



Zwischen fachlichen Standards und wilder Innovation: Zur Begutachtung von Big Data- und KI-Projekten in Forschungs-Ethikkommissionen

Andreas Brenneis  · Petra Gehring · Annegret Lamadé

Eingegangen: 1. Dezember 2023 / Angenommen: 21. Februar 2024
© The Author(s) 2024

Zusammenfassung In der Medizin (wie auch in anderen Fächern) wächst generell die Bedeutung informatischer Expertise. Schon das erfordert eine Kultur interdisziplinärer Begutachtung, auf welche medizinische Ethikkommissionen nicht gut eingestellt sind. Mit dem Einsatz von Big Data- und KI-Verfahren (sei es eigens entwickelter, sei es in Form von „Tools“) sind weitere Herausforderungen für die forschungsethische Begutachtung verbunden. Die Autoren schildern die Problemlage und schlagen vor, sie durch Verfahrensänderungen zu lösen. Besser als die Befassung zweier Kommissionen scheint eine Begutachtung geeignet zu sein, die von vornherein interdisziplinär verfährt. Das zöge allerdings die Forderung nach sich, die Zusammensetzung medizinischer Ethikkommissionen zu ändern. Weil Fragen nach der Organisation forschungsethischer Begutachtungen auch solche nach Kriterien aufrufen, denen diese zu folgen hat, stellt der Beitrag zudem erste Schritte zu einer Konsolidierung der Kriterien für eine forschungsethische Begutachtung von Big Data- und KI-Projekten vor.

Schlüsselwörter KI-Forschung · Datengetriebene Medizin · Ethikkommissionen · Bewertungskriterien · Dual Use · Forschungsförderung

Dieser Artikel erscheint im Rahmen des Themenheftes „Ethik in der datenintensiven medizinischen Forschung“.

✉ Dr. Andreas Brenneis · Prof. Dr. Petra Gehring
Institut für Philosophie, Technische Universität Darmstadt, Residenzschloss 1, 64283 Darmstadt,
Deutschland
E-Mail: andreas.brenneis@tu-darmstadt.de

Annegret Lamadé
Institut für das Recht der Digitalisierung, Philipps Universität Marburg, Marburg, Deutschland

Between academic standards and wild innovation: assessing big data and artificial intelligence projects in research ethics committees

Abstract

Definition of the problem In medicine, as well as in other disciplines, computer science expertise is becoming increasingly important. This requires a culture of interdisciplinary assessment, for which medical ethics committees are not well prepared. The use of big data and artificial intelligence (AI) methods (whether developed in-house or in the form of “tools”) pose further challenges for research ethics reviews. *Arguments* This paper describes the problems and suggests solving them through procedural changes.

Conclusion An assessment that is interdisciplinary from the outset appears to be more suitable than having two commissions with different expertise. However, this would require that the composition of medical ethics committees be altered. In addition, the article recommends initial measures to be taken during research ethics reviews of big data and AI projects in order to consolidate the review process and ensure standardization of the criteria.

Keywords Artificial intelligence research · Data-driven medicine · Ethics committees · Evaluation criteria · Dual use · Research funding

Einleitung

Das derzeit rasch wachsende Gewicht datenanalytischer Verfahren sowie von KI-Tools lässt Schwierigkeiten in der forschungsethischen Begutachtungspraxis entstehen: Man benötigt in der Ethikbegutachtung die Expertise mehrerer Disziplinen. In der Medizin (wie in ähnlicher Weise auch in anderen Fächern) wächst zum einen die Bedeutung informatischer Sachverhalte hinsichtlich forschungsethisch relevanter Details ganz generell – Aspekte des Datenmanagements, Schnittstellen zu kommerziellen Softwareanbietern, Anonymisierungs- und Pseudonymisierungsverfahren, Funktionsweisen von Algorithmen u. a. m. müssen betrachtet werden. Schon das erfordert eine Kultur interdisziplinärer Begutachtung, auf welche medizinische Ethikkommissionen nicht gut eingestellt sind. Mit datengetriebenen Analysen (Big Data) sowie mit KI-Verfahren (länger schon etwa Machine Learning, jetzt aber auch generativer KI) hat sich zum anderen die Lage speziell hinsichtlich – spätestens, wenn es um Produktentwicklung geht – nun auch gesetzlich geforderter Gütekriterien verschärft: Biases und Diskriminierungsgefahren sollten evaluiert werden können, „Black Box“-Probleme kommerzieller Software-Tools gilt es zu beurteilen, Tracking und Datenabflüsse werden brisanter. In einem Klima „wilder“ Innovation in der forschenden Informatik (wir nutzen den pointierten Ausdruck „wild“ in Anlehnung an Latour 1994, 1996¹), sind die Implikationen digitaler Werkzeuge aus der Sicht von Anwendungsdomänen dabei zunehmend schwerer zu erfassen. Was internationale

