

Zulassung KI-basierter Clinical Decision Support Systems unter der Medical Device Regulation

Von *Simone Kuhlmann*¹

I. Einleitung

Ärztliches Entscheiden ist Dreh- und Angelpunkt jeglichen ärztlichen Tuns. Von der ärztlichen Entscheidung hängt ab, ob es einer medizinischen Intervention bedarf, welcher Intervention es bedarf und wie diese Intervention im Einzelnen ausgestaltet ist. Seit den 1970er Jahren werden zur Unterstützung solcher Entscheidungen zunehmend sog. computergestützte Clinical Decision Support Systeme (CDSS) – zu Deutsch Entscheidungsunterstützungssysteme – herangezogen.² Hierbei handelt es sich um Computersysteme, die vor allem Ärzte, aber auch andere an der Versorgung beteiligte Berufe³ bei Entscheidungsaufgaben in dem Moment, in dem sie anfallen, insb. aber am Ort der Versorgung unterstützen sollen.⁴ Ihr Ziel ist es, medizinisches Wissen in Abhängigkeit zu individuellen Patientendaten darzustellen⁵ oder aufzubereiten und so neues Wissen zu generieren, um daraus in einem zweiten Schritt Handlungs- und Entscheidungsempfehlungen ableiten zu können. Neben einer Standardisierung der Behandlung von Patientinnen zur Vermeidung von Über- oder Unterbehandlungen bieten sie den Vorteil, medizinisches Wissen kontextbezogen schnell verfügbar machen zu können, was vor allem angesichts der stetig anwachsenden

¹ Die Autorin dankt Herrn Florian Lucks, studentischer Mitarbeiter am Zentrum für Recht in der digitalen Transformation, für die stete Diskussionsbereitschaft bei der Erstellung des Beitrags.

² R. Sutton/D. Pincock/D. Baumgart et al., An overview of clinical decision support systems: benefits, risks, and strategies for success, *npj Digit Med* 2020, 3:17; zur Historie von CDSS vgl. R. Greenes, *Clinical Decision Support. The Road to Broad Adoption*, 2. Aufl. 2014, S. 49 ff.

³ Z. B. Pflege, Logopäden, Psychologinnen.

⁴ E. Berner/T. La Lande, Overview of Clinical Decision Support Systems, in: *Clinical Decision Support Systems, Theory and Practice*, 2016, S. 1; P. Börm, Leitlinienbasierter Clinical Decision Support – Anforderungen an evidenzbasierte Entscheidungsunterstützungssysteme, *OP-Journal* 2021, 37, S. 28.

⁵ Zu den darstellenden Systemen gehören vor allem Dokumentations- und Monitoringssysteme bspw. auf Intensivstationen, vgl. hierzu J. Steinwendner, Klinische Entscheidungsunterstützungssysteme: von der Datenrepräsentation zur künstlichen Intelligenz, in: *Innovation und Innovationsmanagement im Gesundheitswesen*, 2020, S. 687 f.

Fülle von Erkenntnissen zu einzelnen Erkrankungen und deren Entstehung und Behandlung für die Versorgung von nicht unerheblicher Bedeutung ist.⁶

Ein wachsender Anteil von CDSS basiert heute auf Verfahren Künstlicher Intelligenz (KI).⁷ Davon verspricht man sich neben einer Entlastung des medizinischen Personals eine präzisere und individualisiertere Diagnostik und Therapieauswahl und damit insgesamt eine verbesserte medizinische Versorgung. Gleichwohl werden mit dem Einsatz KI-basierter CDSS aufgrund ihrer Funktionsweisen und ihrer Opazität Risiken assoziiert, was die Frage der Bedingungen ihrer Zulassung für die medizinische Versorgung aufwirft. Der Beitrag identifiziert und analysiert daher nach einer knappen technischen Einführung und Beschreibung der Anwendungsfelder (II.) die spezifischen Risiken des Einsatzes KI-basierter CDSS (III.), leitet darauf basierend spezifische Anforderungen für den Einsatz solcher Systeme in der medizinischen Versorgung her (IV.) und untersucht, inwieweit diese bereits durch den bestehenden Rechtsrahmen der Medical Device Regulation (MDR) abgesichert werden (V.).

II. KI-basierte CDSS in der medizinischen Versorgung

KI-basierte Systeme meinen nach dem heutigen Stand vor allem Systeme, die mittels statistischer Verfahren, wie maschineller Lernverfahren, entwickelt worden sind. Sie dienen etwa der Klassifikation bzw. Bestimmung einer Kategorie (bspw. zur Detektion suspekter Gebiete bzw. *regions-of-interest* in MRT/CT-Scans etwa im Bereich der Krebs- oder Schlaganfall-Diagnostik) oder zur Regression, also der Vorhersage eines quantitativen Zahlenwerts (z. B. eines Laborwerts oder zur Prognose eines Krankheitsverlaufs). Zu den in der Medizin eingesetzten Methoden maschinellen Lernens gehören neben Cluster-Analysen und *Support Vector Machines Deep Learning*, *Neuronale Netze*, *Nearest Neighbor Classifiers* sowie *Lineare* und *Logistic Regression*.⁸

⁶ Steinwendner (Fn. 5), S. 690.

⁷ Sutton/Pincock/Baumgart et al. (Fn. 2); Stellungnahme der Zentralen Ethikkommission (ZEKO) bei der Bundesärztekammer zur Entscheidungsunterstützung ärztlicher Tätigkeit durch Künstliche Intelligenz, Deutsches Ärzteblatt, Jg. 118, Heft 33–34, 23.8.2021, A3; I. Schneider, Diskriminierungsgefahren und Regulationsansätze bei der medizinischen Nutzung von KI, Zeitschrift für medizinische Ethik 2021, S. 327 (330).

⁸ S. Stoppacher/P. Müllner, Software als Medizinprodukt, in: Anforderungen an Medizinprodukte, 2021, S. 169; L. Pechmann/M. Mildner/T. Suthau/M. Leucker, Regulatorische Anforderungen an Lösungen der künstlichen Intelligenz im Gesundheitswesen, in: Künstliche Intelligenz im Gesundheitswesen, 2022, S. 184 f.; zu den verschiedenen Methoden im Detail siehe B. Ozaydin/I. Hardin/D. Chhieng, Data Mining and Clinical Decision Support Systems, in: Clinical Decision Support Systems, Theory and Practice, 2016, S. 51 ff.

Der Unterschied zu herkömmlichen sog. wissensbasierten CDSS⁹ liegt darin, dass KI-basierte Systeme nicht auf formalisierten, in Regeln verfassten medizinischen Erkenntnissen und Annahmen – häufig beschrieben in eindeutigen *If-Then*-Regeln – basieren,¹⁰ die auf jahrzehntlang entstandenes Expertenwissen in diesem Bereich fußen, wie es etwa in medizinischen Leitlinien ärztlicher Fachgremien oder Fachverbände niedergelegt ist. Also, wenn der Laborwert $p > x$ ist und zusätzlich Symptome a und b vorliegen, dann weist das auf die mögliche Diagnose Y oder Z hin. Stattdessen agieren KI-basierte Systeme anhand von Regeln, die sie durch Auswertung von Daten erlernt haben.¹¹ Medizinische Kenntnisse und Annahmen werden also nicht mehr von Programmierern in klare Handlungsanweisungen gecodet, sondern es werden Modelle mithilfe einer Menge von für die zu bewältigende Aufgabenstellung relevanten, existierenden medizinischen Beispieldatensätzen trainiert, die dann für neue Daten Vorhersagen und Annahmen treffen können.¹² Entsprechend werden sie als „nicht wissensbasierte CDSS“ bezeichnet.¹³

Das Problem dieser Systeme ist, dass die gelernten Regeln lediglich auf Korrelationen basieren. Sie erkennen keine kausalen Zusammenhänge, d. h. konkrete Beziehungen zwischen Ursache und Wirkung, selbst wenn diese vorliegen.¹⁴ Das gehäufte Auftreten von Retinopathie bei Diabetespatientinnen bspw. kann, muss aber nicht auf einen kausalen Zusammenhang zwischen diesen Erkrankungen hindeuten. Hinzu kommt, dass die Systeme – je nach verwendeter Methode (Neuronales Netz, Deep Learning) – sehr komplex sind und das Wie und Warum des Zustandekommens des von ihnen ausgeworfenen Outputs vom Menschen nicht oder nur noch begrenzt nachvollzogen werden kann (sog. Black-Box-Problematik).¹⁵ Insbesondere bei

⁹ Vgl. allgemein zu den unterschiedlichen Formen von CDSS *Steinwendner* (Fn. 5), S. 686; *Börm* (Fn. 4), S. 29; zur Taxonomie der verschiedenen CDSS vgl. auch *J. Oshero/J. Teich/D. Levick et al.*, Improving outcomes with clinical decision support, an implementer's guide, Healthcare Information and Management Systems Society (HIMSS), 2012.

¹⁰ *S. Spooner*, Mathematical Foundations of Decision Support Systems, in: *Clinical Decision Support Systems, Theory and Practice*, 2016, S. 37; *Ozaydin/Hardin/Chhieng* (Fn. 8), S. 46.

¹¹ Vgl. *G. Marakas*, Decision Support Systems in the 21st century, 1999; *Berner/La Lande* (Fn. 4), S. 5; *Börm* (Fn. 4), S. 29.

¹² *Ozaydin/Hardin/Chhieng* (Fn. 8), S. 48; vgl. dazu auch allgemein *J. Kleesiek/J. Murray/C. Strack/G. Kaissis/R. Braren*, Wie funktioniert maschinelles Lernen?, *Der Radiologe* 2020, S. 24 (25 f.).

¹³ *Spooner* (Fn. 10), S. 36 f.; *Börm* (Fn. 4), S. 29.

¹⁴ *S. Hännold/N. Schlee/D. Antweiler/K. Beckh*, Die Nachvollziehbarkeit von KI-Anwendungen in der Medizin. Eine Betrachtung aus juristischer Perspektive mit Beispielszenarien, *MedR* 2021, S. 516 (518).

