AI-D des Hippokrates

David Schneeberger

A. Einleitung

Ὄμνυμι Ἀπόλλωνα ἰητρὸν, καὶ Ἀσκληπιὸν, καὶ Ὑγείαν, καὶ Πανάκειαν, καὶ θεοὺς πάντας τε καὶ πάσας, ἵστορας ποιεύμενος, ἐπιτελέα ποιήσειν κατὰ δύναμιν καὶ κρίσιν ἐμὴν ὅρκον τόνδε καὶ ξυγγραφὴν τήνδε. Die Verständigung zwischen Ärztin und Patientin stößt – wie dieses Zitat aus dem Text des hippokratischen Eides metaphorisch illustrieren soll – mitunter aufgrund des fehlenden gemeinsamen Vokabulars oft auf Kommunikationsschwierigkeiten.

Ohne eine verständliche Aufklärung über die Diagnose oder Behandlungsoptionen bleibt die Entscheidungsgewalt in den Händen der Ärztin. Um solchen Informations- und daraus resultierenden Machtasymmetrien zwischen Ärztin und Patientin vorzubeugen, zielt die moderne Medizin anstatt paternalistischer Bevormundung auf eine partnerschaftliche Entscheidungsfindung ab.[1] Damit eine Patientin ihr Recht auf Selbstbestimmung ausüben kann, müssen ihr alle Informationen vermittelt werden, die notwendig sind, um eine fundierte Entscheidung zu treffen (sog. *informed consent*). Eine Zustimmung zur Behandlung kann nur dann gültig erfolgen, wenn ihr die Tragweite ihrer Entscheidung bewusst ist.[2] Fehlt es daran, ist die Behandlung rechtswidrig, selbst wenn sie medizinisch indiziert ist und sorgfaltsgemäß durchgeführt wird.[3]

Dieses Konzept der partnerschaftlichen Entscheidungsfindung zwischen Ärztin und Patientin wandelt sich jedoch durch die Verbreitung von Künstlicher Intelligenz (KI), insbesondere in der Unterform des Machine Learning (ML). Denn da ML-Systeme u.a. in der Diagnose, z.B. von Hautkrankheiten oder diabetischer Retinopathie,[4] bei der Prävention von

1 *M. Neumayr* in: M. Neumayr/R. Resch/F. Wallner, Gmundner Kommentar zum Gesundheitsrecht, 2. Aufl., Wien 2022, Einleitung ABGB Rn. 36.
2 *F. Wallner*, Medizinrecht, 2. Aufl., Wien 2022, Rn. 481.
3 *Neumayr* (Fn. 1), Einleitung ABGB Rn. 32.
4 *E. Topol*, Deep Medicine. How Artificial Intelligence can Make Healthcare Human Again, New York 2019, S. 41 ff.; siehe auch die Beiträge in A. Manzei-Gorsky/C. Schubert/J. von Hayek (Hrsg.), Digitalisierung und Gesundheit, Baden-Baden

Krankheiten, z.B. durch die Feststellung von für Krankheiten indikative Biomarker, sowie im Bereich der Medikamentenentwicklung oder Gensequenzierung zunehmend Einsatz finden,[5] treten sie metaphorisch als dritte Partei an die Seite der Ärztin und Patientin und sind somit prototypisch für die verstärkte Dominanz der Naturwissenschaften in der Medizin.[6] In Form von „verkörperten" KI-Systemen, d.h. Robotik, sind auch bspw. chirurgische Roboter, sozial oder physisch assistierende Roboter sowie Pflegedienstroboter bereits im Einsatz.[7]

Vor allem aufgrund verschiedener technischer Eigenschaften (einiger) der verwendeten Modelle, die eine Interpretation des Entscheidungsprozesses erschweren (sog. „Black-Box-Problematik"),[8] ergibt sich ein Spannungsverhältnis zwischen der Aufklärung[9] als Bedingung für Selbstbestimmung und dem Einsatz intelligenter Medizinprodukte. Den Ausgangspunkt der Überlegungen bildet der Einsatz von ML-Modellen in einem *decision-support*-Szenario (z.B. Empfehlung einer Diagnose oder einer Behandlungsmethode). Das Aufklärungsgespräch verbleibt dabei in der Hand der Ärztin, während das ML-Modell und sein Output zum Gegenstand der Aufklärung werden. Die Einführung eines vollautomatisierten „Dr. Robot" oder eine automatisierte Aufklärung[10] ist dabei zum gegenwärtigen Stand auch aus rechtlichen Gründen – ärztliche Tätigkeit ist auf-

2022. Ein weiterer Anwendungsbereich liegt in der Immunologie z.B. bei der Bekämpfung der COVID-19-Pandemie, *L. J. Catania*, AI for Immunology, Boca Raton, 2021.

5 *D. Roth-Isigkeit*, Unionsrechtliche Transparenzanforderungen an intelligente Medizinprodukte, GesR 2022, 278 (279); *C. Katzenmeier*, Big Data, E-Health, M-Health, KI und Robotik in der Medizin, MedR 2019, 259 (259). Daneben existieren auch Monitoringsysteme, Modelle zur Spracherkennung oder Systeme, die bei der Recherche unterstützen, siehe *D. Linardatos*, Intelligente Medizinprodukte und Datenschutz, CR 2022, 367 (368).

6 *Katzenmeier*, Data (Fn. 5), 270 f.

7 *E. Fosch-Villaronga/H. Drukarch*, AI for Healthcare Robotics, Boca Raton, 2022, S. 24 f.

8 *A. D. Selbst/S. Barocas*, The Intuitive Appeal of Explainable Machines, Fordham Law Review 2018, 1085 (1094 ff.).

9 Zur Einordnung der Aufklärung als eine Form von „Transparenz", vgl. *A. Kiseleva/D. Kotzinos/P. de Hert*, Transparency of AI in Healthcare as a Multilayered System of Accountabilities, Frontiers in Artificial Intelligence 2022, 1 (11 ff.).

10 *C. Rahn*, Ärztliche Aufklärung durch Künstliche Intelligenz, Hamburg 2022.

grund des Arztvorbehalts auf Menschen beschränkt[11] – unwahrscheinlich und daher nicht Gegenstand der Betrachtungen.

B. Medizinische Aufklärung

I. Rechtsgrundlagen

Die österreichische Rechtsordnung kennt zwei Grundlagen für die ärztliche Aufklärungspflicht: zum einen gesetzliche Regelungen, wobei die Aufklärungspflicht nicht umfassend geregelt ist und in der Rechtsordnung verstreut nur einige wenige Bestimmungen existieren, zum anderen den Behandlungsvertrag zwischen der Patientin und der behandelnden Ärztin bzw. dem Krankenanstaltenträger.[12] Die Rechtslage mit einer fehlenden Regelung der Aufklärung im Allgemeinen ähnelt damit der deutschen vor 2013, als die Grundsätze der Aufklärung in § 630c Abs. 2 S. 1 BGB und § 630e BGB kodifiziert wurden.[13]

Die Thematik der Aufklärung über „intelligente Medizinprodukte" hat einen hohen Überschneidungsgrad mit dem Medizinprodukterecht, geregelt in der MPVO,[14] und dem geplanten Artificial Intelligence Act (AIA), der diese teilweise ergänzt. Medizinprodukte, die aufgrund ihrer Risikoklasse einer *ex-ante*-Konformitätsprüfung aufgrund harmonisierter Rechtsvorschriften wie der MPVO unterliegen, fallen als Hochrisikosysteme unter den AIA (Art. 6 Abs. 1 AIA i.V.m. Anhang II Z. 11, 12).[15] Daneben be-

---

11 *G. Ganzger/L. Vock*, Artificial Intelligence in der ärztlichen Entscheidungsfindung, JMG 2019, 153 (157 f.); *Linardatos*, Medizinprodukte (Fn. 5), 368 f; *Roth-Isigkeit*, Transparenzanforderungen (Fn. 5), 282.