¹ Bruno Latour hat in den 1990er-Jahren die epistemischen Besonderheiten der durch die marktlichen Perspektiven der Gentechnik getriebenen Biologie bzw. Biochemie analysiert. Eine im weitesten Sinne ähnlich marktgetriebene Dynamik scheint – bei allen Unterschieden im Detail – in der Epoche des digitalen

Forschungsgemeinschaften für akzeptabel halten, wandelt sich schnell und entspricht keineswegs immer in der EU bzw. Deutschland geltenden Standards.² Seit neuestem sind zudem auch Dual Use-Fragen dringlicher geworden (vgl. DFG 2023). Im Bereich von Software und Datenströmen können solche Fragen besonders schwer zu beurteilen sein.

Medizinische Forschungsethik steht vor dem Hintergrund der stetig zunehmenden Möglich- wie Wirklichkeiten der Datenverarbeitung vor der Herausforderung, entsprechende Forschungsvorhaben ethisch zu evaluieren. Sollen sich medizinische Ethikkommissionen nun also schrittweise zu allumfassend kompetenten – nennen wir es: Superkommissionen entwickeln, die alle genannten Aspekte abfragen und mitbegutachten? Dies würde auf die Schaffung von vornherein interdisziplinär zusammengesetzter Kommissionen hinauslaufen. Ein alternativer Lösungsansatz für die Organisation der forschungsethischen Begutachtung datenintensiver medizinischer Forschung könnte die verstärkte Kooperation von medizinischen mit nichtmedizinischen Ethikkommissionen sein. Ziel dieses Beitrages ist es, beide Optionen vorzustellen und zu diskutieren.

Ergänzen möchten wir unsere Überlegungen zur Organisation von Ethikkommissionen durch Vorschläge zu – für die Forschung geeigneten – Kriterien der Ethikbegutachtung von anwendungsnaher datenintensiver Forschung (etwa mit Big Data- oder KI-Methoden). Diese sollten in erster Linie der Qualitätssicherung von Forschung dienen, gerade auch angesichts der sachlich notwendigen Unterscheidung zwischen der ethischen Bewertung von Big Data- oder KI-Produkten und der forschungsethischen Bewertung von Big Data- und KI-Projekten. Als Beispiel wird die *Handreichung zur forschungsethischen Begutachtung von KI-Forschungsprojekten* des hessischen Zentrums verantwortungsbewusste Digitalisierung vorgestellt.

Zur Bedeutung informatischer Expertise für forschungsethische Begutachtungen – auch in der Medizin

Forschungsanstrengungen im Bereich datengetriebener Medizin und digitaler Gesundheitsversorgung versprechen zahlreiche Innovationen und Verbesserungen für das Gesundheitswesen (WHO 2021; Acatech 2020; Plattform Lernende Systeme 2019). Durch die Nutzung von KI werden etwa in den Bereichen der personalisierten Medizin, bei der Unterstützung von Diagnosen und bei sektorübergreifenden und vernetzten Behandlungsformen, aber auch genereller bei der Optimierung von Abläufen wie etwa der Kommunikation zwischen medizinischem Personal und Patient:innen oder dem Entgegenwirken medizinischer Unterversorgung große Chancen auf Verbesserung gesehen (Beck et al. 2023; Budde 2020). Datengetriebene Medizin kann darüber hinaus bei der Befundung seltener Krankheiten unterstützen oder zur bestmöglichen Auswahl von Patient:innen für eine potenzielle Teilnahme

Wandels im Grundsatz auch die forschende, sich parallel aber bereits auf die unternehmerische Verwertung fokussierende Informatik zu treiben.