¹⁵ Vgl. hierzu Bericht der Enquete-Kommission Künstliche Intelligenz – Gesellschaftliche Verantwortung und wirtschaftliche, soziale und ökologische Potenziale, BT-Drs. 19/23700, S. 64; Stellungnahme ZEKO (Fn. 7), A 3. *U. Gassner/U. Juknat*, in: *E-Health Digital Health*, 2022, Kap. 3 Rn. 374; *J. Bjerring/J. Busch*, Artificial Intelligence and Patient-Centered Decision-Making, *Philosophy & Technology* 2021, S. 349 (353 ff.); *bitkom*, Blick in die Black-

künstlichen neuronalen Netzen (KNN), die regelmäßig eine Million von optimierten Parametern aufweisen und daher besonders gute Ergebnisse hervorbringen, ist für den Menschen i. d. R. nicht mehr begreiflich, welche Merkmale bzw. hochdimensionalen Kombinationen von Merkmalen für das Ergebnis entscheidend sind.¹⁶ Das schränkt jedoch die Möglichkeit der Validierung solcher Systeme ein und begründet letztendlich eine gewisse Skepsis diesen gegenüber. Zumal auch nur bedingt nachvollzogen werden kann, ob der Output des Systems wirklich das aussagt, wie wir ihn interpretieren. Nichtsdestotrotz werden zunehmend solche Systeme aufgrund ihrer Vorzüge und Potentiale gegenüber herkömmlichen wissensbasierten CDSS entwickelt. Sie ermöglichen es, nicht nur bislang unbekannte statistische Zusammenhänge bspw. von Erkrankungen zu identifizieren, die dann ggfs. weiter erforscht werden können.¹⁷ Durch ihre Fähigkeit, mit einer großen Menge an personalisierten Patientendaten zu operieren, erlauben sie zudem eine deutlich präzisere und individualisierte Diagnostik sowie Therapieauswahl und damit eine Verbesserung der Versorgung insgesamt.¹⁸ Insbesondere dort, wo Versorgungslücken etwa durch Fachärztemangel bestehen, könnten solche Systeme helfen, weil sie auch von Nicht-Fachärztinnen bedient werden könnten.¹⁹ Hinzu kommt, dass derartige Systeme besser mit Unsicherheiten umgehen können²⁰ und – sofern gewollt – in der Lage sind, fortlaufend zu lernen und sich anzupassen. Anders als bei herkömmlichen, wissensbasierten Systemen müssen die vielen Regeln, aus denen ein solches System besteht, also nicht mühsam entsprechend dem sich weiterentwickelnden medizinischen Wissen fortlaufend aktualisiert werden.

KI-basierte CDSS kommen bereits sowohl zur Diagnostik, im therapeutischen Bereich als auch zur Erstellung von klinischen Prognosen zur Anwendung, wobei vor allem letzteres aus medizinethischer Sicht nicht unumstritten ist.²¹ Denn hierbei geht es um die Berechnung von Wahrscheinlichkeiten für unerwünschte medizinische Zwischenfälle (bspw. kardiovaskuläre Ereignisse oder postoperative Komplikationen²²) oder die Überlebensdauer von Patienten, die etwa an einer Krebserkrankung

box. Nachvollziehbarkeit von KI-Algorithmen in der Praxis, 2019, S. 12; Börm (Fn. 4), S. 29; Spooner (Fn. 10), S. 37.

¹⁶ Vgl. dazu im Detail Kleesiek/Murray/Strack/Kaisis/Braren (Fn. 12), S. 26 und S. 29.

¹⁷ Stellungnahme ZEKO (Fn. 7), A 4.

¹⁸ Vgl. Stellungnahme der ZEKO (Fn. 7), A 1 f.

¹⁹ Z.B. erlaubt ein vom EU-finanzierten Projekt IScan entwickelter KI-basierter Netzhautscanner die Diagnostik Diabetischer Retinopathie ohne besonderes Fachwissen und Gerätetechnik, so dass er sowohl für die hausärztliche Versorgung als auch für Apotheken als geeignet gilt, <https://cordis.europa.eu/article/id/425627-ai-powered-retinal-scanner-picks-up-early-signs-of-diabetes/de> (zuletzt abgerufen am 16.2.2023).

²⁰ Ozaydin/Hardin/Chhieng (Fn. 8), S. 48.

²¹ Stellungnahme ZEKO (Fn. 7), A 2; siehe weitere Beispiele für CDSS bei Ozaydin/Hardin/Chhieng (Fn. 810), S. 47 f.; C. Johner, Regulatorische Anforderungen an Medizinprodukte, in: Die Zukunft der Medizin, 2019, S. 367.

²² Vgl. bspw. die seit 2021 zugelassene Software *x-c-bleeding*, die Muster von Nachblutungen oder Nierenversagen nach Herzoperationen erkennt und Alarm schlägt, bevor schwere

leiden, auf Basis von vorliegenden Falldaten von anderen Patientinnen.²³ Dabei können Krankheitsverläufe – abhängig von zahlreichen Variablen (z. B. Krankenhausumgebung, persönliches Umfeld) – individuell stark variieren, was die Berücksichtigung solcher Prognosen bei Therapieentscheidungen fraglich macht. Im Gegensatz dazu werden vor allem Diagnostik-unterstützende CDSS sowohl vor dem Hintergrund ihres Potentials, finanzielle und personelle Ressourcen zu schonen, als auch ihrer Präzision als echter Gewinn für den klinischen Alltag angesehen.²⁴ So haben Studien insbesondere bei Aufgaben, bei denen es um die Detektion von Abweichungen und Auffälligkeiten in Bildern oder anderen Messdaten geht (z. B. Arrhythmie-Detektion im EKG²⁵, Erkennung von MS-Läsionen am Gehirn anhand von MRT-Scans, Detektion von diabetischer Retinopathie auf Basis von Netzhaut-Scans²⁶, Hautkrebs-Screening²⁷), gezeigt, dass ihre Ergebnisse mit denen von Ärzten vergleichbar oder z. T. sogar besser als diese sind.²⁸ CDSS, die im therapeutischen Bereich zum Einsatz kommen, dienen hingegen in der Regel dazu, die Therapie mit dem individuell höchsten Nutzen für die jeweilige Patientin zu identifizieren.²⁹ Dies können Systeme sein, die anhand von Patientinnendaten die für den Patienten individuell wirksamste Therapie bestimmen (z. B. Therapie mit einem CRT-D oder mit einem ICD bei Herzinsuffizienz³⁰; Bestrahlungsplan; Dosisberechnung) oder die die Präzision von invasiven Eingriffen erhöhen (z. B. durch Verbesserung der endoskopischen Navigation durch präzise Identifikation erkrankten Gewebes³¹; Berechnung des Einstichwinkels eines chirurgischen Instruments). Insgesamt muss jedoch festgehalten werden, dass all diese Systeme für die Lösung einer bestimmter (Teil-)Aufgabe ent-

Komplikationen eintreten, PM des Berlin Institute of Health v. 8.7.2021, <https://www.bihelth.org/de/aktuell/kuenstliche-intelligenz-auf-der-intensivstation> (zuletzt abgerufen am 16.2.2023).

²³ Stellungnahme ZEK (Fn. 7), A 2.

²⁴ Gassner/Juknat (Fn. 15), Kap. 3 Rn. 353 ff.

²⁵ A. Hannun/P. Rajpurkar/M. Haghighpanahi/G. Tison/C. Bourn/M. Turakhia/A. Ng, Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using deep neural network, *Nat Med* 2019; 25(1), S. 65 ff.

²⁶ V. Gulshan/L. Peng/M. Coram *et al.*, Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs, *JAMA*. 2016, 316(22), S. 2402–2410.

²⁷ A. Esteva/B. Kuprel/R. Novoa/J. Ko/S. Swetter/H. Blau/S. Thrun, Dermatologist-level classification of skin cancer with deep neural networks, *Nature* 2017, S. 115 ff.

²⁸ Mit weiteren Beispielen Sutton/Pincock/Baumgart *et al.* (Fn. 2), S. 6; Gassner/Juknat (Fn. 15), Kap. 3 Rn. 353 ff.; Stellungnahme ZEK (Fn. 7), A 3.

²⁹ B. Zippel-Schultz/D. Müller-Wieland/A. Remppis *et al.*, Künstliche Intelligenz in der Kardiologie, *Herzschr Elektrophys* 2021, S. 89 (94).

³⁰ M. Cikes/S. Sanchez-Martinez/B. Claggett *et al.*, Machine Learning-based phenotyping in heart failure to identify responders to cardiac resynchronization therapy, *Eur J Heart Fail*. 2019, 21 (1), S. 74–85.

³¹ Stellungnahme ZEK (Fn. 7), A 2; siehe exemplarisch C. Cantwell/Y. Mohamied/K. Tzortzis *et al.*, Rethinking multiscale cardiac electrophysiology with machine learning and predictive modelling, *Computers in Biology and Medicine* 2019, S. 339–351.

wickelt worden sind (z. B. Detektion von MS-Läsionen anhand von MRT-Scans, Detektion von Hautläsionen anhand von Fotos, Identifikation von Arrhythmien im EKG, etc.). Ihr Einsatz für ein anderes – wenn auch ähnlich gelagertes – Problem kommt daher bislang nicht in Betracht.³²

III. Risiken des Einsatzes KI-basierter CDSS

Die Anwendungsbereiche von KI-basierten CDSS machen deutlich, wie entscheidend eine gute Performance dieser Systeme für Patientinnen ist. Denn auch, wenn sie keine unmittelbare Wirkung auf den Patienten haben, weil es sich – jedenfalls nach aktuellem Stand – letztendlich um Entscheidungsunterstützungssysteme handelt, die Ärztinnen lediglich Vorschläge oder eine Orientierungshilfe bietet, kann ein fehlerhafter Output eines solchen Systems existenziell für den Patienten sein, zumindest aber zu einer nicht unerheblichen gesundheitlichen Beeinträchtigung führen, etwa wenn ein nicht passender Therapieansatz empfohlen wird oder Areale fälschlicherweise als krankes Gewebe identifiziert und daraufhin chirurgisch entfernt werden. Das gilt insbesondere vor dem Hintergrund, dass zum einen Untersuchungen zeigen, dass Ärztinnen verstärkt unhinterfragt auf Vorschläge von automatisierten Entscheidungssystemen vertrauen, sobald sie in einigen Anwendungsszenarien erfolgreich waren – sie also einem sog. *automation bias* unterliegen;³³ zum anderen, weil es Ärzten bei KI-gestützten Systemen nicht möglich ist, zu verstehen, wie und warum ein bestimmter Output vom System generiert wurde, was sie dann ggfs. noch zu einer korrigierenden Handlung veranlassen könnte.³⁴ Zwar bestehen unter den Begriffen XML/XAI (*Explainable AI*) Ansätze insbesondere im Bereich der *computer vision*, den Einfluss verschiedener Eingabedaten (bspw. Pixel in einem Bild) auf die Entscheidung des Modells zu bestimmen und anzuzeigen.³⁵ Bei komplexen Gewichtungsvorgängen, insb. beim *Deep Learning*, ist hingegen unklar, ob diese jemals nachvollzogen werden können.³⁶

Risiken durch den Einsatz KI-basierter CDSS bestehen für Patientinnen daher insb. dann, wenn das System für den Nutzer nicht erkennbar diskriminierende Out-

³² Z. B. kann ein Computerprogramm, das speziell für die Erkennung von Tumoren in der Lunge entwickelt wurde, ein Pneumothorax im MRT-Scan übersehen.

³³ R. Challen/J. Denny/M. Pitt et al., Artificial intelligence, bias and clinical safety, *BMJ Qual Saf* 2019, S. 231 (234); T. Tsai/D. Fridsma/G. Gatti, Computer decision support as a source of interpretation error: the case of electrocardiograms, *J Am Med Inform Assoc.* 2003, S. 478–483, doi: 10.1197/jamia.M1279.

³⁴ Vgl. zur Blackbox-Problematik bereits Fn. 15 m.w.N.

³⁵ Vgl. zu den verschiedenen existierenden Ansätzen *bikom* (Fn. 15), S. 13 ff.; M. Höhne, Nachvollziehbare Künstliche Intelligenz: Methoden, Chancen und Risiken, *DuD* 2021, S. 453 ff.; A. Adadi/M. Berrada, Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI), *IEEE* 2018, S. 52138 ff.; A. Holzinger, Interpretierbare KI, c't 2018, S. 136 (137 ff.).