12 *M. Memmer*, Aufklärung, in: G. Aigner/A. Kletečka/M. Kletečka-Pulker/M. Memmer (Hrsg.), Handbuch Medizinrecht für die Praxis, Wien 2022, I.3.

13 *C. Katzenmeier*, Aufklärungspflicht und Einwilligung, in: A. Laufs/C. Katzenmeier/V. Lipp (Hrsg.), Arztrecht, 8. Aufl., München 2021, Kap V. Rn. 3 f; *C. Katzenmeier* in: W. Hau/R. Poseck (Hrsg.), BeckOK BGB, 62. Edition, München 2022, § 630e BGB Rn. 1.

14 Software fällt dabei unter die Definition des Medizinprodukts gem. Art. 2 Abs. 1 MPVO, *K. Helle*, Intelligente Medizinprodukte, MedR 2020, 993 (994); *D. Schneeberger/K. Stöger/A. Holzinger*, The European Legal Framework for Medical AI, in A. Holzinger/P. Kieseberg/A. M. Tjoa/E. Weippl (Hrsg.), CD-MAKE 2020, Cham 2020, S. 209 (216 f.).

15 *Roth-Isigkeit*, Transparenzanforderungen (Fn. 5), 283. Software-Produkte, die Informationen liefern, die zu Entscheidungen für diagnostische oder therapeutische Zwecke herangezogen werden, fallen zumindest in die Risikoklasse IIa (Regel 11 Anhang VIII MPVO), *S. Jabri*, Artificial Intelligence and Healthcare, in: T.

steht aufgrund der Verarbeitung von Gesundheitsdaten typischerweise eine Anwendbarkeit der DSGVO, wobei Art. 22 DSGVO und die damit verbundenen „speziellen" Informationspflichten und das Auskunftsrecht im medizinischen Kontext häufig nur eine eingeschränkte Rolle spielen, da Vollautomatisierung bereits an berufsrechtliche Grenzen stößt.[16]

II. Aufklärungsformen

Auch wenn häufig nur von „Aufklärung" gesprochen wird, lässt sich diese nach ihrem Gegenstand und ihrer Funktion in zwei Hauptformen unterteilen: erstens die Selbstbestimmungsaufklärung, die, indem sie das notwendige Wissen verschafft, die Grundlage für die Entscheidung der Patientin zur Einwilligung oder Ablehnung einer Behandlung bildet; zweitens die Sicherungsaufklärung, die, nachdem die Patientin in die konkrete ärztliche Maßnahme eingewilligt hat, zum Tragen kommt. Diese stellt einen Teil der ärztlichen Behandlung dar, soll die bestmögliche Mitwirkung der Patientin bewirken und dient damit der Sicherstellung des Heilerfolgs.[17] Der Fokus dieses Beitrags liegt auf der Selbstbestimmungsaufklärung.

Im Rahmen der Selbstbestimmungsaufklärung ist der Patientin jenes Wissen zu vermitteln, das notwendig ist, um abschätzen zu können, worin sie einwilligt bzw. welche Folgen die Ablehnung einer Behandlung nach sich zieht. Sie ist über die Diagnose, den Therapieverlauf sowie alternative Methoden und über die Risiken der in Aussicht genommenen Maßnahme aufzuklären. Die Selbstbestimmungsaufklärung lässt sich demnach in die Diagnoseaufklärung, die Verlaufsaufklärung sowie die Risikoaufklärung unterteilen.[18]

Wischmeyer/T. Rademacher (Hrsg.), Regulating Artificial Intelligence, Cham 2020, S. 307 (323).

16 *Roth-Isigkeit*, Transparenzanforderungen (Fn. 5), 282; *K. Stöger*, Explainability und „informed consent" im Medizinrecht, in P. Leyens/I. Eisenberger/R. Niemann (Hrsg.), Smart Regulation, Tübingen 2021, S. 143 (144 ff.).

17 *D. Engljähringer*, Ärztliche Aufklärungspflicht vor medizinischen Eingriffen, Wien 1996, S. 7 ff; *H. Jesser-Huß*, Zivilrechtliche Haftung und Fragen der Aufklärung, in: R. Resch/F. Wallner (Hrsg.), Handbuch Medizinrecht, 3. Auflage, Wien 2020, IV. Rn. 78; *M. Kletečka-Pulker/M. Grimm/M. Memmer/L. Stärker/J. Zahrl*, Grundzüge des Medizinrechts, Wien 2019, 54 f; *Memmer*, Aufklärung (Fn. 12), I.3.2; *Neumayr* (Fn. 1), Einleitung ABGB Rn. 35; *K. Prutsch*, Die ärztliche Aufklärung. Handbuch für Ärzte, Juristen und Patienten, 2. Aufl., Wien 2004, S. 136.

18 *S. Laimer*, Ausmaß und Grenzen der Selbstbestimmungsaufklärung, RdM 2022, 250 (251); *Memmer*, Aufklärung (Fn. 12), I.3.2.1.

1. Diagnoseaufklärung

Bei der Diagnoseaufklärung muss die Ärztin die Patientin über den erhobenen Befund informieren.[19] Als Ausgangspunkt für das Selbstbestimmungsrecht ist die Patientin darüber aufzuklären, dass sie überhaupt krank ist und an welcher Krankheit sie leidet.[20] Sie hat erst stattzufinden, wenn die Diagnose gesichert ist, da eine bloße Verdachtsdiagnose möglicherweise zu Verunsicherung führt.[21]

In der Literatur ist bereits die Frage umstritten, ob über den Einsatz eines ML-Systems, z.B. zur Diagnose, aufgeklärt werden muss. Dies ist m.E. in Hinblick auf Art. 22 DSGVO und die damit verknüpften Informationspflichten in Art. 13–14 DSGVO umso eher zu bejahen, je mehr die Diagnoseentscheidung in die „Hände der Maschine" gelegt wird. Wenn eine Ärztin nur noch als „Sprachrohr" einer automatisierten Diagnose fungiert und keine überprüfende Funktion wahrnimmt, ist eine Information über den Einsatz von KI datenschutzrechtlich notwendig. Faktisch würde dieses *rubber stamping* jedoch, wie bereits erwähnt, gegen berufsrechtliche Grenzen verstoßen. Ist die ML-Diagnose nur eine „zweite Meinung", d.h. nur ein Faktor unter vielen, und hat somit nur geringen Anteil an der Diagnose, ist eine solche Bekanntgabe schwieriger zu argumentieren.[22]

Da es sich bei ML zum gegenwärtigen Stand noch um eine nicht weitverbreitete „Außenseitermethode"[23] mit noch unbekannten Risiken handelt, sollte m.E., um die Selbstbestimmungsrechte der Patientinnen zu wahren, darüber aufgeklärt werden. Auch aus medizinethischer Sicht wurde gefordert, den KI-Einsatz nicht zu verbergen.[24] Wenn sich KI jedoch als Stand der Wissenschaft eingebürgert hat, ist es wahrscheinlich, dass – wie

19 *Memmer*, Aufklärung (Fn. 12), I.3.2.1.1; *Neumayr* (Fn. 1), Einleitung ABGB Rn. 38.

