informationen, Empfehlungen und Vorentscheidungen müssen, beispielsweise in der Gesundheitsplanung, vor dem Hintergrund politischer und lokaler Rahmenbedingungen bewertet werden. Bedeutsame Diagnosen oder Intervention aus Algorithmen, die für ein Individuum ein Life Event bedeuten können, bedürfen einer informierten und Individuen-sensiblen Bewertung und Einordnung auf Basis der Fachexpertise des/der im Gesundheitsberuf Tätigen.

Beim Einsatz von Algorithmus-basierten DiGAs soll die hörend-sprechende Medizin nicht vernachlässigt werden, um das Verhältnis zum Individuum als zentrales Element für eine gelingende Diagnostik und Therapie zu erhalten (Cohen et al., 2014). Dies erhöht den Respekt vor der Menschenwürde, der/die PatientIn fühlt sich als Individuum begriffen und nicht als Daten-subjekt. In der Folge ist auch die Adhärenz

gegenüber Therapie- und Verhaltensempfehlungen größer (siehe SI 2.6). Bei Public Health-Interventionen, beispielsweise in der Gesundheitsplanung, müssen Belange der beteiligten Stakeholder bei der Entscheidungsfindung berücksichtigt werden.

Bei der Zertifizierung medizinischer Software (gemäß Art.2 MDR Aufgabe der Benannten Stellen) als auch beim Health Technology Assessment (gemäß §139e SGB V Aufgabe des BfArM<sup>10</sup>) sollten klare und einheitliche Kriterien erarbeitet und angewendet werden. Es ist zu prüfen, ob Algorithmen in Gesundheitsanwendungen, die nicht den Kriterien eines Medizinproduktes entsprechen, jedoch in der gesundheitlichen/medizinischen Praxis eingesetzt werden, ebenfalls einen einheitlichen Qualitätsstandard bedürfen. Bei der Konstruktion der Algorithmen ist zu beachten, dass aktuelle Leitlinien und Empfehlungen der quantitativen Datenauswertungen, wie bspw. die Gute Praxis Sekundärdatenanalyse (GPS)<sup>11</sup> beachtet werden und Datenbasen transparent ausgewiesen werden. Zertifizierungsverfahren nach der DSGVO könnte perspektivisch eine Orientierung bieten.

Es ist zu klären, ab wann das Risiko von negativen Folgen bei der Nutzung von AI in

kritischer Weise erhöht wird. Diese ist abhängig von der Bedeutung der Information, die durch den Algorithmus generiert wird, die festgestellte Diagnostizität/Leistungsfähigkeit der algorithmischen digitalen Gesundheitsanwendung (DiGA) und das Wissen des/der im Gesundheitsberuf Tätigen, der Fähigkeit des/der im Gesundheitsberuf Tätigen Ausreißer, Fehler und untypische Ergebnisse zu erkennen und die Fähigkeit des/der im Gesundheitsberuf Tätigen die Arbeitsweise einer DiGa und ihrer Grenzen zu erklären.

In Zusammenarbeit mit Fachpersonal aus dem Gesundheitswesen sollen Individuen befähigt werden eine gesundheitsbezogene Entscheidung bspw. zu einer Therapie zu treffen (Chewning et al., 2012). Trotzdem wird die Hauptverantwortung bei in Gesundheitsberufen Tätigen verbleiben, die Ergebnisse und deren Sinnhaftigkeit prüfen und dem/der Patienten/in eine Beschreibung des Leistungsvermögens geben. Die für Interventionen benötigte Vertrauensbasis des/der Patienten/in bzw. des/der Entscheidungsträgers/in gegenüber den in Gesundheitsberufen Tätigen wird durch einen Algorithmus nicht ersetzt werden können (Stevenson et al., 2007). Für eine gute Nutzungs-Praxis, in welcher

Algorithmen neue Informationen aus komplexen Zusammenhängen (Multikausalitäten, Bildererkennung etc.) ableiten, die den in Gesundheitsberufen Tätigen zusätzliche Informationen bereitstellen, muss unter Kenntnis i) der Funktionsweise des Algo-

rithmus, ii) der Fehlerursachen in der Beurteilung und iii) der Grenzen der Anwendung, die Algorithmus-basierte Information in angemessener Weise an Individuen vermittelt und in eine mögliche Intervention einbezogen werden können.

## *Welche Maßnahmen sind für welche Ziele sinnvoll*

Für eine Qualitätskontrolle der Algorithmen wird gegenwärtig medizinische Software zertifiziert und für die als Medizinprodukte eingestuft Gesundheitsanwendungen ein Health Technology Assessment durchgeführt. Für die Entwicklung von Kriterien zur Beurteilung von Nutzen, Risiko und Kosten digitaler Gesundheitsanwendungen ist ein transdisziplinärer Prozess notwendig, um wissenschaftliche Erkenntnisse mit praktischer Einsatzfähigkeit zu verbinden (Antes, 2018). In diesen Prozess sind zentrale Akteure der im Systemmodell dargestellten Bereiche (in Gesundheitsberufen Tätige, Individuen, Ökonomie, Technologie, Ethik/Jura, Verwaltung, organisatorische Strukturen) einzubinden.