³⁶ Gassner/Juknat (Fn. 15), Kap. 3 Rn. 374.

puts zu Lasten bestimmter Patientinnengruppen hervorbringt, bspw. mit der Folge, dass diese schlechteren Zugang zu einer medizinischen Behandlung haben. Das kann insb. dann der Fall sein, wenn – was nicht unwahrscheinlich ist – die zum Training des Systems verwendeten Datensätze bereits mit einem *Bias* aufgrund gesellschaftlicher Vorannahmen, Vorurteile etc. behaftet sind, die dann vom System fortgeschrieben und reproduziert werden.³⁷ Diskriminierungen können aber auch dadurch entstehen, dass die genutzten Datensätze einseitig geprägt sind, also bestimmte Personengruppen (z. B. Kinder, Frauen, *People of Color*) darin statistisch unterrepräsentiert sind (etwa weil es von dieser Personengruppe nicht genügend Datensätze gibt)³⁸, so dass Spezifika dieser Gruppe nicht ausreichend gelernt werden konnten.³⁹ Bekanntestes Beispiel ist hier die schlechtere Performance von Systemen zur Erkennung von Hautkrebs bei *People of Color*, da die Datensätze, mit denen das System trainiert worden ist, hauptsächlich von Menschen mit heller Hautfarbe stammten.⁴⁰ Risiken gehen aber auch von sonstigen fehlerhaften Entscheidungen von CDSS aus, wobei fehlerhaft immer im Kontext der jeweiligen Aufgabe zu verstehen ist und insb. bei Prognoseentscheidungen nur schwer zu operationalisieren ist. Die Fehler können daraus resultieren, dass gewichtige medizinische Faktoren für das zu lösende Problem vom System nicht berücksichtigt wurden (z. B. das Alter) oder umgekehrt Faktoren besonders ins Gewicht fallen, die nicht relevant sind, aber zufällig zu einer Korrelation führen. Aber auch Fehler in den Daten, wie bspw. Messfehler, oder deren fehlerhafte Aufbereitung, z. B. durch fehlerhafte Annotationen, können sich in negativer Weise auf die „Richtigkeit“ des Outputs auswirken.⁴¹ Schließlich können Risiken immer auch durch die Wahl eines von vornherein für die zu lösende Aufgabe (Klassifikation, Segmentierung etc.) nicht geeigneten Modells oder durch den nichtbestimmungsgemäßen Gebrauch durch den Anwender hervorgerufen werden. Denn die Qualität der Ergebnisse und letztlich auch die Gefahrenpotentiale solcher Technologien hängen entscheidend immer auch davon ab, wie gut sie in den organisationalen Kontext der Organisation, in dem sie zur Anwendung kommen, eingebettet sind.⁴² Werden solche Systeme also bspw. auf Daten angewendet, die nicht Bestand-

³⁷ Bericht der Enquete-Kommission Künstliche Intelligenz (Fn. 15), BT-Drs. 19/23700, S. 61; Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS, Leitfaden zur Gestaltung vertrauenswürdiger Künstlicher Intelligenz. KI-Prüfkatalog, 2021, S. 37; Schneider (Fn. 7), S. 331 f.

³⁸ Vgl. zu den unterschiedlichen Gründen R. Baumgartner, Künstliche Intelligenz in der Medizin: Diskriminierung oder Fairness?, in: Diskriminierung und Antidiskriminierung, 2021, S. 156.

³⁹ Bericht der Enquete-Kommission Künstliche Intelligenz (Fn. 15), BT-Drs. 19/23700, S. 61; Stellungnahme ZEKO (Fn. 7), A 7; Gassner/Juknat (Fn. 15), Kap. 3 Rn. 371.

⁴⁰ A. Adamson/A. Smith, Machine Learning and Health Care Disparities, JAMA. Dermatology 2018, S. 1247; W. Glauser, AI in health care: Improving outcomes or threatening equity?, CMAJ 2020, 1 E21 f.

⁴¹ Stellungnahme ZEKO (Fn. 7), A 7.

⁴² Siehe dazu auch Deutsche Normungsroadmap Künstliche Intelligenz, Ausgabe 2, 2022, S. 229.

teil des Trainingsdatensatzes waren (bspw. Scans von MRT-Geräten anderer Hersteller oder anderer Generation; andere Population) oder für Aufgaben, für die sie nicht entwickelt worden sind, können sie Fehlklassifikationen oder fehlerhafte Prognosen trotz fehlerfreier Entwicklung hervorbringen.

IV. Anforderungen an den Einsatz KI-basierter CDSS

Daraus lassen sich Anforderungen und Qualitätsstandards ableiten, die an den Einsatz KI-basierter CDSS in der Versorgung zu stellen sind und sich bereits in diversen Positionspapieren⁴³ zu vertrauenswürdiger KI im Allgemeinen einschließlich dem Entwurf des AI-Acts⁴⁴ finden, im medizinischen Bereich aber aufgrund des besonderen Schadenspotentials von besonderer Relevanz sind. Die Anforderungen lassen sich unterteilen in Qualitätsanforderungen, die das KI-System selbst betreffen (1.), sowie Anforderungen an die Implementierung in den Klinikalltag (2.).

1. Qualitätsanforderungen an das KI-System

Entscheidendes Qualitätsmerkmal für den Einsatz solcher Systeme ist zunächst, dass sie sicher, zuverlässig und technisch robust funktionieren. Das bedeutet, es muss zum einen sichergestellt sein, dass das System mit einer Reihe von Eingaben und in verschiedenen Situationen einwandfrei funktioniert, insb. dass die Ausgaben selbst unter kleinen Veränderungen der Eingabedaten sowie bei Unsicherheiten konstant bleiben (Robustheit).⁴⁵ Zudem muss das System in der Lage sein, seine Ergebnisse auch durch andere Anwenderinnen bei gleichen Bedingungen zu reproduzieren.⁴⁶ Denn nur so kann gewährleistet werden, dass sie zuverlässig agieren und auch von verschiedenen Einrichtungen verwendet werden können. Dazu gehört auch, dass diese Systeme – wie im Übrigen alle Softwaresysteme – möglichst sicher sind, sowohl was deren eigene Funktionalität als auch ihre Integrität nach außen betrifft (Sicherheit und Integrität). Angriffe und Sicherheitsverletzungen, etwa in Form von Datenfälschungen oder Systemmanipulationen, müssen durch entsprechende Si-

⁴³ Vgl. *Hochrangige Expertengruppe für Künstliche Intelligenz*, Ethikleitlinien für eine vertrauenswürdige KI, 2019; *dies.*, The Assessment List For Trustworthy Artificial Intelligence (ALTAI), 2020; *Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS*, Vertrauenswürdiger Einsatz von Künstlicher Intelligenz, 2019, S. 15 ff.; explizit für Medizinproduktesoftware vgl. *European Coordination Committee of Radiological, Electromedical and Healthcare IT Industry (COCIR)*, Artificial Intelligence in EU Medical Device Regulation, 2021.

⁴⁴ Vorschlag für eine Verordnung zur Feststellung harmonisierter Vorschriften für künstliche Intelligenz und zur Änderung bestimmter Rechtsakte der Union, COM(2021) 206 final.

⁴⁵ Vgl. *Fraunhofer IAIS*, Leitfaden zur Gestaltung vertrauenswürdiger Künstlicher Intelligenz (Fn. 37), S. 25; Ethikleitlinien für eine vertrauenswürdige KI (Fn. 43), S. 21; *COCIR* (Fn. 43), S. 25.

⁴⁶ Vgl. Ethikleitlinien für eine vertrauenswürdige KI (Fn.43), S. 21; *COCIR* (Fn. 43), S. 24.

cherheitsvorkehrungen verhindert oder zumindest begrenzt werden.⁴⁷ Weiterhin ist elementar, dass die hervorgebrachten Ergebnisse sowohl ein hohes Maß an Genauigkeit als auch Präzision aufweisen, d. h. Objekte so oft wie möglich zutreffend klassifiziert bzw. Vorhersagen richtig getroffen werden.⁴⁸ Nur so kann – angesichts der bislang weitgehend fehlenden Möglichkeit der Kontrolle durch den Nutzer – das Risiko von Fehldiagnosen und -interventionen auf ein angemessenes Maß reduziert werden. Schließlich sollten CDSS-Systeme so beschaffen sein, dass sie Verzerrungen (*Bias*) auf ein Minimum reduzieren und nicht ungerechtfertigt, d. h. ohne sachlichen Grund diskriminieren.⁴⁹ Insbesondere für die Patientengruppen, für die sie entwickelt bzw. mit deren Daten sie trainiert worden sind, sollten sie unvoreingenommen arbeiten und keinen *Bias* enthalten. Das schließt ein, dass sich die Performanz und Qualität des Outputs solcher Systeme in Bezug auf bestimmte Personengruppen (bspw. *People of Color*), für die das System ebenso eingesetzt wird, nicht oder jedenfalls nicht ohne entsprechende Kenntlichmachung verringert.

2. Sachgemäße Implementierung in den Organisationsablauf und Befähigung der Nutzer

Da die Leistungsfähigkeit solcher Systeme immer auch von der richtigen Verwendung im jeweiligen Nutzungskontext abhängt, sollte zudem nicht nur gewährleistet sein, dass solche Systeme an der richtigen Stelle in den jeweiligen Organisationsablauf eingebettet sind, auch der bestimmungs- und sachgemäße Umgang durch die Anwenderin muss sichergestellt sein. Im Hinblick auf CDSS bedeutet das vor allem, dass die Systeme nur ihrer Zweckbestimmung entsprechend verwendet werden, also insb. nicht für eine andere Aufgabenstellung (MS statt Tumordiagnostik) oder eine nicht adressierte Patientinnengruppe eingesetzt werden und die vom System generierten Informationen richtig interpretiert und genutzt werden. Ergebnisse eines lediglich zur Entscheidungsunterstützung entwickelten Systems dürfen also bspw. vom Anwender nicht als Erstdiagnose missverstanden werden; auf Alarme, die das System generiert, muss das verantwortliche Klinikpersonal zuverlässig und angemessen reagieren.

⁴⁷ Vgl. Ethikleitlinien für eine vertrauenswürdige KI (Fn.43), S. 20; vgl. auch *Fraunhofer IAIS*, Leitfaden zur Gestaltung vertrauenswürdiger Künstlicher Intelligenz (Fn. 37), S. 26.

⁴⁸ Vgl. Ethikleitlinien für eine vertrauenswürdige KI (Fn. 43), S. 21; Weißbuch zur Künstlichen Intelligenz – ein europäisches Konzept für Exzellenz und Vertrauen, COM(2020) 65 final, S. 24.

⁴⁹ Vgl. Ethikleitlinien für eine vertrauenswürdige KI (Fn. 43), S. 22 f.; *Fraunhofer IAIS*, Leitfaden zur Gestaltung vertrauenswürdiger Künstlicher Intelligenz (Fn. 37), S. 23.