² Vgl. z. B. die Vorschläge der Datenethikkommission zu einer Ausweitung der *Corporate Digital Responsibility* (Datenethikkommission 2019). Es ist auch eine grundlegende Motivation des *Data Acts* der EU, marktformige Entwicklungsdynamiken einzuhegen.

an klinischen Studien beitragen (Caliebe et al. 2019). In einigen klinischen Anwendungsfeldern sind Systeme zur Unterstützung von Entscheidungen zum Teil schon erprobt. Dazu gehören die Bereiche der Diagnostik (etwa für die Auswertung radiologischer Befunde oder von Literaturrecherchen), der Therapie (für die präoperative Therapieplanung und die Unterstützung intraoperativen Vorgehens durch Verbesserung endoskopischer Navigation), der klinischen Prognostik (z. B. zur Risikoabschätzung bezüglich unerwünschter Zwischenfälle oder Überlebenswahrscheinlichkeiten) und bei der Prädikation von Krankheitsrisiken bei gesunden Menschen (für die Einteilung und erläuternde Beispiele vgl. ZEKO 2021).

Den Chancen stehen aber auch vieldiskutierte Risiken gegenüber. Für einen generellen Überblick und spezifische Handlungsempfehlungen zur Medizin als einem von vier analysierten Sektoren siehe (Deutscher Ethikrat 2023), für eine Übersicht zu durch KI in der Medizin induzierten Haftungsfragen (Katzenmeier 2021). Schon Studien mit großen Datenmengen und diversen Datentypiken bringen hinsichtlich wissenschaftlicher Standards viele Herausforderungen mit sich (Strech 2018). KI-Technologien wie Maschinelles Lernen, Neuronale Netze oder auf großen Sprachmodellen aufsetzende Formen generativer KI verschärfen die Problematik, insofern sie das Problem der Opazität in die Ergebnisfindung einführen. Bei der Nutzung entsprechender Anwendungen sind Transparenz und Nachvollziehbarkeit der Datenverarbeitung und des Entscheidungsweges nicht gegeben. Neben der „technischen“ Opazität kann eine mangelnde Transparenz von KI-Anwendungen im medizinischen Bereich auch dadurch begründet sein, dass die zugrundeliegenden Algorithmen durch Geschäftsgeheimnisse oder geistiges Eigentum geschützt sind.³ Darüber hinaus ist zu bedenken, dass KI-Anwendungen oftmals privatwirtschaftlich entwickelt und finanziert werden und die Entwickelnden dem ärztlichen Berufsethos nicht unbedingt unterliegen.

Vor dem Hintergrund der rasanten Entwicklungen hat etwa die Zentrale Ethikkommission bei der Bundesärztekammer eine Stellungnahme zu IT-basierten Systemen zur Entscheidungsunterstützung bzw. Clinical Decision Support Systems abgegeben (ZEKO 2021). Darin werden sieben Aspekte benannt, unter denen sich ethische Herausforderungen des Medizinsektors durch KI-Anwendungen beschreiben lassen. (1) Vertrauen muss demnach in zumindest dreifacher Hinsicht möglich sein: mit Blick auf die Reliabilität professionell genutzter KI-Systeme, in der Beziehung von medizinischem Personal und Patient:innen und übergeordnet schließlich in einem Systemvertrauen gegenüber dem medizinischen Sektor insgesamt. (2) Verantwortung müssen Ärzt:innen übernehmen können für den fachgerechten Umgang mit Medizingeräten und die kritische Prüfung von Vorschlägen („Automation Bias“) und sie obliegt höherstufig Einrichtungsträgern für die Sicherheit, Validierung und Diskriminierungsfreiheit von eingesetzten Systemen. (3) Die ärztliche Autonomie ist durch den schon genannten Automation Bias oder die unzureichende Reaktion