Die Güte der Ergebnisse aus Algorithmen sollten möglichst quantitativ, beispielsweise mittels positiv prädiktiven Werten

(PPV) oder Konfidenzintervallen, ausgewiesen und zugänglich gemacht werden<sup>12</sup> (Gigerenzer, 2013). Eine eindeutige Differenzierung zwischen Korrelation und Kausalität unter Vermeidung von Scheinkorrelationen muss ausgewiesen werden (Vezyridis & Timmons, 2019). Hilfreich wäre in diesem Zuge auch informiertem Fachpersonal in Gesundheitsberufen zu ermöglichen in die Entscheidungsbäume von regelbasierten Algorithmen, bspw. durch transparente und flexible Grenzwertfestlegungen, eingreifen zu können, um ihre zuvor getroffene Entscheidung datengestützt zu prüfen<sup>13</sup>. Hierzu müssen Kenntnisse der grundlegenden Funktionsweisen von Algorithmen Fachpersonal in Gesundheitsberufen vermittelt werden.

Bei der Anwendung von Algorithmen muss eine individuelle Betrachtung der Gesundheitschancen und -risiken vorgenommen und DiGA, v.a. außerhalb der bildgebenden Diagnostik, als Entscheidungsunterstützung bzw. „Zweitmeinung“ angesehen werden und nicht ausschließlich deterministisch eingesetzt werden. Folglich sollten einer DiGA vordringlich Assistenzfunktionen zukommen und ihre Ergebnisse als ein relevanter Faktor Eingang in eine menschliche Entscheidung finden. Vor einer abschließenden Entscheidung sollte die Intervenierbarkeit des menschlichen Entscheiders, aber auch des Betroffenen, stets gegeben sein. Die in Gesundheitsberufen Tätigen müssen Grundkenntnisse zu Gesundheitsdaten sowie Funktionsweise und Einsatzmöglichkeiten von Algorithmen kennen und informationstheoretische Kenntnisse zur Bewertung der automatisch generierten Ergebnisse besitzen. Die Kenntnisse sollten dabei auf Transparenz und Nachvollziehbarkeit ausgerichtet sein, nicht auf der Konstruktionsbeschreibung des Algorithmus (Code). Dies stärkt das Vertrauen der Nutzer in die algorithmisch produzierten Ergebnisse und ermöglicht eine angemessene Aufklärung von PatientInnen über den

Einsatz von Gesundheitsanwendungen und deren Unsicherheiten und Risiken, auch wenn der/die PatientIn selbstständig eine DiGA genutzt hat (Huss, 2019, siehe SI 2.1). Eine Transparenz von Algorithmen ist hierfür äußerst bedeutsam. Diese beinhaltet die Offenlegung und Kennzeichnung der Algorithmen, der Nachweis der verwendeten Daten sowie die Übersicht über Kontrollmöglichkeiten durch die AnwenderInnen.

Die curriculare Verankerung eines kritisch-reflektierten, in der Praxis nutzbaren, informationstheoretischen und technischen Grundverständnisses in bestehenden und zukünftigen Ausbildungs- und Studiengängen mit Gesundheits-, Daten- und Informatikbezug im Sinne einer human-centered digitalization begegnet falschen Erwartungen an DiGAs<sup>14</sup>, vergrößert das Verständnis für gesundheitliche Ziele und unterstützt den Austausch zwischen Medizin/Public Health und IT. Kurzfristig ermöglicht die Weiterbildung in Gesundheitsberufen Tätiger einen hochwertig-informierten Einsatz DiGA zum Wohle des Individuums und der Allgemeinheit.

## *Begründung für die Orientierung*

Die potentielle Verbesserung der gesundheitlichen Versorgung durch Algorithmen kann durch schwer bzw. falsch zu interpretierende numerische / quantitative Ergebnisse ins Gegenteil verkehrt werden. Daher sollte ein informationstheoretisches und technisches Grundverständnis sowie eine Qualitätsbewertung von DiGAs in Medizin, Public Health und anderen Gesundheitsberufen etabliert werden, die eine Ergebnisreflektion und den richtigen Einsatz zum bestmöglichen Nutzen für Individuen und Bevölkerung sicherstellt.