V. CDSS unter der Medical Device Regulation

Mit der Medizinprodukteverordnung (MDR)⁵⁰ existiert bereits ein Rechtsrahmen, der Gesundheitsschutz und Patientensicherheit insb. explizit auch bei Verwendung von Software sicherstellen soll, indem nur solche Produkte in Verkehr gebracht oder in Betrieb genommen werden dürfen, die bestimmten Sicherheitsanforderungen genügen. Inwieweit dieser Rechtsrahmen allerdings auch geeignet ist, den besonderen Funktionslogiken KI-basierter Software gerecht zu werden und ob er die an diese geknüpften Qualitätserwartungen hinreichend adressiert, soll im Folgenden – auch mit Blick darauf, dass sich die Frage des Inverkehrbringens selbst bei Inkrafttreten des AI-Act auch weiterhin im Wesentlichen nach den Maßgaben der MDR richten wird⁵¹ – untersucht werden. Spezielle, die Besonderheiten von KI-basierter Software adressierende Bestimmungen enthält die MDR bis dato jedenfalls nicht⁵² und sind auch angesichts der erst für 2027 geplanten Evaluation vorerst nicht zu erwarten, weshalb es entscheidend auf den darin enthaltenden Rechtsrahmen für Software-Medizinprodukte ankommt.

I. CDSS als Medizinprodukt i. S. d. MDR

CDSS werden in aller Regel als Medizinprodukte i. S. d. MDR einzuordnen sein, die dann nur unter Einhaltung der Anforderungen der MDR in Verkehr gebracht werden dürfen.⁵³ Denn nach Art. 2 Nr. 1 MDR sind alle Gegenstände – explizit auch Software als solche – Medizinprodukte, die dem Hersteller zufolge für Menschen bestimmt sind und der Erfüllung eines spezifischen medizinischen Zwecks dienen, was bei CDSS, die spezifisch zur Unterstützung der Diagnostik, Prognose, Überwachung oder zur Therapie entwickelt worden sind, regelmäßig der Fall sein wird.⁵⁴ Der

⁵⁰ Verordnung (EU) 2017/745 des Europäischen Parlaments und des Rates vom 5. April 2017 über Medizinprodukte, zur Änderung der Richtlinie 2001/83/EG, der Verordnung (EG) Nr. 178/2002 und der Verordnung (EG) Nr. 1223/2009 und zur Aufhebung der Richtlinien 90/385 EWG und 93/42/EWG des Rates.

⁵¹ Noch nicht abschließend geklärt ist allerdings, ob die Vorgaben des AI-Acts ergänzend in die Bewertung nach der MDR einbezogen werden oder ob es der Durchführung eines weiteren, zusätzlichen Konformitätsbewertungsverfahrens nach dem AI-Act bedarf, vgl. *M. Heil*, Innovationsermöglichungsrecht oder Innovationshemmnis? Regulatorische Herausforderungen für KI-basierte Medizinprodukte-Software in der EU, in: FS Gassner, 2022, S. 456; *M. Fuderer*, Doppelte Konformitätsbewertung bei KI-basierten Medizinprodukten?, MPR 2022, S. 121 ff.

⁵² *U. Gassner*, Intelligente Medizinprodukte – Regulierungsperspektiven und Zertifizierungspraxis, MPR 2021, S. 41 (44); *Heil* (Fn. 51), S. 451; Deutsche Normungsroadmap KI (Fn. 42), S. 229; *Z. Schreitmüller*, Regulierung intelligenter Medizinprodukte – Eine Analyse unter besonderer Berücksichtigung der MPVO und DSGVO, 2023, S. 109.

⁵³ Siehe dazu auch *H. Dettling*, Künstliche Intelligenz und digitale Unterstützung ärztlicher Entscheidung in Diagnostik und Therapie, PharmR 2019, 633 (636); Stellungnahme ZEKO (Fn. 7), A 8.

⁵⁴ *Steinwendner* (Fn. 5), S. 694.

EuGH ordnete jedenfalls eine Software, die Patientendaten mit Medikamenten abgleicht, die die Ärztin verschreiben möchte, und so in der Lage ist, ihm in automatisierter Form eine Analyse über etwaige Kontraindikationen, Wechselwirkungen von Medikamenten und Überdosierungen zu liefern, als Medizinprodukt ein.⁵⁵ Nicht entscheidend ist, ob die Software der Steuerung oder Beeinflussung eines anderen Medizinprodukts dient – auch eine sog. *Stand-Alone-Software* kann Medizinprodukt i. S. d. MDR sein.⁵⁶ Ebenso wenig kommt es darauf an, ob die Software unmittelbar im oder am menschlichen Körper wirkt; vielmehr ist nach der Rspr. des EuGH der Verwendungszweck des Produkts für die Einordnung maßgeblich.⁵⁷

2. (Qualitäts-)Anforderungen an Medizinproduktesoftware nach der MDR

Bereits heute knüpft die MDR das Inverkehrbringen von Software – unabhängig ob KI-basiert oder nicht – regulatorisch an die Erfüllung bestimmte (Qualitäts-)Anforderungen, die die unter IV. für KI-basierte CDSS herausgearbeiteten Erwartungen an deren Performance adressieren. Insbesondere die in Anhang I der Verordnung normierten grundlegenden Sicherheits- und Leistungsanforderungen, denen Medizinprodukte nach Art. 5 Abs. 2 MDR zu genügen haben, formulieren entsprechende Anforderungen.⁵⁸ So müssen Softwareprodukte zunächst zum einen konkret so ausgelegt sein, dass *Wiederholbarkeit*, *Zuverlässigkeit* und *Leistung* entsprechend ihrer bestimmungsgemäßen Verwendung gewährleistet sind (Anhang I 17.1. MDR). Zum anderen müssen sie so entwickelt und hergestellt sein, dass sie dem *Stand der Technik* entsprechen, wobei sie die Grundsätze des *Software-Lebenszyklus*, des *Risikomanagements* einschließlich der *Informationssicherheit*, zu berücksichtigen haben (Anhang I 17.2. MDR). Formuliert sind damit folglich nicht nur bereits Anforderungen an die Zuverlässigkeit und Fehlertoleranz von Software, die überdies dem anerkannten, gegenwärtigen Entwicklungsstand des technischen Könnens zu entsprechen hat, sondern auch Erwartungen an die IT-Sicherheit solcher Systeme. So soll durch den Einschub „einschließlich Informationssicherheit“ in Anhang I 17.2. MDR gewährleistet werden, dass bereits im Rahmen des nach Art. 10 Abs. 2, Anhang I Nr. 3 und 4 MDR für alle Medizinprodukte grds. einzurichtenden und anzuwendenden Risikomanagements sichergestellt wird, dass es nicht durch IT-Sicherheitslücken zu Patientenrisiken kommt.⁵⁹

⁵⁵ EuGH, EuZW 2018, S. 166 (167).

⁵⁶ Vgl. Definition von Medical Device Software in MDCG 2019–11, Guidance on Qualification and Classification of Software Regulation (EU) 2017/745 – MDR and Regulation (EU) 2017/745. S. 6 f.; Gassner (Fn. 52), S. 42.

⁵⁷ EuGH, EuZW 2018, S. 166 (167).

⁵⁸ So auch i.E. COCIR (Fn. 43), S. 21 ff.

⁵⁹ M. Hastenteufel/S. Renaud, Software als Medizinprodukt. Entwicklung und Zulassung von Software in der Medizintechnik, 2019, S. 8, 94 f.

Weiterhin setzen die Leistungsanforderungen eine *Validierung und Verifizierung* der Software voraus (Anhang I 17.2. MDR). Das bedeutet, vor Inverkehrbringen muss nicht nur nachgewiesen sein, dass die Software die versprochenen Nutzungsziele entsprechend ihrer Zweckbestimmung erreicht;⁶⁰ ein Bildbearbeitungsalgorithmus MS-Läsionen also bspw. mit ausreichender Wahrscheinlichkeit auch tatsächlich in MRT-Scans erkennt (Validierung). Durch die Bereitstellung eines objektiven Nachweises muss auch bestätigt sein, dass spezifische an die konkrete Software zu stellende Eigenschaften (z. B. Interoperabilität) nach dem Stand der Technik erfüllt sind (Verifikation). Schließlich sieht die MDR speziell für Produkte, die der Diagnose dienen oder denen eine Messfunktion innewohnt, die Gewährleistung ausreichender Genauigkeit, Präzision und Stabilität vor (Anhang I 15.1. MDR). Damit wird – zumindest was diagnostische Systeme betrifft – konkret den unter IV. beschriebenen Performanz-Kriterien, was die Häufigkeit der Richtigkeit der Ergebnisse betrifft, Rechnung getragen.

3. Nachweis der Erfüllung der Sicherheits- und Leistungsanforderungen

Wann diese im Anhang I der Verordnung aufgeführten Sicherheits- und Leistungsanforderungen im Einzelnen als erfüllt anzusehen sind – also eine Software bspw. als *zuverlässig* i.S.d MDR gilt oder eine hinreichende Validierung vorliegt – regelt die Verordnung selbst allerdings nicht.⁶¹ Dies erfolgt vielmehr durch eine ganze Reihe von medizin(product)spezifischen technischen Normen und Standards sowie Prozessnormen seitens nationaler und internationaler Normungsorganisationen,⁶² die insoweit die abstrakten Anforderungen der MDR – z. T. spezifiziert für einen konkreten Produkttyp – konkretisieren und insoweit materiell-rechtlich ausfüllen.⁶³ So spezifiziert bspw. die ISO 14971⁶⁴ Maßgaben an das gemäß Art. 10 Abs. 2, Anhang I Nr. 3 und 4 MDR für alle Medizinprodukte grds. einzurichtende und anzuwendende Risikomanagementsystem, die ISO 13485⁶⁵ zentrale Aspekte des Qualitätsmanagements und die IEC 62304⁶⁶ die Mindestanforderungen an die wichtigsten Software-Lebenszyklus-Prozesse (Anhang I 17.2. MDR).

⁶⁰ Vgl. *Hastenteufel/Renaud* (Fn. 59), S. 171.

⁶¹ *Gassner* (Fn. 52), S. 45.

⁶² Z. B. solche des International Organization for Standardization (ISO) oder des Deutschen Instituts für Normung (DIN); exemplarisch seien hier die ISO 13485 (Medizinprodukte – Qualitätsmanagementsystem – Anforderungen an regulatorische Zwecke), IEC 62304 (Medizingeräte-Software – Software-Lebenszyklus-Prozesse) genannt.

⁶³ Vgl. *Hastenteufel/Renaud* (Fn. 59), S. 43; *Pechmann/Mildner/Suthau/Leucker* (Fn. 8), S. 178; *L. Reitebuch*, *Mobile Health Applications. Rechtliche Rahmenbedingungen von Gesundheits-Apps*, 2022, S. 121.

⁶⁴ *Medical Devices – Application of risk management to medical Devices*.

⁶⁵ *Medizinprodukte – Qualitätsmanagementsysteme – Anforderungen für regulatorische Zwecke*.

⁶⁶ *Health Software – Software life cycle processes*.