³ Burrell (2016) unterscheidet drei Formen der Opazität: Sie diskutiert neben der technischen Opazität, die für Machine Learning-Verfahren spezifisch ist und unter dem Stichwort Black Box erörtert wird („the way algorithms operate at the scale of application“), und der den Schutz von Geschäfts- oder Staatsgeheimnissen betreffenden Opazität („intentional corporate or state secrecy“) als dritte Form mangelndes Verständnis des Programmierens und von Code („technical illiteracy“), wozu im weiteren Sinn auch Datenkompetenz zu zählen ist.

auf Warnsignale („Alert Fatigue“) bedroht und Verschiebungen in der Systemnutzung von der Assistenz bei Entscheidungen zur Entscheidungsübernahme können zu Formen der Unachtsamkeit und verminderter Kontrolle führen. In den Bereich der Autonomie von Ärzt:innen gehört dabei höherstufig auch das Kompetenzfeld der Datenliteracy, zu dem beispielsweise eine Sensibilität gegenüber sich selbst erfüllenden Vorhersagen gehört, die durch KI-Systeme induziert sind. Das kann etwa dann der Fall sein, wenn auf der Basis schlechter Prognosen eine palliative statt einer kurativen Behandlung erfolgt und die ursprüngliche Datenbasis dadurch weiter bestätigt wird (Challen et al. 2019). Weniger autonome Entscheidungen von Ärzt:innen wirken sich auch auf die Standards der Patient:innenautonomie aus.⁴ (4) Kommunikation und Empathie zwischen medizinischem Personal und Patient:innen sind wichtige Faktoren für Wohlergehen und Gesundheit. Durch eine mit stärker automatisierten Abläufen einhergehende Vernachlässigung der Beziehungsebene zugunsten der Präsentation vermeintlich rein sachlicher Informationen könnten diese zwischenmenschlichen Interaktionen aus z. B. ökonomischen Gründen eine Zurücksetzung erfahren. (5) Auch ärztliches Erfahrungswissen ist wesentlich für die Ausübung der Heilkunde, zu der auch die Beurteilung spezifischer Lebensumstände gehört, und kann durch die Aufwertung maschinenlesbarer Daten an Bedeutung verlieren. (6) Diskriminierungsrisiken bestehen aufgrund der Bedeutung von Trainingsdaten und der Möglichkeit, dass diese fehler- oder lückenhaft sind oder falsch verarbeitet, z. B. falsch klassifiziert werden. In Gesellschaften verankerte Formen struktureller Diskriminierung und Stigmatisierung können sich so auch bei medizinischen Behandlungen auswirken. (7) Datensouveränität und Privatheit schließlich sind durch die Möglichkeiten von KI-Systemen ebenfalls herausgefordert, wobei Gesundheitsdaten nach Art. 9 Abs. 1 der Datenschutzgrundverordnung aufgrund ihrer Sensibilität der höchsten Datenschutzstufe unterliegen und sie zudem durch das Sozialgesetzbuch, v. a. §§ 291 ff. SGB V, und die ärztliche Schweigepflicht geschützt sind (Koepe 2020). Der Zuwachs von Gesundheitsdaten durch die Nutzung von KI-Anwendungen im Gesundheitswesen stellt nicht zuletzt den Datenschutz vor neue Herausforderungen: So entstehen Daten beispielsweise durch die Nutzung von Wearables oder KI-Applikationen auf dem Smartphone, aber auch schon bei Einkäufen oder Stichwortsuchen im Internet werden automatisch Daten produziert, die Einschätzungen über den Gesundheitszustand oder zum Teil sogar medizinische Vorhersagen erlauben (Deutscher Ethikrat 2017).