Im Kern leiten sich Orientierungen, Ziele und Maßnahmen aus dem ab, was bereits im konventionellen, nicht digitalen Bereich als Standard akzeptiert und etabliert ist. Dies gilt vor allem für Fragen von Datenschutz und -sicherheit. Obwohl klare Regeln vorgegeben sind, bedarf es im Umfeld der digitalen Medizin meist besonderer Sorgfalt und spezifischer Regelungen, auch vor dem Hintergrund, dass Künstliche Intelligenz (z.B. selbstlernende Algorithmen) zunehmend zu einem „Gestalter“ im Gesundheitssystem wird. Ziel ist, dass in diesem besonders innovativen, aber auch sensiblen Feld möglichst keine Nachteile, beispielsweise in Form von falscher Diagnostik oder Therapieentscheidung, für Gesundheitsversorgung, Gesundheitsschutz,

Gesundheitsförderung und Medizin entstehen.

Hierzu gehört die Förderung eines kritisch-reflektierten, in der Praxis nutzbaren, informationstheoretischen und technischen Grundverständnis in der Ausbildung der Gesundheitsberufe durch spezifische Schulungen von Fachpersonal in Gesundheitsberufen und der curricularen Implementierung in Ausbildung und Studium der Gesundheitsberufe (Medizin, Public Health etc.). Die transparente Ausweisung von Nutzen, Risiken und Kosten ist hierfür ein zentrales Element, aber auch die Stärkung des Verständnisses der Ergebnisgenerierung seitens des Algorithmus und damit das Vertrauen in die algorithmisch gene-

rierten Ergebnisse für in Gesundheitsberufen Tätige als auch für deren Vermittlung gegenüber Patienten.

Das Handeln im Gesundheitssystem darf z.B. in der Patientenversorgung und in der Gesundheitsplanung auch in Zukunft nicht ausschließlich von digitalen Daten determiniert werden. Die Beachtung individueller

Gesundheitszustände und Patientenverhalten in Wissenschaft und Praxis über quantitativ generierte Informationen hinaus, stellt eine nötige Akzentuierung im verantwortungsvollen Umgang mit digitalen Daten dar.

## Literatur zu den wesentlichen Aussagen

- Antes, G. (2018, Januar 2). *Die Medizin im Datenrausch*. F.A.Z. Plus. <https://zeitung.faz.net/feuilleton/2018-01-02/9b583344667f696c3b3aabb13b7424f?GEPC=s2>
- Castelvecchi, D. (2016). Can we open the black box of AI? *Nature News*, 538(7623), 20. <https://doi.org/10.1038/538020a>
- Chewning, B., Bylund, C. L., Shah, B., Arora, N. K., Gueguen, J. A., & Makoul, G. (2012). Patient preferences for shared decisions: A systematic review. *Patient Education and Counseling*, 86(1), 9–18. <https://doi.org/10.1016/j.pec.2011.02.004>
- Cohen, I. G., Amarasingham, R., Shah, A., Xie, B., & Lo, B. (2014). The legal and ethical concerns that arise from using complex predictive analytics in health care. *Health Affairs (Project Hope)*, 33(7), 1139–1147. <https://doi.org/10.1377/hlthaff.2014.0048>
- Flores, A. W., Bechtel, K., & Lowenkamp, C. T. (2016). False Positives, False Negatives, and False Analyses: A Rejoinder to Machine Bias: There's Software Used across the Country to Predict Future Criminals. And It's Biased against Blacks. *Federal Probation*, 80, 38.
- Huss, R. (2019). Therapie mit Erfolg und ohne Nebenwirkungen. In R. Huss (Hrsg.), *Künstliche Intelligenz, Robotik und Big Data in der Medizin* (S. 47–58). Springer. [https://doi.org/10.1007/978-3-662-58151-3\\_4](https://doi.org/10.1007/978-3-662-58151-3_4)
- Jannes, M., Friele, M., Jannes, C., & Woopen, C. (2018). *Algorithmen in der digitalen Gesundheitsversorgung*. <https://www.bertelsmann-stiftung.de/de/publikationen/publikation/did/algorithmen-in-der-digitalen-gesundheitsversorgung> (abgerufen am: 10.07.2020)
- Klingel, A. (2019). *Gesund dank Algorithmen? Chancen und Herausforderungen von Gesundheits-Apps für Patient:innen*. <https://www.bertelsmann-stiftung.de/de/publikationen/publikation/did/gesund-dank-algorithmen> (abgerufen am: 10.07.2020)
- Manrai, A. K., Patel, C. J., & Ioannidis, J. P. A. (2018). In the Era of Precision Medicine and Big Data, Who Is Normal? *JAMA*, 319(19), 1981–1982. <https://doi.org/10.1001/jama.2018.2009>
- Rebitschek, F.G., & Gigerenzer, G. (2020). Einschätzung der Qualität digitaler Gesundheitsangebote: Wie können informierte Entscheidungen gefördert werden? *Bundesgesundheitsbl* 63, 665–673. <https://doi.org/10.1007/s00103-020-03146-3>
- Stevenson, F. A., Kerr, C., Murray, E., & Nazareth, I. (2007). Information from the Internet and the doctor-patient relationship: The patient perspective – a qualitative study. *BMC Family Practice*, 8(1), 47. <https://doi.org/10.1186/1471-2296-8-47>
- Vezyridis, P., & Timmons, S. (2019). Resisting big data exploitations in public healthcare: Free riding or distributive justice? *Sociology of Health & Illness*, 41(8), 1585–1599. <https://doi.org/10.1111/1467-9566.12969>