20 *Prutsch*, Aufklärung (Fn. 17), S. 136 f.

21 *Kletečka-Pulker/Grimm/Memmer/Stärker/Zahrl*, Grundzüge (Fn. 17), S. 57 f; *Memmer*, Aufklärung (Fn. 12), I.3.2.1.1.

22 *G. I. Cohen*, Informed Consent and Medical Artificial Intelligence, The Georgetown Law Journal 2020, 1425 (1442).

23 *D. Schneeberger*, KI-Assistenzsysteme in der Medizin, RdM 2021, 138 (142).

24 *F. Ursin/C. Timmermann/M. Orzechowski/F. Steger*, Diagnosing Diabetic Retinopathy With Artificial Intelligence, Frontiers in Medicine 2021, 1 (2); *WHO*, Ethics and Governance of Artificial Intelligence for Health, 2021, S. 48, abrufbar unter https://www.who.int/publications/i/item/9789240029200 (zuletzt abgerufen: 09.12. 2022).

auch bei anderen Verfahren – nicht mehr explizit über die Verwendung einer KI gestützten Diagnosemethode aufgeklärt werden muss.[25]

Dabei darf nur über eine gesicherte Diagnose aufgeklärt werden. Eine Verdachtsdiagnose ist, wenn ausnahmsweise darüber aufgeklärt wird, als eine solche zu benennen.[26] Das Ergebnis eines ML-Modells ist jedoch technisch zwangsläufig eine durch Induktion gewonnene Wahrscheinlichkeitsaussage. Ein ML-Ergebnis muss insofern als Verdacht präsentiert und darf weder der Ärztin noch der Patientin als gesicherte Diagnose „verkauft" werden. Ein solcher „Diagnosevorschlag" muss daher vor der Aufklärung durch zusätzliche (menschliche) Untersuchungen abgesichert werden. Der Ärztin obliegt damit als überprüfende Instanz, als *human-in-the-loop*, eine Kontrollfunktion. Wie hoch der „Verdacht" hinter einer Diagnose ist, lässt sich technisch durch Angabe eines *confidence scores*, d.h. dem „Vertrauen" eines Modells in sein Ergebnis, quantifizieren. Fehlt eine solche Angabe, könnten subsidiär im Rahmen der Evaluierung ermittelte Fehlermetriken wie Genauigkeit angegeben werden.[27]

Die Aufklärung über den Befund impliziert, dass nicht nur über einen Output eines Black-Box-Modells (z.B. „Hautkrebs Stufe IV") aufgeklärt wird. Ein solches „orakelartiges" Ergebnis hat – da die Diagnose nicht näher begründet wird – m.E. nur wenig Mehrwert in Bezug auf die Selbstbestimmung der Patientin. Jedoch darf mit Blick auf die zunehmende Komplexität der Medizin, die eine Aufklärung in allen Details auch ohne den Einsatz von KI häufig „zur Fiktion"[28] werden lässt, die Anforderung an die Aufklärung nicht überspannt werden. Eine ausführliche Erläuterung einer Diagnose ist auch bei rein „menschlichen Diagnosen" nicht geboten. Ärztinnen generieren Diagnosen oft selbst mittels Erfahrung und Intuition und können diese nicht im Detail begründen. Auch z.B. bei Untersuchungen mittels MRT muss die technische Funktionsweise nicht näher erläutert werden. In Hinblick auf die Verständlichkeit der Aufklärung scheint es im Ergebnis argumentierbar, dass grundlegende Informationen über die we-

25 *F. Molnár-Gábor*, Artificial Intelligence in Healthcare, in: T. Wischmeyer/T. Rademacher (Hrsg.), Regulating Artificial Intelligence, Cham 2020, S. 337 (349 f.); *Stöger*, Explainability (Fn. 16), 150.

26 *Memmer*, Aufklärung (Fn. 12), I.3.2.1.1.

27 *Stöger*, Explainability (Fn. 16), 147.

28 *W. Eberbach*, Wird die ärztliche Aufklärung zur Fiktion? (Teil 1), MedR 2019, 1; *W. Eberbach*, Wird die ärztliche Aufklärung zur Fiktion? (Teil 2), MedR 2019, 111.

sentliche Funktionsweise und mögliche Fehlerquellen für eine informierte Einwilligung ausreichen.[29]

Im Falle einer Hautkrebsdiagnose wäre dies z.B. die Angabe, dass ein Modell mithilfe von Bilddaten trainiert wurde und so erlernt hat, zwischen verschiedenen Formen von Muttermalen zu unterscheiden. Notwendig könnten auch Angaben zur Performanz des Systems und der erfolgten klinischen Bewertung sein. Somit würde die Diagnoseaufklärung gleichgeschaltet werden mit der h.M. zur Auslegung der „involvierten Logik" in Art. 13–15 DSGVO.[30] Eine Aufklärung i.S.v. „Sie haben Hautkrebs, da die Größe des Muttermals 5 mm überschreitet und die Ränder unscharf sind"[31] ist rechtlich nicht erforderlich. Dieser Fokus auf eine Form von Systemtransparenz spiegelt sich m.E. inzwischen auch in Art. 13 AIA wider, der das Hauptaugenmerk auf Gebrauchsanweisungen für Nutzerinnen legt. Diese müssen u.a. Informationen zu den Merkmalen, Fähigkeiten und Leistungsgrenzen eines Hochrisikosystems enthalten. Diese Angaben erlauben den Ärztinnen das notwendige grobe Verständnis für ein System und somit eine grobe Aufklärung.

2. Verlaufsaufklärung

Dieser Teil der Aufklärung, teilweise als Therapie-, Behandlungs- oder Verlaufsaufklärung bezeichnet, soll der Patientin ein Bild von Wesen und Umfang, von der Schwere und Dringlichkeit der geplanten Maßnahmen sowie Informationen über die Erfolgsaussichten und allfällige Folgewirkungen liefern.[32] Damit werden die therapeutischen Möglichkeiten und ihre Alternativen dargelegt.[33] Die Ärztin muss daher das Für und Wider

---

29 *Stöger*, Explainability (Fn. 16), 147 f; international *R. Matulionyte/P. Nolan/F. Magrebi/A. Beheshti*, Should AI-Enabled Medical Devices be Explainable? 2022, abrufbar unter https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4140234 (zuletzt abgerufen: 09.12. 2022), S. 27 f; *D. Schiff/J. Borenstein*, How Should Clinicians Communicate with Patients About the Roles of Artificially Intelligent Team Members? AMA Journal of Ethics 2019, 138 (140 f.).