⁴ Eine patient:innengerechte Aufbereitung von medizinischen Informationen ist darüber hinaus insbesondere auch dann angeraten, wenn gar keine ausgebildeten oder approbierten Fachkräfte bei einer Diagnose beteiligt sind, wenn also etwa Chatbots oder Gesundheits-Apps die entsprechenden Daten zur Verfügung stellen. Generell sind in diesem Zusammenhang auch Risiken zu bedenken, die dadurch entstehen, dass KI-Systeme von Laien dazu gebraucht werden können, sich selbstständig über Krankheiten zu informieren – und evtl. ärztliche Hilfe dadurch nicht in Anspruch nehmen. Eine ähnlich gelagerte mögliche Folgeerscheinung durch den vermehrten Einsatz von KI- und datengetriebener Medizin ist die Möglichkeit und die daraus abgeleitete machtförmige Normalität von Technologien beispielsweise zur Überwachung von Vitalwerten durch Wearables. Daran ansetzende Befürchtungen sehen für die individuelle Ebene gefühlte Zwänge zur Optimierung des eigenen Lebensstils sowie zu zwanghaften Formen der Selbstüberwachung. Und damit einhergehend in gesellschaftlicher Perspektive eine neue Form sozial eingeforderter Eigenverantwortung inklusive eines Schuldprinzips für mangelhafte Gesundheitswerte – welches das Solidaritätsprinzip unterminiert.

Die genannten Aspekte sowie die Hinweise zu aktuell diskutierten Haftungsfragen und zur Opazität, aber etwa auch die Einstufung des Medizinsektors als Hochrisikobereich im *AI Act* (s. unten) zeigen, dass die medizinische Praxis und insbesondere die medizinische Forschung aktuell – jenseits der bekannten Problematiken des klassischen Datenschutzes – auch vor daten- und forschungsethischen Herausforderungen steht, wenn sie sich Big Data- und KI-Verfahren zunutze machen will. Diese Lage wollen wir im nächsten Abschnitt illustrieren.

Beispiele für forschungsethische Herausforderungen durch datengetriebene Forschung in der Medizin

Die Arbeit von Forschungs-Ethikkommissionen ist aus guten Gründen vertraulich, das Praxiswissen aus der Kommissionsarbeit findet so nicht hinreichend zurück in diesbezügliche Forschung. Auch wir stützen uns auf anekdotische Evidenz aus nur wenigen, selbst erlebten Fällen. Drei von diesen stellen wir hier idealtypisch verfremdet dar, so dass sie in ihrer Typik fassbar, aber nicht rückverfolgbar sind.

- Im Fall A handelt es sich um ein Verbundvorhaben, das ein zu entwickelndes digitales Medizinprodukt im Zuge des Entwicklungsprozesses mehrfach (denn die Optimierung der Lösung erfordert mehrfache Erhebung vergleichbarer Daten) an vulnerablen Patienten testen will. Zum einen wird die informatische Dimension (KI-Mustererkennung in großen Datenmengen) optimiert, zum anderen werden die erhobenen Daten medizinisch ausgewertet, um die durch das Produkt induzierten Therapieeffekte messbar zu machen (also auch für den gewählten Ansatz überhaupt erst hinreichend präzise zu modellieren).
- Im Fall B sollen ebenfalls in einem interdisziplinären Verbund große Datenmengen aus psychiatrischen Therapiegesprächen automatisiert ausgewertet werden. Auch hier ist ein medizinisches Interesse leitend, die anfallenden Sprachmuster mit Diagnosen zu korrelieren. Die sprachtechnologische Seite der Forschung verfügt jedoch ebenfalls noch nicht über medizinisch anerkannte Standards und ist also Teil des Experiments.
- Im Fall C sollen informatische Standardwerkzeuge kommerzieller Softwareanbieter eingesetzt werden, um verknüpft Telekommunikationsdaten zu erheben, während zugleich qualitative Interviews stattfinden sollen, die sich unter anderem auf medizinische Sachverhalte und die Versorgungslage in Katastrophen-/Krisenregionen beziehen. Die verknüpften Daten werden softwaregestützt umfassend (also auch hinsichtlich der Metadaten zur Einwahl) sowie in medizinischer Hinsicht evaluiert, um bislang noch nicht beschriebene Formen der improvisierten Krisenkommunikation erstmals präziser zu beschreiben.