Für KI-basierte Systeme enthalten diese Normen bislang jedoch keine spezifischen Anforderungen. Insgesamt fehlt es – ebenso wie auf legislativer Ebene – bis dato noch an harmonisierten Normen, die einzuhaltende Prozessschritte oder konkrete Anforderungen an die Entwicklung und Testverfahren, mithilfe derer die Konformität mit der MDR nachgewiesen werden kann, beschreiben. Zwar existieren inzwischen einige horizontale Regelwerke (z. B. Normen der IEEE-7000er Serie)⁶⁷ sowie eine unübersichtliche Reihe weiterer entsprechender Initiativen und Entwürfe.⁶⁸ Kernproblem ist jedoch, dass in der Fachdisziplin selbst noch keine Einigkeit über die Methodik der Validierung und Verifizierung von KI-Systemen herrscht und es somit nach wie vor an verbindlich beschriebenen Prüfverfahren fehlt. Hersteller von Medizinproduktesoftware sind somit darauf angewiesen, selbst Nachweise für die Konformität ihre Produkte zu erbringen. Da KI-Systeme letztendlich vergleichbar einer Software sind, kann auch hier zunächst auf die harmonisierten Normen für Software zurückgegriffen werden, ergänzt durch bestehende inoffizielle Leitlinien für den Konformitätsbewertungsprozess KI-basierter Systeme, wie bspw. den Fragenkatalog „Künstliche Intelligenz bei Medizinprodukten“ der Interessengemeinschaft der Benannten Stellen für Medizinprodukte in Deutschland (IG NB)⁶⁹. Er gilt mit seinen Anforderungen zum jetzigen Zeitpunkt in Deutschland als maßgebliche Benchmark für den Bewertungsprozess⁷⁰ und soll daher im Folgenden näher betrachtet werden.

a) Klinische Bewertung

Wie bei jedem anderen Medizinprodukt hat die Bestätigung, dass die Software die beschriebenen grundlegenden Sicherheits- und Leistungsanforderungen erfüllt, im Rahmen einer klinischen Bewertung zu erfolgen, Artt. 5 Abs. 3, 10 Abs. 3, 61 Abs. 1 MDR. Beschrieben ist damit ein systematischer und geplanter Prozess der Generierung, Sammlung, Analyse und Bewertung klinischer Daten in einem Umfang, der für den Nachweis der Erfüllung der Sicherheits- und Leistungsanforderungen einschließlich des klinischen Nutzens des Produkts ausreicht, vgl. Art. 2 Nr. 44 MDR. Für Medizinprodukte-Software findet sich eine gezielte Anleitung, wie eine klinische Bewertung den Anforderungen der MDR genügen kann, in der *Guidance on Clinical Evaluation (MDR)/Performance Evaluation (IVDR) of Medical*

⁶⁷ IEEE 7001:2021 (Standard for Transparency of Autonomous Systems); IEEE 7002:2022 (Standard for Data Privacy Process); IEEE 7007:2021 (Ontological Standard for Ethically driven Robotics and Automation Systems); IEEE 7005:2021 (Transparent Employer Data Governance).

⁶⁸ Einen guten Überblick bietend *Gassner* (Fn. 52), S. 45 ff.; *D. Reinsch*, Regulatorische Anforderungen an Medizinprodukte mit Maschine Learning, abrufbar unter <https://www.johner-institut.de/blog/regulatory-affairs/regulatorische-anforderungen-an-medizinprodukte-mit-machine-learning/> (zuletzt abgerufen am 25.2.23).

⁶⁹ *Interessengemeinschaft der Benannten Stellen für Medizinprodukte in Deutschland (IG NB)*, Questionnaire „Artificial Intelligence (AI) in medical devices“, Version 4 v. 9.6.2022.

⁷⁰ *Gassner* (Fn. 52), S. 47; Deutsche Normungsroadmap KI (Fn. 42), S. 226.

*Software der Medical Device Coordination Group (MDCG 2020–1)*⁷¹, deren Prinzipien im Grundsatz auch für die Nachweisführung bei KI-basierter Software – jedenfalls sofern diese nicht weiterlernend ist – geeignet sind.⁷²

Im Kern sieht diese Anleitung für die Ermittlung der klinischen Evidenz drei Schlüsselemente vor: den Nachweis wissenschaftlicher Validität, die technische und analytische Leistungsfähigkeit sowie den Nachweis der klinischen Leistungsfähigkeit.⁷³ Mit dem Beweis *wissenschaftlicher Validität* hat der Hersteller den Nachweis zu erbringen, dass der Output der Software (z. B. Messwert, Score) mit einem klinischen oder physiologischen Zustand verknüpft ist; der ausgewiesene Risikoscore also bspw. tatsächlich ein Maß für ein bestimmtes Risiko ist.⁷⁴ Damit wird sichergestellt, dass nur Systeme zum Einsatz kommen, bei denen nachweisbar ist, dass deren berechnete Ausgabedaten auch eine klinische Aussagekraft haben (z. B. bei einem Gerät, welches Herzinfarkte anhand von Hebungen der ST-Strecke im EKG erkennt, dass diese Hebungen mit Herzinfarkten assoziiert sind). Der Nachweis gilt als geführt, wenn die Medizinproduktesoftware der klinischen Situation, dem Zustand, der Indikation oder dem Parameter entspricht, die in ihrer Zweckbestimmung definiert sind.⁷⁵ Dies setzt voraus, dass die wissenschaftliche Validität klinisch anerkannt oder begründet ist. Als Nachweis können *peer reviewed articles*, Leitlinien wissenschaftlicher medizinischer Fachgesellschaften, veröffentlichte Ergebnisse anderer klinischer Studien etc. ausreichen.⁷⁶ Bei KI-basierten Systemen wird sich der Nachweis aus den benannten Quellen jedoch regelmäßig nicht ohne weiteres führen lassen, jedenfalls sofern nicht bereits Daten zu vergleichbaren Produkten vorliegen. Da bei diesen Systemen vielfach nicht nachvollzogen werden kann, wie und warum ein Output auf Basis der Inputdaten generiert wird, In- und Output also nicht linear miteinander verbunden sind, wird sich die klinische Aussagekraft der von ihnen hervorgebrachten Ergebnisse vielfach nicht unmittelbar aus anderen, insb. nicht zu KI-basierten Produkten existierenden Studien oder Fachveröffentlichungen ergeben können. Insofern kann es für den Hersteller erforderlich sein, eine Sekundärdatenanalyse oder eine klinische Studie bzw. eine Leistungsstudie durchzuführen, sofern

⁷¹ Basierend auf Konzepten der Guidance on Clinical Evaluation of Software as a Medical Device (SaMD) des International Medical Device Regulators Forum (IMDRF) zur Verbesserung globaler Konvergenz, vgl. MDCG 2020–1, S. 3.

⁷² So auch *COCIR* (Fn. 43), S. 22; *Pechmann/Mildner/Suthau/Leucker* (Fn. 8), S. 188.

⁷³ Siehe hierzu auch *U. Gassner/T. Schreiegg*, Klinische Bewertung und Leistungsbewertung von Medizinprodukte-Software – normative Integration und globale Konvergenz, MPR 2020, S. 104 (106 f.); *Hastenteufel/Renaud* (Fn. 59), S. 63 f.; *B. Martin*, Klinische Bewertung von Software: Drei Beweise für die Konformität, abrufbar unter <https://www.johner-institut.de/blog/regulatory-affairs/klinische-bewertung-von-software/> (zuletzt abgerufen am 27.1.23).

⁷⁴ MDCG 2020–1, S. 10.

⁷⁵ *Gassner/Schreiegg* (Fn. 73), S. 107.

⁷⁶ MDCG 2020–1, S. 12.

er dies nicht – wie vielfach üblich – ohnehin bereits im Rahmen der Entwicklung getan hat.⁷⁷

Mit der zweiten Nachweiskomponente, der *technischen und analytischen Leistungsfähigkeit*, muss hingegen das technische Leistungsvermögen demonstriert werden. Konkret gilt es die Fähigkeit der Software, aus den Eingabedaten genau, zuverlässig und präzise den beabsichtigten Output zu erzeugen, nachzuweisen.⁷⁸ Für auf maschinellem Lernverfahren basierende Systeme kommen dabei je nach Aufgabenstellung verschiedene Performanz-Metriken in Betracht, anhand derer sich deren technische Leistungsfähigkeit quantifizieren und beurteilen lässt: *Accuracy*, Spezifität und Sensitivität, *Precision and Recall*, *F1-Score* bspw. für Klassifikationsaufgaben, *Mean Squared Error* und *Mean Absolute Error* für Regressionsprobleme.⁷⁹ Zudem kann mithilfe einer *Loss*-Funktion Aufschluss über die Qualität des Trainings (z. B. mögliches *Overfitting*) erlangt werden.⁸⁰ Daneben gilt es, die Güte und die Grenzen der Vorhersage eines solchen Systems (bspw. an medizinischen Randfällen im Vergleich zu Normalbefunden) zu analysieren und möglichst adäquat zu beschreiben.⁸¹

Der dritte Nachweis über die *klinische Leistungsfähigkeit* zielt schließlich auf den Realeinsatz der Software. Der Hersteller muss nachweisen, dass die Software die Fähigkeit besitzt, klinisch relevante Ergebnisse entsprechend dem beabsichtigten Verwendungszweck zu liefern,⁸² z. B. muss ein KI-basiertes System zur Kariesdiagnostik tatsächlich in einer zahnmedizinischen Praxis Karies auf Röntgenbildern detektieren. Dazu gehören u. a. Informationen über die klinische/diagnostische Sensitivität und/oder Spezifität, die Anzahl der Patienten, die aufgrund einer Erkrankung oder präventiv behandelt werden müssen, um ein zusätzliches Ereignis wie (Folge-)Erkrankung oder Tod zu vermeiden sowie über das positive wie negative Wahrscheinlichkeitsverhältnis.⁸³ Diese Daten können durch präklinische Tests, eigenständige klinische Prüfungen (Art. 2 Nr. 45 MDR) oder eine klinische Leistungsstudie generiert werden, wobei die Frage, welche Nachweise konkret zu erbringen sind, jeweils

⁷⁷ Vgl. *Gassner/Schreiegg* (Fn. 73), S. 107, für den Fall, dass keine anderen Daten vorhanden sind.

⁷⁸ Dazu zählen im Einzelnen Software-Eigenschaften wie Vertraulichkeit, Integrität, Zuverlässigkeit, Genauigkeit, analytische Sensitivität, Nachweisgrenze, Quantifizierungsgrenze, analytische Spezifität, Linearität, Grenzwert(e), Messintervall, Abwesenheit inakzeptabler Cybersicherheitslücken etc., vgl. *Gassner/Schreiegg* (Fn. 73), S. 107.

⁷⁹ Vgl. *Fraunhofer IAIS*, Leitfaden zur Gestaltung vertrauenswürdiger Künstlicher Intelligenz (Fn. 37), S. 91; siehe dazu auch Normungsroadmap KI (Fn. 42), S. 230.

⁸⁰ *Fraunhofer IAIS*, Leitfaden zur Gestaltung vertrauenswürdiger Künstlicher Intelligenz (Fn. 37), S. 90.

