

Künstliche Intelligenz im ärztlichen Alltag: ethische Herausforderungen

M. Herrmann¹, E. Winkler¹

In den letzten Jahren hat die Entwicklung medizinischer KI-Anwendungen Hoffnungen in eine Revolution der Versorgung geweckt: Immer akkuratere KI-Diagnosen wie auch neuartige Formen der Behandlungssteuerung dringen in alle Bereiche der Medizin vor – und es lockt die Aussicht auf immense Ressourceneinsparungen. So bietet Microsoft inzwischen ein KI-Tool an, welches das Schreiben von Arztbriefen überflüssig macht, indem es Patientengespräche dokumentiert und dabei zugleich den Arztbrief vorverfasst – ohne dass Monitor und Tastatur den Arzt vom Patienten ablenken [1].

All dies wurde dadurch möglich, dass nicht mehr jeder Schritt der Datenverarbeitung programmiert werden muss. Vielmehr entwickeln sich moderne KI-Algorithmen anhand medizinischer Trainingsdaten selbstständig in solch einer Tiefe weiter, wie sie sich von einem menschlichen Entwickler nicht programmieren, sogar mitunter nicht einmal nachvollziehen lassen.

Doch es ist sehr schnell die Erkenntnis gereift, dass moderne KI-Anwendungen ethischer Begleitung bedürfen. Aufsehen erregte es, als eine Bilderkennungs-KI von Facebook eine Gruppe junger Menschen schwarzer Hautfarbe als Affen kategorisierte [2]. Solch schreiende KI-Fehler sind aber beileibe kein Problem, das sich auf Life Style Apps beschränkt. Auch medizinische KI-Lösungen werden von diskriminierenden Unregelmäßigkeiten geplagt: So erkennt beispielsweise eine KI bei dunkelhäutigen Menschen im Vergleich zu

hellhäutigen erst unter deutlich höherer Symptomlast ein hohes Gesundheitsrisiko [3].

Die ethischen Aspekte des KI-Einsatzes in der Medizin in den Blick zu nehmen, ist daher unerlässlich.

KI und ärztliche Letztverantwortung

Furore machende KI-Fehler haben den Ruf nach einer menschlichen Kontrollinstanz laut werden lassen. In der Medizin lautet die Forderung, dass die menschliche Letztverantwortung für den KI-Einsatz beim behandelnden Arzt liegen solle [4]. Es liege an ihm, medizinische KI-Fehler zu erkennen und zu kontern.

„Kann ein Arzt die Letztverantwortung dafür tragen, aus einer Position der Ungewissheit heraus KI-Fehler von KI-Entdeckungen zu unterscheiden?“

Lange vor der Ankunft moderner KI-Modelle wurde bereits eingewandt, dass algorithmische Entscheidungshilfesysteme einer Black Box gleichen [5]. Wie ein Algorithmus zu seiner Empfehlung kommt, ist nicht einsichtig – und eine Rückfrage nach den Gründen, wie bei Kollegen, ist nicht möglich. Wie soll ein Arzt eine solche Empfehlung kritisch beurteilen, wenn die zugrundeliegende Rechtfertigung unzugänglich ist? Daher steht die Forderung nach einer Erklärbarkeit (Explainability) von KI-Systemen im Raum: Für den Endnutzer – in der Medizin für den Arzt – soll nachvollziehbar sein, wie die KI zu ihrer Ent-

scheidungsempfehlung gelangt. Dies wird beispielsweise dadurch versucht, dass für den Arzt hervorgehoben wird, welche Daten des Patienten ausschlaggebend für die Entscheidungsempfehlung der KI sind. In der Bildgebung geschieht dies anhand sogenannter Heat Maps: Wenn eine KI einen Fall als Lungentzündung kategorisiert, färbt sie dabei auf dem Röntgenbild diejenigen Bereiche der Lunge farbig auffällig ein, die für ihre Kategorisierung ausschlaggebend sind. Der Arzt kann dies nun genauer überprüfen und die Letztentscheidung über die Diagnose treffen.

Es ist jedoch zweifelhaft, ob eine solche Form der Erklärbarkeit wirklich zu einer kritischen Auseinandersetzung mit KI-Entscheidungsempfehlungen führt. Denn solche Heat Maps liefern keine kausale Erklärung für die Empfehlung. Sie zeigen nur an, welche Daten für die Berechnung relevant waren. Das „Warum“ bleibt unerklärt.

Angenommen eine Heat Map zeigt an, dass Bereiche der Bildgebung für die Berechnung relevant waren, die mit der bekannten Krankheitssymptomatik in keinerlei Verbindung stehen – beispielsweise, wenn eine KI ein Lungenkarzinom anzeigt, aber dabei auch Bereiche der Bildgebung hervorhebt, die weitab der Lunge sind und von denen kein Lehrbuch etwas im Zusammenhang mit Lungenkrebs erwähnt. Dann könnte man dies einfach als einen offensichtlichen KI-Fehler abtun. Allerdings ist gerade eine der Stärken moderner KI-Modelle, dass sie viel mehr Daten berücksichtigen können als es ein Mensch je vermag. Sie können dadurch Zusammenhänge erfassen, die der heutigen Medizin unbekannt sind.