30 *D. Schönberger*, Artificial intelligence in healthcare, International Journal of Law and Information Technology 2019, 171, (190); a.A. *K. Astromskė/E. Peičius/P. Astromskis*, Ethical and legal challenges of informed consent applying artificial intelligence in medical diagnostic consultations, AI & SOCIETY 2021, 509 (515 f.).

31 *Linardatos*, Medizinprodukte (Fn. 5), 371 spricht von der „inneren, medizinspezifischen Logik".

32 *Memmer*, Aufklärung (Fn. 12), I.3.2.1.2.

33 *Neumayr* (Fn. 1), Einleitung ABGB Rn. 39.

einer therapeutischen Maßnahme nachvollziehbar vermitteln, sodass die Patientin die Tragweite ihrer Entscheidung überschauen kann.[34]

Wenn auch ML-Systeme bislang primär in der Diagnose Einsatz finden, so existieren auch im Bereich der Therapie bereits erste Anwendungen. Bspw. wurde ein System für die Diagnose und Behandlung von Sepsis auf der Intensivstation konstruiert, das die passende Medikamentendosierung wählt.[35] Erste vollautomatisierte Chirurgieroboter werden experimentell erprobt.[36]

Dabei stellt sich die Frage, ob für die Verlaufsaufklärung eine Erklärung notwendig ist, warum ein Modell z.B. bei Brustkrebs der Option der chirurgischen Behandlung eine höhere Erfolgswahrscheinlichkeit zuweist als anderen Behandlungsalternativen. Denn nach manchen Stimmen wie *J. C. Bjerring* und *J. Busch* reicht eine Darlegung der Vor- und Nachteile nicht aus. Eine Erklärung, warum eine Therapiemethode effizienter ist, wäre daher notwendig.[37] Auf rechtlicher Ebene ist diese Annahme jedoch zu hinterfragen. Denn die Wahl einer Therapiemethode obliegt der Ärztin, solange sie dem Stand der medizinischen Wissenschaft entspricht. Sie muss nicht ungefragt über potentiell mögliche Behandlungsalternativen aufklären.[38] Sie muss z.B. die Wahl oder Dosierung von Medikamenten nicht umfassend rechtfertigen. Selbst wenn eine „entschlossene Patientin" eine bestimmte Maßnahme wünscht, muss die Ärztin zwar die potentiell geringeren Erfolgschancen, nicht aber etwaige Alternativen darlegen.[39]

Festzustellen ist jedoch wiederum, dass beim Einsatz von ML-Systemen, die gegenwärtig noch als „Außenseitermethoden" anzusehen sind, jedenfalls über den Umstand aufzuklären ist, dass ML eingesetzt wird. Notwendig ist der Hinweis, dass es sich um eine neue oder unerprobte

34 *Prutsch*, Aufklärung (Fn. 17), S. 139 f.

35 *M. Komorowski/L. A. Celi/O. Badawi/A. C. Gordon/A. Aldo Faisal*, The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care, Nat Med 2018, 1716.

36 *H. Saeidi/J. D. Opfermann/M. Kam/S. Wei/S. Leonard/M. H. Hsieh/J. U. Kang/A. Krieger*, Autonomous robotic laparoscopic surgery for intestinal anastomosis, Science Robotics 2022, abrufbar unter https://pubmed.ncbi.nlm.nih.gov/35080901/ (zuletzt abgerufen: 09.12.2022)

37 *J. C. Bjerring/J. Busch*, Artificial Intelligence and Patient-Centered Decision-Making, Philosophy & Technology 2021, 349 (362 f.).

38 *Neumayr* (Fn. 1), Einleitung ABGB Rn. 40.

39 *Neumayr* (Fn. 1), Einleitung ABGB Rn. 40.

Behandlungsmethode handelt und deshalb die Risiken noch nicht voll abschätzbar sind.[40]

Die Aufklärung über „Art und Wesen" einer Behandlung ist, da nur Grundinformationen notwendig sind (z.B. „Medikamentengabe, deren Dosierung durch KI unterstützt wird"), auch bei Black-Box-Modellen möglich. Die notwendige Aufklärung über „den Umfang", die „Schwere einer Behandlung", die „Erfolgsaussichten und Versagerquote" einer Methode, die „mit der Behandlung verbundenen Belastungen", die „Folgen der Behandlung" und die Gefahren der Unterlassung der Behandlung stellen Informationen dar, die nicht durch gängige *eXplainable-AI*-Techniken (XAI),[41] sondern auf empirischem Weg, primär durch die klinische Bewertung, gewonnen werden können. In einigen Fällen dürften die dafür notwendigen Informationen bereits vorliegen, bspw. wenn ein Modell auf Auswahl und Dosierung zwischen verschiedenen, bereits erprobten Medikamenten eingeschränkt ist.

Die notwendigen Angaben korrespondieren dabei teilweise mit der in Anh. I Regel 23.4 MPVO normierten Gebrauchsanweisung für Medizinprodukte. So informiert bspw. die „Zweckbestimmung des Produkts mit einer genauen Angabe der Indikationen, Kontraindikationen, Patientenzielgruppe[n]" bereits über „Art und Wesen" einer Behandlung. Da die MPVO nicht speziell auf „intelligente Medizinprodukte" abgestimmt ist, könnte Art. 13 Abs. 2, 3 AIA die Anforderungen an Gebrauchsanweisungen weiter konkretisieren. So korrespondieren die dort normierten Angaben über das Maß „an Genauigkeit […] für das das Hochrisiko-KI-System getestet und validiert wurde und das zu erwarten ist" mit den „Erfolgsaussichten" einer Methode.[42] Die weiteren notwendigen Angaben zur Behandlung erfordern dabei m.E., wie bereits dargelegt, empirische Evaluierung.

Besteht eine „echte Auswahl", d.h. Methoden, die gleichwertig sind, aber unterschiedliche Risiken und Erfolgschancen haben, sind die Vor- und Nachteile der Alternativen gemeinsam abzuwägen.[43] So bieten sich

40 *C. Katzenmeier*, KI in der Medizin – Haftungsfragen, MedR 2021, 859 (861); *G. Spindler*, Medizin und IT, insbesondere Arzthaftungs- und IT-Sicherheitsrecht, in: C. Katzenmeier (Hrsg.), Festschrift für Dieter Hart. Medizin – Recht – Wissenschaft, Berlin 2020, S. 581 (602). Allgemein zu „Außenseitermethoden" *Memmer*, Aufklärung (Fn. 12), I.3.3.2.3.

41 *C. Molnar*, Interpretable Machine Learning, 2. Aufl., München 2022.

42 Siehe auch Anh. I Regel 15.1 MPVO zu diagnostischen Produkten.

43 *Memmer*, Aufklärung (Fn. 12), I.3.2.1.2.2.