⁸¹ Siehe Deutsche Normungsroadmap KI (Fn. 42), S. 230; Fragenkatalog *IG NB* (Fn. 69), B Unterkapitel 4 lit. c) Frage 3 und 4.

⁸² MDCG 2020–1, S. 10 und 12.

⁸³ MDCG 2020–1, S. 14.

vom intendierten Nutzen der Software abhängt.⁸⁴ Es besteht aber auch die Möglichkeit, auf Daten von Äquivalenten zurückzugreifen. Außerdem muss die Validierung der klinischen Leistung zeigen, dass die Medizinproduktesoftware benutzerfreundlich ist, also bspw. ein benutzerfreundliches Interface aufweist.

b) Ergänzende Nachweise speziell für KI-basierte Systeme

Wie die Sicherheit- und Leistungsfähigkeit speziell von KI-basierter Medizinproduktesoftware durch Hersteller nachgewiesen werden kann, findet sich für Deutschland exemplarisch im – wenn auch nicht verbindlichen – Fragenkatalog „Künstliche Intelligenz bei Medizinprodukten“ der IG NB⁸⁵. Er verfolgt einen prozessorientierten Ansatz, d. h., statt spezifischen Anforderungen an die Medizinproduktesoftware definiert er Leitlinien für alle relevanten Prozesse und Phasen des Lebenszyklus einer KI-Software. Dabei referenziert er weitgehend auf die bereits existierenden technischen Normen sowie die beschriebenen Vorgaben des Anhang I der MDR und ergänzt diese durch spezifische für die Validierung von KI-Systemen relevante Komponenten. Insbesondere die im zweiten Kapitel des Fragenkatalogs (Anforderungen an die Produktentwicklung) aufgeführten Anforderungen an das *Datenmanagement* und die *Modellentwicklung* dienen der Mitigation von Risiken, die spezifisch für mithilfe maschinellen Lernverfahrens entwickelter Systeme sind, und die dem entsprechend bislang nicht von den für Software allgemein existierenden verbindlichen Normen adressiert sind.

So wird mit dem Datenmanagement sowohl dem bei KI-Systemen bestehenden Risiko einer möglichen mangelnden Leistungsfähigkeit des Modells als auch möglichen Diskriminierungen begegnet. Denn da beides entscheidend von der Auswahl und Aufbereitung der für das Training verwendeten Daten abhängt, ist die Sicherstellung qualitativ hochwertiger und für das Anwendungsziel relevanter Datensätze von entscheidender Bedeutung.⁸⁶ Entsprechend folgerichtig statuiert der Fragenkatalog daher – ähnlich dem AI-Act-Entwurf⁸⁷ – Anforderungen an die Qualität, Auswahl und Aufbereitung der Daten. Hierzu zählen vor allem Maßgaben an die Sammlung der Trainings-, Validierungs- und Testdatensätze (z. B. ausreichende Anzahl an Datensätze, Repräsentativität der Datensätze für die Zielpopulation, Analyse von möglichen *Bias* in den Datensätzen) sowie Prozessanforderungen an das *Labeling* sowie

⁸⁴ MDCG 2020–1, S. 11.

⁸⁵ Siehe Fn. 69.

⁸⁶ Vgl. hierzu *Fraunhofer IAIS*, Leitfaden zur Gestaltung vertrauenswürdiger Künstlicher Intelligenz (Fn. 37), S. 19 f.; *Kleesiek/Murray/Strack/Kaisiss/Braren* (Fn. 12), S. 26; *K. Strasser/B. Niedermayer*, Unvoreingenommenheit von Künstliche-Intelligenz-Systemen. Die Rolle von Datenqualität und Bias für den verantwortungsvollen Einsatz von künstlicher Intelligenz, in: *CSR und Künstliche Intelligenz*, 2021, S. 121 ff.

⁸⁷ Vgl. nur Art. 10 AI-Act-Entwurf (Fn. 44).

das Verfahren zur (Vor-)Verarbeitung von Daten.⁸⁸ So soll insb. für das überwachte Lernen sichergestellt werden, dass durch Anforderungen an das eingesetzte Personal (z. B. Anzahl der Personen, Kompetenz, entsprechende Schulungen) und den Prozess des *Labelings* (Verfahren zur Überprüfung, z. B. 4-Augen-Prinzip) das *Labeling* korrekt erfolgt. Zudem schließt das Datenmanagement datenschutzrechtliche Aspekte mit ein: Hersteller haben bspw. ein Verfahren zu etablieren, mit dem die Daten vor der Nutzung anonymisiert bzw. pseudonymisiert werden, um so den Anforderungen der DSGVO gerecht zu werden.⁸⁹ Nicht geklärt ist mit den Fragen zum Datenmanagement u. a. allerdings, ob bspw. auch synthetische Datensätze genutzt werden dürfen (oder gar müssen, um Repräsentativität zu erzeugen) und ab wann eine Datenmenge und -auswahl als ausreichend (repräsentativ) für eine bestimmte Anwendung betrachtet werden kann.⁹⁰

Auch mit den Fragen zur Modellentwicklung, mit der die Folgephase des Software-Lebenszyklus einer KI beschrieben ist, soll die Leistungsfähigkeit des Systems weitergehend abgesichert werden. Da dessen Qualitätseigenschaften maßgeblich durch die Wahl einer bestimmten Modellarchitektur für die konkrete Aufgabenstellung und durch das Training des Modells bestimmt werden, hat der Hersteller die von ihm bei der Modellentwicklung getroffenen Entscheidungen anhand von bestimmten Kriterien darzulegen. Dazu zählt zum einen insbesondere die Begründung der Auswahl der gewählten Modellarchitektur für den vorgesehenen *use-case* (z. B. die Wahl von *Convolutional Neural Networks* in Form von *Encoder-Decoder*-Architekturen für semantische Segmentierungen) orientiert am Stand der Technik, wobei stets zu begründen ist, warum nicht auf einfachere und interpretierbare Modelle zurückgegriffen wird.⁹¹ Zum anderen hat der Hersteller darzulegen, wie das Training des Modells erfolgt ist bzw. wie er dieses anhand von Gütemaßen bewertet hat.⁹² Hierfür hat er u. a. anzugeben, in welchem Verhältnis und nach welcher Stratifizierung er die ihm zur Verfügung stehenden Datensätze in Trainings-, Validierungs- und Testdaten aufgeteilt hat, einschließlich ob er sichergestellt hat, dass die Testdaten nicht bereits für das Training bzw. die Validierung des Modells verwendet wurden, da nur so eine valide Qualitätskontrolle des Modells durch Testung möglich ist.⁹³

Daneben finden sich in den weiteren Kapiteln des Fragenkatalogs vereinzelt weitere KI-spezifische Anforderungen, die die bereits für normale Software etablierten Nachweise und Prozessschritte um Einzelkomponenten ergänzen. Hierzu zählt etwa

⁸⁸ Vgl. Fragen unter B Unterkapitel 3 (Datenmanagement) lit. a-b des Fragenkatalogs (Fn. 69).

⁸⁹ Vgl. B Unterkapitel 3 (Datenmanagement) lit. a Frage 9 des Fragenkatalogs (Fn. 69).

⁹⁰ Vgl. Deutsche Normungsroadmap KI (Fn. 42), S. 241.

⁹¹ Vgl. B Unterkapitel 4 (Modellentwicklung) lit. c Frage 5 des Fragenkatalogs (Fn. 69); siehe auch entsprechend Deutsche Normungsroadmap KI (Fn. 42), S. 229 f.

⁹² Vgl. Fragen unter B Unterkapitel 4 (Modellentwicklung) lit. b-c des Fragenkatalogs (Fn. 69).

⁹³ Vgl. Fragen unter B Unterkapitel 4 (Modellentwicklung) lit. a des Fragenkatalogs (Fn. 69).

die Analyse, Identifikation und Bewertung von speziell durch die Verwendung von KI-Systemen auftretenden Risiken im Rahmen des einzusetzenden Risikomanagementsystems (Art. 10 Abs. 2, Anhang I 3 MDR)⁹⁴, die Darlegung von implementierten Sicherungsmaßnahmen gegen KI-spezifische Cyberattacken (wie *Poisoning Attacks*, *Evasion Attacks* oder *Modell Extraction*)⁹⁵ etwa im Rahmen des allgemeinen Nachweises der technischen Leistungsfähigkeit oder die Bestimmung der Kompetenzen mit Bezug zur KI der beteiligten Rollen (z. B. Entwickler, Statistiker, Modellspezialisten) durch den Hersteller.⁹⁶

c) Konformitätsbewertungsverfahren

Die Erfüllung der in der MDR genannten Anforderungen – insb. was die Sicherheits- und Leistungsanforderungen betrifft – muss der Hersteller von CDSS gemäß Art. 52 MDR durch ein Konformitätsbewertungsverfahren nachweisen, bevor er sie in Verkehr bringen kann. Da CDSS regelmäßig als Medizinprodukte der Risikoklasse IIa oder höher einzustufen sein werden,⁹⁷ muss die Konformitätsbewertung gemäß Art. 52 Abs. 3–6 i. V. m. Anhängen IX–XI MDR unter Mitwirkung eines akkreditierten privaten Sachverständigeninstituts als benannte Stelle erfolgen, die die Einhaltung der gesetzlichen Vorgaben durch Audits beim Hersteller sicherstellt und die Zertifizierung anhand von harmonisierten Normen vornimmt.⁹⁸

Die Zuordnung von CDSS mindestens zur Risikoklasse IIa folgt aus der (zu) weit gefassten Regel 11 des Anhangs VIII der MDR. Diese klassifiziert Software – unabhängig ob KI-basiert oder nicht – bereits dann als Software der Risikoklasse IIa, wenn sie dazu bestimmt ist, Informationen zu liefern, die zu Entscheidungen für diagnostische oder therapeutische Zwecke herangezogen werden. CDSS, deren Zwecke gerade in der Generierung von Informationen liegen, um Ärztinnen bei der Diagnostik oder Therapieentscheidung zu unterstützen, werden somit aber in aller Regel mindestens der Risikoklasse IIa zuzuordnen sein.⁹⁹ Können die auf Basis der gelieferten Informationen getroffenen Entscheidungen eine schwerwiegende (aber reversible) Verschlechterung des Gesundheitszustands einer Person oder einen chirurgischen Eingriff verursachen, sind sie sogar der Klasse IIb zuzuordnen. Können Tod oder

⁹⁴ Vgl. Fragen unter B Unterkapitel 1 (Zweckbestimmung und Stakeholder-Anforderungen) lit. d des Fragenkatalogs (Fn. 69).

⁹⁵ Vgl. Fragen unter B Unterkapitel 2 (Software-Anforderungen) lit. e des Fragenkatalogs (Fn. 69).

⁹⁶ Vgl. Fragen unter A Unterkapitel 3 (Kompetenzen bei der Entwicklung) des Fragenkatalogs (Fn. 69).

⁹⁷ Vgl. *Heil* (Fn. 51), S. 454; kritisch zu den weitreichenden fast alle Arten von Software umfassenden Klassifizierungsregeln vgl. *J. Prütting/T. Wolk*, Software unter dem Regime der europäischen Medizinprodukteverordnung (2017/745/EU), *MedR* 2020, S. 359 (363).