¹ Nationales Centrum für Tumorerkrankungen Heidelberg (NCT)

Da nun eine Erkrankung wie Krebs systemweite Spuren hinterlässt, stellt sich für den behandelnden Arzt die Frage: Handelt es sich um ein bisher unbekanntes Symptom, das die KI aufgedeckt hat, oder ist es wieder einmal ein Beispiel für einen schreienden KI-Fehler?

In der Medizin geschieht es sehr oft, dass Ärzte aus einer Position der Ungewissheit heraus entscheiden müssen. Doch kann ein Arzt die Letztverantwortung dafür tragen, aus einer Position der Ungewissheit heraus KI-Fehler von KI-Entdeckungen zu unterscheiden? Auch wenn es sicher richtig ist, dass der Arzt Letztentscheider ist, so stellt sich die Frage, wo genau dabei die Grenzen der Letztverantwortung verlaufen sollten.

Eine Alternative sind KI-Systeme, deren Empfehlungen mit denen von Menschen vergleichbar sind – beispielsweise, wenn eine KI zur Melanomerkennung nicht nur anzeigt, welche Hautstellen für ihre Berechnung ausschlaggebend sind, sondern diese auch in einer Differenzialdiagnose mit Lehrbuchkategorien beschreibt [6]. Hier liegt eine Form der Erklärbarkeit vor, bei der menschlich zugängliche Kategorien zum Tragen kommen. Doch auch eine solche menschliche Erklärbarkeit hat ihren Preis: Sie verlangt, das ungeheure Potenzial von KI aufzugeben, alle Daten zu berücksichtigen und somit außerhalb bekannter menschlicher Kategorien Erkenntnisse zu gewinnen. In der Medizinethik ist daher die Diskussion noch nicht abgeschlossen, inwiefern Erklärbarkeit auf Kosten der Genauigkeit gehen darf.

Die Psychologie des Zusammenspiels von Arzt und KI

Wenn das Zusammenspiel von Mensch und Maschine gelingt, können Ärzte ihre Performance und somit die Qualität der Patientenversorgung merklich

verbessern [7]. Einem solchen Gelingen stehen mindestens zwei psychologische Stolpersteine im Weg: Automation Bias und Algorithm Aversion Bias.

Der Automation Bias bezeichnet die Neigung, maschineller Entscheidungsempfehlung unkritisch zu folgen. Ein blindes Vertrauen in die fortschrittliche Technik führt dazu, das eigene ärztliche Erfahrungswissen zu ignorieren. Da die Technik nun einmal nicht perfekt ist, riskiert man so vermeidbare Fehler.

„Eine KI kann immer nur Therapieoptionen bieten, die dann im Gespräch von Arzt und Patient zu erörtern sind“

Genau die gleiche Konsequenz hat das Gegenstück zum Automation Bias: der Algorithm Aversion Bias. Hier wird die KI-Empfehlung durchweg zu Gunsten der eigenen Expertise ignoriert. Was vermag schon eine KI einen erfahrenen Facharzt zu lehren? – so der Gedanke. Doch wie die Technik nicht perfekt ist, so ist man es selbst ebenfalls nicht. Wieder werden vermeidbare Fehler riskiert.

In diesem Zusammenspiel von Mensch und Maschine ist entscheidend, dass Ärzte andere Fehler machen als eine KI [8]. Gerade weil die Fehler so unterschiedlicher Art sein können, besteht die große Chance, sich gegenseitig zu ergänzen. Dies setzt jedoch voraus, dass der Letztentscheider, also der Arzt, beide Seiten ernst nimmt – sich und die Maschine.

Wenn seine Einschätzung von der Berechnung der KI abweicht, ist es daher unerlässlich, dies zum Ausgang einer inhaltlichen Auseinandersetzung

zu nehmen. Die Frage darf dann nicht lauten, welche der beiden Seiten generell besser ist. Vielmehr müssen alle Daten des vorliegenden Falls nochmals kritisch durchgegangen und anhand dieser die Entscheidung gefällt werden. Es hat sich gezeigt, dass solch ein Vorgehen im Zusammenspiel mit einer KI die besten Ergebnisse hervorbringt [9]. Daraus folgt aber, dass KI in ihrer Bedienung nicht simpel ist – wie intuitiv diese auch sein mag. Vielmehr müssen Ärzte im korrekten kritischen Umgang mit dieser neuen Technologie geschult werden.

Die Arzt-Patienten-Beziehung beim Einsatz von KI

Patientenautonomie ist einer der größten Werte der Medizinethik. Man ist in den letzten Jahren von dem Bild weggekommen, dass der Arzt dem Patienten vorschreibt, was zu tun ist. Die gemeinsame Entscheidungsfindung ist das Ideal, das an die Stelle einer paternalistischen Medizin getreten ist. Der Eintritt der KI in die Arzt-Patienten-Beziehung weckt jedoch die Sorge, dass dadurch die Uhr wieder zurückgedreht werden könnte.