Vergleichsstudien an, um die Vor- und Nachteile einer „menschlichen" und einer „KI-Behandlung" zu kontrastieren.[44]

Um die Erfolgsaussichten von Alternativen zu vergleichen, reichen jedoch übliche Performanzmetriken wie Genauigkeit (*accuracy*),[45] die häufig nur anhand des Testdatensatzes und nicht „im Feld" ermittelt werden, nicht aus. Denn die im „Labor" erzielten Vorteile lassen sich vielfach nicht auf die Realität übertragen. Die so ermittelte Genauigkeit kann potentiell einen fehlerhaften Entscheidungsprozess des Modells i.S.e. „Kluger-Hans"-Effekts[46] verbergen. Neben der Genauigkeit müssen der Patientin daher auch andere Performanzangaben wie die *false-positive-* oder *false-negative-*Rate verständlich mitgeteilt werden.[47]

Um einer Patientin eine echte Wahl zu ermöglichen, sind daneben weiterführende Informationen über Vor- und Nachteile, eine verschieden starke Intensität des Eingriffs, die unterschiedlichen Risiken, Schmerzbelastungen und Erfolgsaussichten zu erteilen.[48] So könnte bspw. darüber aufzuklären sein, dass eine automatisierte Operation weniger invasiv ist und die Konvaleszenz rascher erfolgt. Wiederum ist wahrscheinlich, dass sich eine KI-gestützte Behandlung in Zukunft zum medizinischen Standard entwickeln und somit nicht ungefragt über Alternativen zu dieser aufzuklären sein wird.

3. Risikoaufklärung

Als dritte Komponente der Selbstbestimmungsaufklärung fungiert die Risikoaufklärung. Die Ärztin hat über mögliche Gefahren der Behandlung aufzuklären, damit die Patientin eine Vorstellung erhält, welche Komplikationen selbst dann auftreten können, wenn die Behandlung entsprechend *lege artis* und mit größter ärztlicher Sorgfalt durchgeführt wird.[49] Ihr muss die Möglichkeit unvermeidbarer medizinischer Fehlschläge oder Folgeschäden mitgeteilt werden, soweit sie wissenschaftlich bekannt sind

---

44 *Schiff/Borenstein*, Clinicians (Fn. 29), 140 f.

45 *F. Cabitza*, Breeding electric zebras in the fields of Medicine, 2017, https://arxiv.org/abs/1701.04077, S. 5 (zuletzt abgerufen: 09.12.2022).

46 *H. Chang Yoon/R. Torrance/N. Scheinerman*, Machine learning in medicine, J Med Ethics 2021, 1 (2).

47 *P. Hacker/R. Krestel/S. Grundmann/F. Naumann*, Explainable AI under contract and tort law, Artificial Intelligence and Law 2020, 415 (421).

48 *Memmer*, Aufklärung (Fn. 12), I.3.2.1.2.2.

49 *Memmer*, Aufklärung (Fn. 12), I.3.2.1.3.

und einigermaßen ernsthaft in Betracht kommen.[50] Mit der Einwilligung nimmt die Patientin die Verwirklichung dieser Risiken in Kauf.[51] Die Grenzen zwischen der Risikoaufklärung und der Verlaufsaufklärung sind dabei, wie die obigen Ausführungen zu den Erfolgschancen einer Behandlung zeigen, teilweise fließend.

Zunächst ist im Rahmen der Risikoaufklärung über sog. „eingriffsspezifische Gefahren" aufzuklären. Die „Typizität" ergibt sich nicht aus der Komplikationshäufigkeit, sondern aus dem Umstand, dass das Risiko speziell dem geplanten Eingriff anhaftet und auch bei Anwendung allergrößter Sorgfalt und fehlerfreier Durchführung nicht sicher zu vermeiden ist.[52] Bei KI-Systemen muss daher auch über schwer erkenn-, aber erwartbare Fehlleistungen aufgeklärt werden.[53] Da nur auf zum Zeitpunkt der Behandlung bekannte, typische Gefahren hinzuweisen ist, ist m.E. wiederum der Hinweis auf KI als „Außenseitermethode" notwendig, deren Risiken sich noch nicht vollständig abschätzen lassen.

Dabei ist wahrscheinlich, dass bspw. automatisierte Chirurgie andere „typische" Gefahren (z.B. Übertragung von Elektroschocks, unbemerkt abbrechende Instrumententeile)[54] aufweist, über die aufgeklärt werden muss. Damit korrespondierend muss die Patientin im Rahmen der Sicherungsaufklärung darüber aufgeklärt werden, dass z.B. eine unerwartete Bewegung eine Verletzung durch die autonomen Roboterarme hervorrufen könnte.[55]

Hauptquelle für die Risikoaufklärung könnten wiederum die Gebrauchsanweisungen nach der MPVO, primär aber Art. 13 Abs. 2, 3 AIA, darstellen. Diese haben Informationen über „alle bekannten und vorhersehbaren Umstände, die sich auf das erwartete Maß an Genauigkeit […] auswirken" oder „[…] die zu Risiken für die Gesundheit und Sicherheit oder die Grundrechte führen" können und Informationen zur „Leistung bezüglich der Personen oder Personengruppen, auf die das System bestimmungsgemäß angewandt werden soll", zu enthalten. Daher müsste angege-

---

50 *Engljähringer*, Aufklärungspflicht (Fn. 17), S. 12; *Prutsch*, Aufklärung (Fn. 17), S. 142 f.

51 *Jesser-Huß*, Haftung (Fn. 17), Rn. 83.

52 *Neumayr* (Fn. 1), Einleitung ABGB Rn. 44.

53 *E. Paar/K. Stöger*, Medizinische KI, in: Fritz/Tomaschek (Hrsg.), Konnektivität, Münster 2021, S. 85 (89).

54 *E. Silvestrini*, da Vinci Robotic Surgery Complications, 2021, abrufbar unter https://www.drugwatch.com/davinci-surgery/complications/ (zuletzt abgerufen: 09.12.2022).

55 *L. Blechschmitt*, Die straf- und zivilrechtliche Haftung des Arztes beim Einsatz roboterassistierter Chirurgie, Baden-Baden 2017, S. 75.

ben werden, wenn z.B. die Beleuchtung im Raum negativen Einfluss auf ein Bilderkennungssystem haben könnte, wenn ein chirurgisches System bei einer bestimmten Haltung der Patientin zu unerwarteten Schnittbewegungen neigt oder ein System in Bezug auf eine Patientengruppe, z.B. mangels entsprechender Trainingsdaten, eine geringere Performanz aufweist. Unter die „bekannten oder vorhersehbaren Umstände[,][…] die zu Risiken für die Gesundheit" führen können, fallen m.E. auch sog. „patientenspezifische Risiken". Diese ergeben sich spezifisch aus in der Sphäre der Patientin liegenden Faktoren (bspw. körperliche Merkmale wie Alter, Gewicht, Allgemeinkonstitution, anatomische Regelwidrigkeiten).[56] Daher könnte eine Aufklärung darüber notwendig sein, dass z.B. anatomische Anomalien bei einer Behandlung durch ML zu erhöhten Behandlungsrisiken führen.

Unter Umständen hat eine Ärztin auch über die personelle und apparative Ausstattung aufzuklären. Da nicht jeder Patientin eine Behandlung mit den neuesten Apparaten geboten werden kann, muss sie – wenn die neuen Methoden gravierende Vorteile bringen – über diese informiert werden, um ihr die Option zu eröffnen, diese Behandlung, z.B. in einem anderen Krankenhaus, in Anspruch zu nehmen.[57] Zukünftig muss daher darüber aufgeklärt werden, dass eine KI-gestützte Behandlung als neue Methode in einer anderen Praxis angeboten wird oder dass ein Krankenhaus über bessere KI-Modelle verfügt.