- 
- <sup>1</sup> Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115–118. <https://doi.org/10.1038/nature21056>
  - <sup>2</sup> Cinar, A. (2019). Automated Insulin Delivery Algorithms. *Diabetes Spectrum*, 32(3), 209–214. <https://doi.org/10.2337/ds18-0100>
  - <sup>3</sup> Madabhushi, A., & Lee, G. (2016). Image analysis and machine learning in digital pathology: Challenges and opportunities. *Medical image analysis*, 33, 170–175. <https://doi.org/10.1016/j.media.2016.06.037>
  - <sup>4</sup> Siehe auch Antes, G. (2019, April 2). *Digital Health. Künstliche Dummheit statt künstliche Intelligenz?* ÄrzteZeitung. <https://www.aerztezeitung.de/Wirtschaft/Kuenstliche-Dummheit-statt-kuenstliche-Intelligenz-256806.html>
  - <sup>5</sup> Albrecht, U.-V. (2016). *Chancen und Risiken von Gesundheits-Apps (CHARISMHA)*. Medizinische Hochschule Hannover. <http://www.digibib.tu-bs.de/?docid=00060000> (abgerufen am: 10.4.2020)
  - <sup>6</sup> Haverinen, J., Keränen, N., Falkenbach, P., Maijala, A., Kolehmainen, T., & Reponen, J. (2019). Digi-HTA: Health technology assessment framework for digital healthcare services. *Finnish Journal of EHealth and EWelfare*, 11(4), 326–341–326–341. <https://doi.org/10.23996/fjhw.82538>
  - <sup>7</sup> Bauer, M., Glenn, T., Monteith, S., Bauer, R., Whybrow, P. C., & Geddes, J. (2017). Ethical perspectives on recommending digital technology for patients with mental illness. *International Journal of Bipolar Disorders*, 5(1), 6. <https://doi.org/10.1186/s40345-017-0073-9>
  - <sup>8</sup> WHO (1948). Constitution of the World Health Organization. Geneva.
  - <sup>9</sup> UN Generalversammlung (1966). International Covenant on Economic, Social and Cultural Rights, 16 December 1966, United Nations, Treaty Series, vol. 993.
  - <sup>10</sup> Eine systematische, evidenzbasierte Bewertung medizinischer Verfahren und Technologien im Hinblick auf deren Effekte auf die Gesundheitsversorgung wird nicht durch das BfArM durchgeführt. Das BfArM überprüft die Angaben des Herstellers in ähnlicher Weise wie die Arzneimittelbewertung gemäß AMNOG ([https://www.bfarm.de/DE/Medizinprodukte/DVG/\\_node.html](https://www.bfarm.de/DE/Medizinprodukte/DVG/_node.html)).
  - <sup>11</sup> Swart, E., Gothe, H., Geyer, S., Jaunzeme, J., Maier, B., Grobe, T. G., & Ihle, P. (2015). Gute Praxis Sekundärdatenanalyse (GPS): Leitlinien und Empfehlungen. *Das Gesundheitswesen*, 77(02), 120–126. <https://doi.org/10.1055/s-0034-1396815>
  - <sup>12</sup> Gigerenzer, G. (2013). *Risiko: Wie man die richtigen Entscheidungen trifft*. Bertelsmann.
  - <sup>13</sup> GVG - Gesellschaft für Versicherungswissenschaft (ohne Datum). GVG- Positionspapier zu Kernanforderungen an eine künstliche Intelligenz im Gesundheitswesen.

---

<https://gvg.org/allgemein/publikationen/positionen/gvg-facharbeitsgruppe-digitalisierung-position-zu-kernanforderungen-an-eine-kuenstliche-intelligenz-im-gesundheitswesen/#more-4907> (Abrufdatum: 13.07.2020)

- <sup>14</sup> Die Unintended Side Effects resultieren vermutlich aus falschen Erwartungen, in der die multiplen Mensch-Umwelt Beziehungen im Spannungsfeld von AI-DiGA – in Gesundheitsberufen Tätige - Individuen ausgeblendet werden. Grenzen und Potentiale der AI-DiGAs werden von in Gesundheitsberufen Tätigen teilweise über- aber auch unterschätzt.