Es ist in jedem Fall eine kritische Aufklärung durch den Arzt darüber notwendig, dass KI zum Einsatz kommt und dass ihr nicht zwingend zu folgen ist, nur weil es sich um eine KI handelt. Gerade bei Therapieempfehlungen ist eine kritische Distanz notwendig, da diese immer wertebasiert sind und kein generelles Falsch und Richtig kennen. Welche Therapieform zu wählen ist, hängt vom einzelnen Patienten ab – und zwar nicht nur von seiner Physis, sondern eben auch von seinen Überzeugungen.

Die Werte und Präferenzen eines Patienten lassen sich jedoch nicht in einer einfachen Abfrage ermitteln. Denn oft treten diese erst in einem hermeneuti-

schen Prozess zutage – sie zeigen sich erst im Patientengespräch. Eine KI kann daher in solchen Fällen immer nur Therapieoptionen bieten, die dann im Gespräch von Arzt und Patient zu erörtern sind. Es braucht also mehr denn je den Arzt als Vermittler und Vertrauensperson, damit nicht ein neuer „Computer-Paternalismus“ an die Stelle der Patientenautonomie tritt.

Daher wäre es grundverkehrt, die Ressourceneinsparungen durch KI zum Anlass zu nehmen, die Interaktion zwischen Arzt und Patient zugunsten einer zwischen KI und Patient zu verschieben. Vielmehr gilt es, die eingesparten Ressourcen zugunsten der seit Jahren vernachlässigten sprechenden Medizin einzusetzen.

Daten: der heikle Treibstoff von KI-Systemen

Die Durchbrüche der KI-Entwicklung der letzten Jahre wären nicht ohne entsprechende Daten möglich gewesen. Da sich eine KI anhand möglichst großer Trainingsdatensätze selbstständig weiterentwickelt, stellen Daten den Treibstoff dieser Technologie dar.

Doch der Einsatz von Patientendaten ist heikel. Zurecht ist deren Verwendung stark reglementiert, da ein Missbrauch katastrophale Auswirkungen für das Individuum nach sich ziehen kann. Es braucht daher eine ethisch akzeptable und rechtlich abgesicherte Form der Zustimmung von Patienten, um ihre Daten zu Trainingszwecken zu verwenden.

Im Besonderen braucht es klar definierte Nutzungsregeln, wenn Daten

aus dem solidarfinanzierten Gesundheitswesen zum Training einer von einem privaten Unternehmen entwickelten KI herangezogen werden. Als Google 2015 die Daten von 1,6 Millionen Patienten des National Health Service (NHS) erhielt, hatte dies einen öffentlichen Aufschrei sowie eine Sammelklage zur Folge [10]. Um die Bereitschaft zum dringend benötigten Teilen von Gesundheitsdaten zu Forschungszwecken nicht zu gefährden und ganz besonders um Patientenrechte nicht zu verletzen, ist die Entwicklung eines entsprechenden ethischen und rechtlichen Rahmens unerlässlich.

Aber es stellt nicht nur ein Problem dar, wie man ethisch akzeptabel Trainingsdaten zur KI-Entwicklung gewinnt, sondern auch deren Repräsentativität hat nicht unerhebliche ethische Konsequenzen. Da es sich meist um Daten aus dem Gesundheitssystem handelt, bilden sie auch den Status Quo dieses Systems ab. Das bedeutet aber, dass die KI eben diesen Status Quo lernt – mit all seinen Verzerrungen und Ungerechtigkeiten. Wenn eine KI bei dunkelhäutigen Menschen erst bei deutlich höherer Symptomlast als bei hellhäutigen eine Behandlungsempfehlung anzeigt, dann hängt dies nicht damit zusammen, dass der von den Entwicklern geschriebene Programmcode bereits solch diskriminierende Unterscheidungen enthält. Der Grund dafür ist, dass die KI dies aus den Trainingsdaten lernt. Die KI verinnerlicht die Missstände, die schon im Gesundheitssystem enthalten sind.

Das Zusammenspiel von Ethik und Technik

Selbst wenn man die gegenwärtig überschwängliche Euphorie herausrechnet, ist das Potenzial medizinischer KI-Anwendungen nicht weniger als enorm. Das darf aber nicht vergessen lassen, dass die Herausforderungen nicht minder gewichtig sind. Sie reichen von Fragen der Interaktion von Arzt und KI, über die Gefahr der Patientenbevormundung bis hin zu einer Technik, welche nicht die Altlasten der Vergangenheit mit sich schleppen darf. Nur dann wird sich das volle medizinische Potenzial dieser technischen Revolution ausschöpfen lassen, wenn sie in allen Entwicklungsschritten auch entsprechende ethische Begleitung erfährt. ■

Literatur unter www.slaek.de → Über Uns →
Presse → Ärzteblatt



Korrespondierende Autorin
Prof. Dr. med. Dr. phil. Eva Winkler
Geschäftsführende Direktorin NCT Heidelberg
Leiterin der Sektion für Translationale
Medizinethik; Universitätsklinikum
Heidelberg (UKHD), Medizinische Fakultät,
Universität Heidelberg
Vorsitzende der Zentralen Ethikkommission
bei der BÄK (ZEKO)
E-Mail eva.winkler@nct-heidelberg.de