Als Zwischenfazit ist festzustellen, dass sich die Aufklärung, die der Patientin Selbstbestimmung ermöglichen soll, in der Praxis häufig auf den Hinweis auf den Einsatz von KI und auf eine verständliche Darlegung der „Gebrauchsanweisung" reduziert. Dies umfasst eine allgemeine Systembeschreibung in Verbindung mit Angaben zur Leistung, die mittels klinischer Bewertung gesichert sein sollte. Eine präzisere Aufklärung, die potentiell den Einsatz von XAI-Techniken notwendig macht, ist nach dieser Interpretation nicht Kernbestandteil der Aufklärung. Diese Interpretation der medizinischen Aufklärung kann jedoch kritisch hinterfragt werden. Eine Reihe von potentiellen Argumenten lassen sich vorbringen.

56 *Memmer*, Aufklärung (Fn. 12), I.3.2.1.3.2; *Neumayr* (Fn. 1), Einleitung ABGB Rn. 45.

57 *Memmer*, Aufklärung (Fn. 12), I.3.2.1.3.3.

C. Aufklärung im (technischen) Wandel

I. AIA und Transparenz

Welche Rolle Transparenz oder Interpretierbarkeit im AIA zukommt, ist zum gegenwärtigen Zeitpunkt Gegenstand von Diskussionen. Art. 13 Abs. 1 AIA verlangt zwar, dass „Hochrisiko-KI-Systeme […] so konzipiert und entwickelt [werden], dass ihr Betrieb hinreichend transparent ist, damit die Nutzer die Ergebnisse des Systems angemessen interpretieren und verwenden können", wobei die Transparenz „[…] auf eine geeignete Art und in einem angemessenen Maß gewährleistet [wird], damit die Nutzer und Anbieter ihre in Kapitel 3 dieses Titels festgelegten einschlägigen Pflichten erfüllen können."

Jedoch verzichtet der AIA auf eine nähere Definition der verwendeten Begriffe Transparenz und Interpretierbarkeit. Die konkrete Interpretation und der Bezug zu Termini wie Erklärbarkeit, Rechtfertigung oder Accountability bleibt damit offen.[58] Es ist jedoch kaum zu erwarten, dass Art. 13 Abs. 1 AIA die Verwendung von White-Box-Modellen oder die detaillierte Erklärung einer konkreten Entscheidung erfordert.[59] Vielmehr ist wahrscheinlich, dass die nähere Implementation von „Transparenz" im Ermessen der Nutzerinnen liegt bzw. erst von den Europäischen Normungsorganisationen durch harmonisierte Standards mit Leben erfüllt wird.

Zahlreiche Fragen wirft die *human-in-the-loop*-Regelung gem. Art. 14 AIA auf. Dieser normiert die Einrichtung einer wirksamen menschlichen Aufsicht. Die zuständigen Personen müssen gem. Art. 14 Abs. 3 AIA die Fähigkeiten und Grenzen eines Hochrisikosystems vollständig verstehen und die Ergebnisse richtig interpretieren können. Als zentraler Aspekt sind „[…] insbesondere […] die vorhandenen Interpretationswerkzeuge und -methoden zu berücksichtigen". Diese Maßnahmen, um die Interpretation zu erleichtern, sind auch Teil der Gebrauchsanweisung (Art. 13 Abs. 3 lit. d AIA). Denn nach Art. 29 Abs. 4 AIA müssen die Nutzerinnen ein System anhand der Gebrauchsanweisung überwachen sowie das Vor-

58 *D. Bomhard/M. Merkle*, Europäische KI-Verordnung, RDi 2021, 276 (280); *M. Ebers*, Standardisierung Künstlicher Intelligenz und KI-Verordnungsvorschlag, RDi 2021, 588 (590); *P. Hacker/J.-H. Passoth*, Varieties of AI Explanations Under the Law, in A. Holzinger/R. Goebel/R. Fong/T. Moon/K.-R. Müller/W. Samek (Hrsg.), xxAI, Cham 2022, S. 343 (358 ff.).

59 So wurde bspw nach *Roth-Isigkeit*, Transparenzanforderungen (Fn. 5), 284 zugunsten einer Systemtransparenz, abgesichert durch Begleitpflichten, auf eine Erklärung einer konkreten Entscheidung verzichtet.

liegen eines Risikos oder einer Fehlfunktion melden und die Verwendung aussetzen können.

Ähnlich wird in der medizinrechtlichen Literatur aufgrund der ärztlichen Sorgfalt (in Österreich: § 49 ÄrzteG und § 8 Abs. 2 KAKuG) argumentiert, dass Ergebnisse eines ML-Systems wie Diagnosen einer Plausibilitätsprüfung unterzogen werden müssen.[60] Art. 14 AIA reflektiert damit nach *Buchner* den Gedanken, dass automatisierte Systeme akzeptabel sind, wenn sie von Menschen verstanden und hinterfragt, kontrolliert, überstimmt oder abgeschaltet werden können.[61]

Die Rolle der Interpretationswerkzeuge in Bezug auf Art. 14 AIA ist jedoch m.E. noch zu undefiniert. Ist das Ziel die Vermeidung grober Fehler, z.B. offensichtlicher Fehldiagnosen, sind XAI-Methoden potentiell überflüssig. Ist das Ziel die Aufdeckung subtiler Fehler, können diese Fehler mittels XAI-Methoden nicht zuverlässig und rasch genug evaluiert werden, um eine Aufsicht sicherzustellen. Denn wie in der Literatur[62] festgehalten wurde, erlaubt bspw. eine *heatmap*, die die relevanten Regionen z.B. auf einem Röntgenbild hervorhebt, keinen Schluss darauf, ob medizinisch relevante Faktoren für das Ergebnis ausschlaggebend waren und die Diagnose normativ akzeptiert werden kann.[63] Die Folge könnte sein, dass eine Ärztin aufgrund der eigenen (begrenzten) Fachkenntnisse[64] ihr als logisch erscheinende Ergebnisse bestätigt und damit potentiell Fehler vermeidet, aber auch ihren persönlichen Bias einbringt.

Art. 14 AIA deutet m.E. stärker als Art. 13 Abs. 1 AIA an, dass XAI-Methoden eine bedeutende Rolle im Gefüge des AIA spielen. Welche genau dies im Gefüge der *human-in-the-loop*-Regelung ist, z.B. ob Art. 14 AIA die Pflicht, XAI zu implementieren, beinhaltet, bedarf noch einer Klarstellung im Gesetzgebungsprozess.

---

60 *Paar/Stöger*, KI (Fn. 53), 86.

61 *B. Buchner*, Artificial intelligence as a challenge for the law, Int. Cybersecur. Law Rev. 2022, 181 (184).

62 *M. Ghassemi/L. Oakden-Rayner/A. L. Beam*, The false hope of current approaches to explainable artificial intelligence in health care, Lancet Digit Health 2021, E745; a.A. *J. Amann/A. Blasimme/E. Vayena/D. Frey/V. I. Madai*, Explainability for artificial intelligence in healthcare, BMC Medical Informatics and Decision Making 2020, 1 (2 f.); *S. Reddy*, Explainability and artificial intelligence in medicine, Lancet Digit Health 2022, E745.