⁹⁸ *Dettling* (Fn. 53), S. 638; vgl. dazu auch *Hastenteufel/Renaud* (Fn. 59), S. 18 f.; *Prütting/Wolk* (Fn. 97), S. 361.

⁹⁹ Ebenso *Dettling* (Fn. 53), S. 638.

eine irreversible Verschlechterung des Gesundheitszustands einer Person die Folge sein, unterfallen sie der Risikoklasse III. Offen ist dabei allerdings, wie unmittelbar der Zusammenhang zwischen Einsatz der Software zur Informationsgewinnung und Folge (Tod, Gesundheitsschädigung etc.) sein muss. Denn denkbar ist, dass diese sich erst am Ende einer durch die von einer Software gelieferten Informationen angestoßenen Kausalkette realisieren, z. B. ein falsch befundeter MRT-Scan, der den Arzt dazu veranlasst, von einer Intervention abzusehen, oder langfristig wegen Nichtbehandlung zu einer Folgeerkrankung und nicht reversiblen Schäden führt. Selbst eine Software, die bspw. lediglich der Aufzeichnung und Übermittlung von Fieberwerten dient, müsste aufgrund möglicher mittelbarer Folgen bei Fehlfunktion dementsprechend in Risikoklasse IIb oder gar III eingestuft werden, was vom eigentlichen risikobasierten Ansatz der MDR nicht mehr viel übriglassen würde.¹⁰⁰ Entscheidend für die Zuweisung zur Risikoklasse ist aber letztendlich die Zweckbestimmung durch den Hersteller (vgl. Anhang VIII 3.1. MDR).¹⁰¹ Technische Eigenschaften und objektive Nutzungsmöglichkeiten der gelieferten Informationen in einem anderen, vom Hersteller nicht vorgesehenen Kontext, vermögen die Zuordnung allein nicht verändern.

d) Inverkehrbringen von im Betrieb weiterlernender CDSS unter der MDR?

Anhand dieser Maßgaben können bislang allerdings ausschließlich Systeme in einem eingefrorenen Zustand in Verkehr gebracht werden.¹⁰² Adaptive Systeme, die im laufenden Betrieb weiterlernen und sich weiterentwickeln und sich somit – je nach Standort und Nutzung – vom ursprünglichen Produkt unterscheiden können, gelten als nicht zertifizierbar bzw. bedürfen nach Auffassung der Praxis bei jeder substantziellen Veränderung einer erneuten Zertifizierung, da sich die Bewertung der Konformität des Produkts mit der MDR stets auf einen bestimmten technischen Zustand mit entsprechenden Funktionen beziehe.¹⁰³

Diese Handhabung entspricht jedoch nicht gänzlich den Vorgaben der MDR.¹⁰⁴ Denn analysiert man die MDR, dann wird man feststellen, dass diese Änderungen von Medizinprodukten nach dem Zeitpunkt ihrer Zertifizierung nicht per se aus-

¹⁰⁰ U. Gassner, Software als Medizinprodukt – zwischen Regulierung und Selbstregulierung, MPR 2016, S. 109 (112); Prütting/Wolk (Fn. 97), S. 362.

¹⁰¹ Vgl. hierzu und zum folgenden Prütting/Wolk (Fn. 97), S. 362; Reitebuch (Fn. 63), S. 107 f.

¹⁰² Vgl. IG NB Fragenkatalog (Fn. 69), S. 3; Deutsche Normungsroadmap KI (Fn. 42), S. 226; Heil (Fn. 51), S. 462 f.

¹⁰³ Besonders die Gewährleistung der Wiederholbarkeit und Zuverlässigkeit der Leistung gelten als problematisch, vgl. Hastenteufel/Renaud (Fn. 59), S. 96; Stoppacher/Müllner (Fn. 8), S. 169.

¹⁰⁴ Ebenso Gassner (Fn. 52), S. 48; Schreitmüller (Fn. 52), S. 149 ff.; das Inverkehrbringen weiterlernender Systeme ebenfalls nicht gänzlich ausschließend Reitebuch (Fn. 63), S. 127 f.

schließt bzw. stets eine erneute Zertifizierung erfordert. Vielmehr etabliert sie eine Reihe von Vorgaben, wie mit Änderungen zu verfahren ist, zumal solche gerade auch für herkömmliche Softwareprodukte (insb. durch Updates) nicht unüblich sind.¹⁰⁵ So haben Hersteller Änderungen etwa an der Auslegung des Produkts oder an seinen Merkmalen zeitgerecht angemessen zu berücksichtigen (Art. 10 Abs. 9 UAbs. 1 S. 2 MDR); im Rahmen des von ihnen einzurichtenden Qualitätsmanagements müssen stets auch Verfahren für das Management von Änderungen an dem vom System erfassten Produkten vorgesehen sein (Art. 10 Abs. 9 UAbs. 3 lit. a MDR). Zudem sind die benannten Stellen vom Hersteller über Änderungen des genehmigten Produkts zu informieren, welche dann auf dieser Grundlage erst darüber entscheiden, ob tatsächlich ein neues Konformitätsbewertungsverfahren erforderlich ist oder ob es ausreicht, einen Nachtrag zu den EU-Bescheinigungen über die Bewertung der technischen Dokumentation auszustellen (vgl. Anhang VII 4.9., Anhang IX 4.10. MDR).¹⁰⁶ Der Informationspflicht unterliegen – anders als es der Wortlaut von Anhang VII 4.9. MDR nahelegt – allerdings nur *wesentliche* Änderungen (vgl. Anhang IX 2.4. MDR), also solche, die eine Änderung der Zweckbestimmung des Produkts bedeuten, oder die die Sicherheits- und Leistungsfähigkeit des Produkts oder die für das Produkt vorgesehenen Anwendungsbedingungen beeinträchtigen können (Anhang IX 4.10. MDR). Alle anderen Änderungen bedürfen nicht zwingend einer Genehmigung durch die benannten Stellen. Offen lässt die MDR allerdings, anhand welcher Kriterien die benannten Stellen darüber entscheiden, ob es der Durchführung eines neuen Konformitätsbewertungsverfahrens bedarf oder ein Nachtrag zu den EU-Bescheinigungen genügt. Lediglich bei wesentlichen Änderungen des Designs oder der Zweckbestimmung eines nach der bisherigen Medizinprodukterichtlinie zertifizierten Produkts besteht nach der Übergangsvorschrift Art. 120 Abs. 3 MDR ein Verbot, dieses ohne erneute Zertifizierung weiterhin in Verkehr zu bringen.¹⁰⁷ Wann das der Fall ist, lässt sich anhand der zu Art. 120 Abs. 3 MDR veröffentlichten *Guidance on significant changes regarding the transitional provision under Article 120 of the MDR* der *Medical Device Coordination Group* (MDCG 2020–3)¹⁰⁸ beantworten, die ihrerseits hinsichtlich Software wiederum auf den Anhang VI 6.5. MDR, welcher bzgl. der UDI-Kennzeichnungen¹⁰⁹ zwischen geringfü-

¹⁰⁵ Vgl. dazu *Hastenteufel/Renaud* (Fn. 59), S. 143.

¹⁰⁶ In letzterem Fall bewertet die benannte Stelle die geplanten Änderungen, teilt dem Hersteller ihre Entscheidung mit und stellt ihm, sofern die Änderungen genehmigt wurden, einen Nachtrag zu der EU-Bescheinigung über die Bewertung der technischen Dokumentation aus.

¹⁰⁷ *Gassner* (Fn. 52), S. 48 f.

¹⁰⁸ *Medical Device Coordination Group*, *Guidance on significant changes regarding the transitional provision under Article 120 of the MDR with regard to devices covered by certificates according to MDD or AIMDD* (MDCG 2020–3).

¹⁰⁹ Die UDI-Kennzeichnung (Unique Device Identification) dient der Identifikation und Nachverfolgung von Medizinprodukten und gilt einschließlich für stand-alone Software, vgl. hierzu im Einzelnen *S. Mayr/A. Thiermann/M. Schrack/C. Kiesselbach*, *Das neue Medizinprodukterecht*. Praxishandbuch zur MP-VO, 2021, § 6 Rn. 12 ff.

gigen und sonstigen Änderungen unterscheidet, referenziert.¹¹⁰ Geringfügige Änderungen, die keiner Genehmigung bedürfen, sind danach solche, die der Fehlerbehebung oder Verbesserung der Betriebseffizienz dienen, einschließlich Sicherheitsupdates. Demgegenüber liegt eine wesentliche Änderung, die einer neuen UDI-DI und regelmäßig einer nochmaligen Konformitätsbewertung bedarf, vor, wenn sich die ursprüngliche Leistung, die Sicherheit oder die bestimmungsgemäße Verwendung oder die Auswertung der Daten auch bei gleichbleibender Zweckbestimmung des Produkts verändert (Anhang VI 6.5.2. MDR). Dazu gehören laut MDR insbesondere neue oder geänderte Algorithmen, Änderungen der Datenbankstrukturen, Betriebsplattformen und Architekturen sowie neue Schnittstellen oder neue Kanäle der Interoperabilität (Anhang VI 6.5.2. MDR).¹¹¹ Wann bei im Betrieb weiterlernenden Systemen die Schwelle zur Wesentlichkeit überschritten sein wird, wird dabei jeweils im Einzelfall zu bestimmen sein. Zweckbestimmung und Modellarchitektur werden sich zwar durch das Training mit neuen Datensätzen regelmäßig nicht verändern, denkbar ist aber, dass sich die Leistungsfähigkeit aufgrund veränderter Gewichtungen verschiebt.

4. Sachgemäße Nutzung und Implementierung in den Klinikalltag

Neben der Erfüllung grundlegender Sicherheits- und Leistungsanforderungen knüpft die MDR das Inverkehrbringen von Produkten zudem an die Minimierung von Risiken, die bspw. aus einem Fehlgebrauch des Produkts durch den Nutzer resultieren können. So haben Hersteller im Rahmen des nach Art. 10 Abs. 2, Anhang I 3 MDR einzuführenden Risikomanagements u. a. Risiken einzuschätzen und zu bewerten, [...] die bei einer vernünftigerweise vorhersehbaren Fehlanwendung auftreten, und diese zu beseitigen oder wenigstens zu minimieren (vgl. Anhang I 3 c-d MDR). Sichergestellt wird damit in Teilen auch die unter IV. 2. beschriebene sachgemäße Verwendung von CDSS, jedenfalls soweit diese im Verantwortungsbereich des Herstellers liegt. Dazu gehört bei Medizinproduktesoftware die Minimierung von möglichen Fehlern von Nutzerinnen im Rahmen des bestimmungsgemäßen Gebrauchs, insbesondere bei der Eingabe (z. B. Messfehlern) oder der Interpretation der Ergebnisse der Software durch eine bessere *Usability* der Software (bspw. durch eine Verbesserung der Steuerungselemente oder der Anzeige).¹¹² Abgesehen davon hat der Hersteller – wie bei jedem anderen Medizinprodukt – Vorkehrungen zu treffen, damit das Produkt überhaupt bestimmungsgemäß gebraucht wird, also bspw. ausschließlich bei der für die Anwendung vorgesehenen Patientengruppe und – falls erforderlich – vom entsprechend geschulten Personal mit medizinischen Kenntnissen.