63 *Roth-Isigkeit*, Transparenzanforderungen (Fn. 5), 280: „Bei komplexen Gewichtungsvorgängen (wie z.B. Diagnosevorschlägen) ist hingegen weitgehend unklar, inwieweit das sichtbar gemachte Ausgabeergebnis des Arbeitsprozesses für Menschen verständlich wäre."

64 *Eberbach*, Aufklärung (Fn. 28), 3.

II. KI-Systeme als „Aspirin 2.0“

Die häufig vorgebrachte Analogie von KI zu Arzneimitteln, Röntgengeräten und anderen Diagnoseinstrumenten, über die rechtlich nicht näher aufgeklärt werden muss bzw. über die eine Ärztin selbst häufig nicht im Detail Bescheid weiß, könnte sich m.E. in Zukunft als unhaltbar erweisen.[65]

Zwar genügt, wie ausgeführt, zum derzeitigen Stand m.E. in Bezug auf Modelle, die auf einen eng begrenzten Zweck, wie Hautkrebserkennung, zugeschnitten sind, eine allgemeine Funktionsbeschreibung, die eine grobe Vorstellung des Modells ermöglicht, als Substrat der Aufklärung.

Zunehmend werden dem technischen Trend entsprechend jedoch komplexe Modelle wie GPT-3 konstruiert, die für verschiedene Zwecke eingesetzt werden. Würde z.B. ein Modell als Teil der Systemmedizin[66] im Rahmen einer Vorsorgeuntersuchung zahlreiche verschiedene Datenquellen wie die Krankenakte, genetische Daten, Umwelteinflüsse aus Sensordaten und die Social-Media-Aktivität einbeziehen, erlaubt eine abstrakte Funktionsbeschreibung keine Vorstellung von der Funktionsweise und dürfte für eine Patientin jeden Bezug zu ihrer Diagnose verlieren. Würde eine Aufklärung über ein System, das über hundert Krankheiten feststellen kann, i.S.v. „Sie werden innerhalb der nächsten Monate mit 96% Wahrscheinlichkeit Schizophrenie ausbilden. Das dafür verwendete Modell wurde mit Daten aus der Krankengeschichte, genetischen und Social-Media-Daten trainiert. Das Modell ist nach Evaluierungen zu 87% genau.“ eine selbstbestimmte Entscheidung ermöglichen? Im Gegensatz zu anderen Methoden scheint es zweifelhaft, ob die Ärztin in dieser Situation auch nur eine grobe Vorstellung von der Funktionsweise eines solchen Modells aufweist, die zum Inhalt der Aufklärung werden könnte.[67] Im Kontrast zu Röntgengeräten ist das theoretische Grundgerüst von Modellen, wie ein Verweis auf Systeme zur Emotionserkennung zeigt,[68] häufig nur gering ausgeprägt oder umstritten.

Auch können XAI-Methoden in solchen Konstellationen eine Aufklärung nur bedingt effektuieren. So könnte in einem Idealszenario eine sog. lokale *post-hoc*-Technik wie LIME oder SHAP, die z.B. die entscheidenden

---

65 *T. P. Quinn/S. Jacobs/M. Senadeera/V. Le/S. Coghlan*, The three ghosts of medical AI, Artificial Intelligence In Medicine 2022, 1 (5 f.).

66 *Katzenmeier*, Data (Fn. 5), 259 f.

67 *Matulionyte/Nolan/Magrebi/Beheshti*, Devices (Fn. 29), S. 21.

68 *M. Veale/F. Zuiderveen Borgesius*, Demystifying the Draft EU Artificial Intelligence Act, CRi 2021, 97 (107 f.).

Symptome als Merkmale für eine konkrete Diagnose visualisieren oder reihen, eine Diagnoseaufklärung ermöglichen.[69] Dies setzt jedoch m.E. voraus, dass die hervorgehobenen Merkmale und ihr Bezug zur Diagnose (kausal) nachvollziehbar sind. Bei Modellen mit tausenden Merkmalen, wie dem Tippverhalten am Handy und komplexen Interaktionen, z.B. zwischen Genetik und Umwelt,[70] erlauben solche Methoden jedoch nur bedingten Einblick und begrenztes Verständnis für ein Ergebnis. Aufgrund der korrelationsbasierten Natur von ML ist generell fraglich, ob es eine Erklärung geben kann, die sich in das kausale Denken von Ärztin und Patientin einfügt und somit ein Hinterfragen der Diagnose ermöglicht.

III. Stand der Wissenschaft und Black Box

Eine Frage, die eng mit der Aufklärung über Behandlungsalternativen verzahnt ist, ist, ob Black-Box-Modelle zum neuen Stand der medizinischen Wissenschaft werden könnten oder ob ihre Opazität dem entgegensteht. Nach *Hacker et. al.* reicht höhere Genauigkeit dafür nicht aus. Vielmehr würde die Evaluierung der unterschiedlichen Risiken, ob im konkreten Fall die Vorteile einer Methode die Nachteile überwiegen, häufig den Einsatz von lokalen XAI-Methoden implizieren. Denn die fehlende Interpretierbarkeit wäre trotz der größeren Leistung opaker Modelle mit Risiken für die Patientin verbunden. Ein Modell könnte bspw. nicht auf das Vorliegen von *false negatives* oder *false positives* evaluiert werden. Die Verwendung von Black-Box-Modellen könnte daher nur in Sonderfällen gerechtfertigt werden, z.B. wenn die höhere Performanz die Risiken übertreffen würde oder wenn Fehler auch ohne Interpretierbarkeit aufgedeckt werden könnten. Wäre Interpretierbarkeit zur Aufdeckung von Fehlern notwendig bzw. würde sie diese erleichtern, wäre der Einsatz interpretierbarer Modelle durch den Stand der Wissenschaft geboten. Ebenso kann nach *Hacker et al.* ein Black-Box-Modell den Stand der Wissenschaft nicht „anheben" und damit verpflichtend werden, da ein solches Modell, dessen Ergebnisse nicht kritisch hinterfragt werden können, nicht sinnvoll in den medizinischen Arbeitsablauf eingegliedert werden könnte.[71]

---

69 Ähnlich im Datenschutzrecht, *T. Hoeren/M. Niehoff*, Artificial Intelligence in Medical Diagnoses and the Right to Explanation, EDPL 2018, 308 (317 f.).

70 *W. N. Price, II.*, Medical Malpractice and Black Box Medicine, in: I. Glenn/H. Fernandez Lynch/E. Vayena/U. Gasser (Hrsg.), Big Data, Health Law and Bioethics, Cambridge 2018, S. 295 (297).

71 *Hacker/Krestel/Grundmann/Naumann*, AI (Fn. 47), 421 f.

Damit wäre in vielen Fällen der Einsatz von XAI-Methoden notwendig, damit ein Modell zum neuen Stand der Wissenschaft werden könnte. M.E. ist den Autoren jedenfalls darin zuzustimmen, dass nicht grundlos ein Black-Box-Modell verwendet werden sollte, wenn ein ähnlich performantes White-Box-Modell zur Verfügung steht.

IV. Generierung von klinischer Evidenz

Auch wenn klinische Bewertung, auf die z.B. in der Risikoaufklärung verwiesen wird, für intelligente Medizinprodukte eine größere Rolle spielen dürfte als XAI-Techniken, sind damit doch Probleme verbunden. Zunächst herrscht ein Mangel an Studien, die klinische Wirksamkeit und nicht Effizienz im „Labor" belegen.[72] Die Qualität existierender Studien gilt als zweifelhaft.[73] Selbst die Anwendbarkeit von randomisierten klinischen Studien auf ML-Systeme wird teilweise hinterfragt.[74] Auch wenn klinische Evidenz existiert, haben Ärztinnen potentiell keinen Zugang dazu, da Studien nicht veröffentlicht werden müssen, wenn Interessen des Geheimnisschutzes dem entgegenstehen.[75]

Ein Verweis auf eine klinische Bewertung als Teil der Aufklärung ist somit gegenwärtig selten realistisch. Konzeptuell können klinische Studien als Validierungsschritt auch nur bedingt die vorhergehende Hypothesenbildung, die bei ML-Systemen häufig fehlt, ersetzen.[76] Bis ausreichend klinische Evidenz vorliegt, könnten XAI-Methoden dazu beitragen, Ärztinnen eine Plausibilitätsprüfung und eine Abschätzung des Risikos zu ermöglichen. Für die Generierung von klinischer Evidenz könnten wiederum Methoden wie *causal inference*,[77] die Kausalitäten und nicht nur Korrelationen modelliert, zur notwendigen Hypothesenbildung beitragen.

---

72 *E. Topol*, High-performance medicine, nature medicine 2019, 44 (45).
73 *S. Scholz*, Evidenzbasierter Einsatz von KI in der Medizin, in: I. Eisenberger/R. Niemann/M. Wendland (Hrsg.), Smart Regulation. Symposium 2021, Tübingen 2022, i.E.
74 *Price*, Malpractice (Fn. 70), S. 301 f.
75 *Scholz*, Einsatz (Fn. 73), i.E.
76 *S. Kundu*, AI in medicine must be explainable, Nat Med 2021, 1328 (1328).
77 *M. Prosperi/Y. Guo/M. Sperrin/J. S. Koopman/J. S. Min/X. He/S. Rich/M. Wang/I. E. Buchan/J. Bian*, Causal inference and counterfactual prediction in machine learning for actionable healthcare, Nature Machine Intelligence 2020, 369.

V. (X)AI-Entwicklungen

Auch wenn, wie erwähnt, derzeitig gebräuchliche XAI-Techniken nur begrenzt den Anforderungen der medizinischen Aufklärung i.S.e. personalisierten, auf die Patientin zugeschnittenen Gesprächs entsprechen, so ist die technische Entwicklung nicht statisch. Diese Entwicklungen könnten in Hinblick auf den Stand der Medizin dazu beitragen, dass nicht nur bessere intelligente Medizinprodukte, sondern auch interpretierbarere Medizinprodukte zunehmend dem *state of the art* entsprechen.

So ist ein Trend zu beobachten, neben neuen XAI-Verfahren auch einheitliche Evaluierungsmetriken zu konzipieren, die helfen, die „Brauchbarkeit" einer Erklärung, z.B. den Grad der „Kausalität" oder die notwendige Anpassung an das „Zielpublikum", zu evaluieren und zu vergleichen.[78]

Auch sog. „neuro-symbolische" Techniken, die die Ansätze von Expertensystemen, deren Regeln durch fachlichen Input z.B. von Ärztinnen manuell konstruiert werden, mit maschinellem Lernen verbinden und somit potentiell bessere Interpretierbarkeit aufweisen,[79] könnten m.E. dazu beitragen, dass das Ideal einer gemeinsamen, partnerschaftlichen Entscheidungsfindung aufrecht bleibt und die Rolle der Ärztin als Entscheidungsträgerin nicht unterminiert wird.

D. Conclusio

Jonathan Zittrain spricht von der potentiellen „intellektuellen Schuld",[80] die wir uns aufladen, wenn wir zwar wissen, dass ein Produkt funktioniert, aber die Antwort auf das „Warum?" in die Zukunft verschieben. Die zwei Abschnitte dieses Beitrags haben versucht, die Frage, ob wir die Schuld einer „Black-Box-Medizin" in Kauf nehmen sollen, aus zwei unterschiedlichen Blickwinkeln zu betrachten. Sollen wir die Idealvorstellung der Auf-

78 *A. Holzinger/G. Langs/H. Denk/K. Zatloukal/H. Müller*, Causability and explainability of artificial intelligence in medicine, WIRE's Data Mining and Knowledge Discovery 2019, 1.

79 Mitteilung der Kommission an das Europäische Parlament, den Rat, den Europäischen Wirtschafts- und Sozialausschuss und den Ausschuss der Regionen. Weißbuch Zur Künstlichen Intelligenz – ein europäisches Konzept für Exzellenz und Vertrauen, COM (2020) 65 final 5.

80 *J. Zittrain*, Intellectual Debt, in: S. Vöneky/P. Kellmeyer/O. Müller/W. Burgard (Hrsg.), The Cambridge Handbook of Responsible Artificial Intelligence, Cambridge/New York/Port Melbourne/New Delhi/Singapore 2022, S. 176.

klärung als Grundbaustein für eine partnerschaftliche Entscheidungsfindung aufgeben? Sollen wir zugunsten des utilitaristischen Lockrufs einer (potentiell) „besseren Medizin für alle" zulassen, dass das persönliche Aufklärungsgespräch[81] durch allgemeine Informationen und standardisierte Aufklärungsbögen[82] zu KI-Systemen abgelöst wird?

Zusammengefasst gilt es, in den Worten der (österreichischen) *Bioethikkommission*, „abzuwägen zwischen den Möglichkeiten neue Diagnosemöglichkeiten zu erhalten […] und der für ärztliches Handeln fundamentalen dialogischen Komponente, die dem Begründen und Einsichtigmachen – ‚telling a story' – großen Wert beimisst."[83] Mit Blick auf den AIA und seine noch zu konkretisierenden Transparenzanforderungen, einer Rechtsprechung, die der Aufklärung eine immer größere Rolle beimisst, und Entwicklungen im XAI-Bereich ist m.E. noch offen, welche intellektuelle Schuld wir zugunsten intelligenter Medizinprodukte rechtlich auf uns nehmen werden.

81 *Memmer*, Aufklärung (Fn. 12), I.3.7.1; *Neumayr* (Fn. 1), Einleitung ABGB Rn. 55.

82 *Memmer*, Aufklärung (Fn. 12), I.3.7.3.1; *Neumayr* (Fn. 1), Einleitung ABGB Rn. 54.

83 *Bioethikkommission*, Ärztliches Handeln im Spannungsfeld von Big Data, Künstlicher Intelligenz und menschlicher Erfahrung, 2020, abrufbar unter https://www.bundeskanzleramt.gv.at/dam/jcr:b3a4fdd0-0d89-430e-ae35-3ad5d991a3e4/200617_Stellungnahme_BigData_Bioethik_A4.pdf (zuletzt abgerufen: 09.12.2022), S. 51.