¹¹⁰ MDCG 2020–3 (Fn. 108), S. 4; eingehend *Gassner* (Fn. 52), S. 49.

¹¹¹ Vgl. auch Chart C der MDCG-3 (Fn. 108), S. 10; *Hastenteufel/Renaud* (Fn. 59), S. 31; vgl. dazu auch *B. Tölle*, Unique Device Identification (UDI), abrufbar unter <https://www.johner-institut.de/blog/regulatory-affairs/unique-device-identification-udi/> (zuletzt abgerufen am 13.2.2023).

¹¹² Vgl. hierzu *Hastenteufel/Renaud* (Fn. 59), S. 89 ff.

Üblicherweise erfolgt dies durch Bereitstellung einer entsprechenden Gebrauchsanleitung, die u. a. die genaue Zweckbestimmung des Produkts mit Angabe der Indikationen, Kontraindikationen, Patientengruppe(n) und vorgesehenen Anwender, Angaben zum erwarteten klinischen Nutzen sowie Leistungsmerkmale des Produkts zu enthalten hat (Anhang I 23.4. MDR).¹¹³ Bei diagnostischen Produkten oder Produkten mit Messfunktion hat der Hersteller zudem die Genauigkeitsgrenzen anzugeben (Anhang I 15.1. MDR). Begegnet wird damit von vornherein u. a. den unter III. beschriebenen bei der Verwendung von CDSS bestehenden Risiken einer etwaigen falschen Nutzung durch den Verwender sowie möglichen auftretenden Missverständnissen über den Output des Produkts und dessen Aussagekraft, die dann im Fall von CDSS zu Fehlentscheidungen bei der Diagnose und Behandlung führen können. Der Fragenkatalog der IG NB enthält zudem für KI-basierte Medizinprodukte weitergehende Spezifika für die Begleitmaterialien zu den Produkten:¹¹⁴ So soll bspw. die Gebrauchsanleitung nicht nur die vorgesehene Patientenpopulation bestimmen, sondern auch explizit Patientinnen, Daten und Anwendungsfälle benennen, bei denen das Produkt nicht verwendet werden darf. Darüber hinaus soll die Gebrauchsanweisungen auch Anforderungen an die Input-Daten (inklusive Formate, Auflösung, Wertebereich etc.) sowie die genaue Zweckbestimmung des Produkts dokumentieren.

Garantiert werden kann mit diesen Vorgaben allerdings weder abschließend der sachgemäße Umgang durch den Nutzer noch, dass solche Systeme vernünftig in den Organisationsablauf eingebettet sind, denn dies liegt im Verantwortungsbereich des jeweiligen Verwenders (Klinik, Arztpraxis etc.), den die MDR mit ihren Bestimmungen jedoch nicht adressiert. Letzteres wird daher vor allem durch das Haftungsrecht abgesichert.

VI. Fazit

Der Einsatz KI-basierter CDSS birgt spezifische Risiken für die Gesundheit und Sicherheit von Patientinnen, auch wenn diese Systeme bestimmungsgemäß bislang nur als Assistenz und Orientierungshilfe konzipiert sind und die Letztentscheidung weiterhin dem behandelnden Arzt obliegt. Mangelnde Genauigkeit, Verwendung für eine nicht im Trainingsdatensatz abgebildete Patientenpopulation aber auch Unwissenheit über die Bedeutung des Outputs sind nur einige Risiken, die Ärztinnen auf Basis der generierten Information zu einer medizinisch nicht indizierten (Be)Handlung veranlassen können.

Mit der MDR existiert jedoch bereits ein Rechtsrahmen, der dem Grunde nach geeignet ist, auch diese von KI-basierten CDSS für die Gesundheit und Patientensicherheit ausgehenden Risiken zu adressieren und weitgehend zu minimieren. Das

¹¹³ Vgl. hierzu auch im Kontext von KI-Produkten *COCIR* (Fn. 43), S. 21.

¹¹⁴ Vgl. Fragen unter B Unterkapitel 5 (Produktentwicklung) lit. b des Fragenkatalogs (Fn. 69).

gilt sowohl für Gefahren, die vom System selbst ausgehen, als auch für solche, die aus einer möglichen unsachgemäßen Nutzung bzw. mangelhaften Einbettung in den jeweiligen Organisationsablauf herrühren können. Da das Regulierungskonzept die Konkretisierung und Herausbildung von Qualitätsstandards für Verfahren und Produkte privaten Normungsgesellschaften überlässt, ist trotz Fehlen spezifischer Bestimmungen für KI-Systeme eine Zertifizierung und ein Inverkehrbringen solcher Systeme unter der MDR möglich, auch wenn es freilich auf Normungsebene so schnell wie möglich verbindlicher Standards bedarf. Bis diese vorliegen, ist eine Nachweisführung der Erfüllung der Anforderungen der MDR und Zertifizierung von KI-Systemen auf Basis des unter V. 3 b) vorgestellten und in Deutschland als Benchmark geltenden Fragenkatalogs der IG NB möglich. Insbesondere KI-gestützte CDSS, die ein begrenztes Problem adressieren, nicht weiterlernen und bei denen bereits ein klinischer Zusammenhang zwischen In- und Output nachgewiesen ist, werden auf dieser Basis zertifizierbar sein.

Ob und inwieweit darüber hinaus tatsächlich noch zusätzlich erforderlich ist, dass KI-basierte Systeme bzw. deren Output – wie vielfach (nicht zuletzt auch gesetzlich)¹¹⁵ gefordert¹¹⁶ – für die Nutzerin (oder gar den Patienten) nachvollziehbar und erklärbar sind, wenn bereits nach den Maßgaben der MDR ihre Sicherheit und Leistungsfähigkeit nachgewiesen ist, bleibt allerdings fraglich. Die MDR erfordert jedenfalls nicht, dass Geräte ihre Entscheidungen und Handlungen der Nutzerin gegenüber erklären, noch schließt sie aus, dass Produkte, deren Wirkmechanismen und Funktionsweisen im Einzelnen unerklärbar sind, in Verkehr gebracht werden, solange sie den Anforderungen der MDR genügen.¹¹⁷ Dies ist insoweit konsequent, als im Fall der Zertifizierung von KI-Systemen unter der MDR dessen Eigenschaften nicht vollkommen opak sind. Denn mit dem Nachweis der Erfüllung der Anforderungen der MDR müssen nicht nur das System und dessen Verhaltensverhalten unter bestimmten Bedingungen vom Hersteller untersucht und verstanden sein,¹¹⁸ dessen Zuverlässigkeit und Leistungsfähigkeit müssen auch anhand von Kriterien beschrieben und validiert sein. Auch heute schon finden sich im Medizinsektor Beispiele von Produkten insb. im Arzneimittelbereich, deren Wirkung zwar durch Studien nachgewiesen, im Einzelnen aber nicht verstanden ist. Daher muss auch mit Blick auf den Nutzen solcher Produkte – erst recht, wenn sie nicht einmal unmittelbar

¹¹⁵ Vgl. bspw. nur Artt. 13 Abs. 2 lit. f, 14 Abs. 2 lit. g DSGVO, § 93 Abs. 1 MStV, Art. 27 DSA.

¹¹⁶ Vgl. Hochrangige Expertengruppe für Künstliche Intelligenz (Fn. 43), S. 22; Fraunhofer IAIS (Fn. 43), S. 17; vgl. hierzu auch *Hänold/Schlee/Antweiler/Beckh* (Fn. 14), S. 518; *D. Roth-Isigkeit*, Unionsrechtliche Transparenzanforderungen an intelligente Medizinprodukte, GesR 2022, S. 278 ff.

¹¹⁷ *COCIR* (Fn. 43), S. 19; *Gassner* (Fn. 52), S. 50; vgl. dazu auch *Schreitmüller* (Fn. 52), S. 134 f.

¹¹⁸ So werden bspw. Design- und Modellentscheidungen, die Wahl der Eingabedaten etc. bewusst in Abhängigkeit davon getroffen, inwieweit diese geeignet für die Lösung der Aufgabenstellung sind, vgl. zu Fehlvorstellungen über die Opazität KI-basierter Systeme *J. Kroll*, The fallacy of inscrutability, *Phil. Trans R. Soc.* 2018, 376 ID: 20180084.

am Patienten wirken, wie im Fall von CDSS – kritisch hinterfragt werden, ob mit der Nachweisführung im Rahmen von Zulassungsverfahren nicht bereits hinreichende Transparenz gegeben ist.

Literatur

- Berner*, Eta S.: Clinical Decision Support Systems: Theory and Practice, 3rd Edition 2016.
- Detting*, Heinz-Uwe: Künstliche Intelligenz und digitale Unterstützung ärztlicher Entscheidungen in Diagnostik und Therapie, PharmR 2019, S. 633–642.
- Gassner*, Ulrich M.: Intelligente Medizinprodukte – Regulierungsperspektiven und Zertifizierungspraxis, MPR 2021, S. 41–51.
- Hastenteufel*, Mark/*Renaud*, Sina: Software als Medizinprodukt. Entwicklung und Zulassung von Software in der Medizintechnik, Wiesbaden 2019.
- Heil*, Marina: Innovationermöglichungsrecht oder Innovationshemmnis? Regulatorische Herausforderungen für KI-basierte Medizinprodukte-Software in der EU, in: Grinbalt, Roman/ Schlotz, Sybille/ Stock, Sophy (Hrsg.), Medizinproduktrecht im Wandel – Festschrift für Ulrich M. Gassner zum 65. Geburtstag, Baden-Baden 2022, S. 447–466.
- Pechmann*, Ludwig/*Mildner*, Martin/*Suthau*, Tim/*Leucker*, Martin: Regulatorische Anforderungen an Lösungen der künstlichen Intelligenz im Gesundheitswesen, in: Pfanntiel, Mario A. (Hrsg.), Künstliche Intelligenz im Gesundheitswesen, 2022, S. 175–198.
- Schreitmüller*, Zeynep: Regulierung intelligenter Medizinprodukte. Eine Analyse unter besonderer Berücksichtigung der MPVO und DSGVO, Baden-Baden 2023.
- Steinwendner*, Joachim: Klinische Entscheidungsunterstützungssysteme: von der Datenrepräsentation zur künstlichen Intelligenz, in: Mario A. Pfanntiel/Kirstin Kassel/Christoph Rasche (Hrsg.), Innovation und Innovationsmanagement im Gesundheitswesen, Wiesbaden 2020, S. 683–698.
- Sutton*, Reed T./*Pincock*, David/*Baumgart*, Daniel C./*Sadowski*, Daniel C./*Fedorak*, Richard N./*Kroeker*, Karen I.: An overview of clinical decision support systems: benefits, risks, and strategies for success, npj Digital Medicine (2020) 3:17.
- Zentrale Ethikkommission bei der Bundesärztekammer (ZEKO)*: Entscheidungsunterstützung ärztlicher Tätigkeit durch Künstlicher Intelligenz, Deutsches Ärzteblatt, Jg. 118, Heft 33–34.