In allen drei Fällen hat man es – nur darauf kommt es im Weiteren an – mit Forschungen zu tun, in welchen Medizin und digitale Methoden „auf Augenhöhe“ innovativ zusammenkommen – soll heißen: so, dass auch der informatische Part der Forschung mittels einer noch nicht standardisierten Methode zum Ergebnis beiträgt. Informatische, datenwissenschaftliche und medizinische Forschungsfragen sind dabei jeweils so miteinander verknüpft, dass der Erfolg der Vorhaben nur aus der gemeinsamen, wechselseitig voneinander abhängenden Forschung resultieren kann.

Im Fall C besteht der informatische Beitrag zwar nur aus der Zulieferung von Telekommunikations-Standarddaten. Gleichwohl ist der Big Data-gestützte Zugang zu den untersuchten ausnahmsweisen Formen des Kommunizierens ein Schlüssel zur Beantwortung der (explorativ gehaltenen) Forschungsfrage. Klar ist, dass in allen drei Fällen Ethikfragen berührt sind.⁵ Jeweils stellt sich also die Frage, wie man hierzu eine forschungsethische Begutachtung organisiert.

In den Fällen A und B handelt es sich um interdisziplinäre Verbundvorhaben, so dass es ein gangbarer Weg sein kann, die Forschungsbeiträge der medizinischen und der informatischen Partner getrennt zu beschreiben, um auf dieser Basis von einer medizinischen und einer nichtmedizinischen („technisch“ kompetenten) Ethik-Kommission Voten zu jeweils einem Teil der geplanten Forschungen zu erbitten. Eine solche Doppelbegutachtung (ggf. in nachrichtlicher Kenntnis auch jeweils des fachlich anderen Antragsteils) verspricht belastbare Ergebnisse, aber die Nachteile liegen auf der Hand: Der Aufwand für die Antragsteller wie auch die Kommissionen ist hoch, verbleibende „Restfragen“ zur jeweils anderen Domäne können zusätzliche Kommunikationsschleifen erforderlich machen, die Voten kommen bei den Forschenden letztlich „additiv“ zusammen (und können sich in Details sogar widersprechen), und auch das Monitoring des Gesamtvorganges sowie die Verantwortlichkeit der beteiligten Kommissionen gegenüber den beteiligten Institutionen werden fragmentiert. Eine „ganzheitliche“ Würdigung fehlt – oder ließe sich erst durch eine Zusatzvereinbarung zwischen den beteiligten Ethikkommissionen herstellen, der zufolge eines der beiden Verfahren das „führende“ und das andere lediglich das „zuarbeitende“ ist. In der Praxis sorgen alle erforderlichen Klärungsschritte für eine Doppelbegutachtung nicht zuletzt für eine zeitliche Verlängerung der Verfahren.

In Fällen vom Typ des Falles C muss vor allem eine datenintensive Methode (mitsamt einer möglicherweise kritischen Erhebungssituation) forschungsethisch bewertet werden, der im engeren Sinne medizinische Part ist nachrangiger Art (auf den Daten, die nicht ohne Weiteres „klinische“ sind, ließen sich auch sozialwissenschaftliche oder psychologische Auswertungen durchführen). Gleichwohl wird man den Begutachtungsbedarf nicht ohne Weiteres auf Informatik-Expertise beschränken, sondern eine – sagen wir: sozialmedizinische Kompetenz beziehen wollen. Infrage käme die Einholung eines Gutachtens oder die Beiziehung einer Person mit einschlägiger Qualifikation durch eine nichtmedizinische Kommission, die den Fall bewertet. Erneut liegen gewisse Nachteile auf der Hand: Die Komplexität der eingesetzten Datenerhebung- und Auswertung – kombiniert mit einer eben doch auch medizinisch relevanten Forschungsfrage – sprengt den Rahmen eines entweder „medizinischen“ oder aber „informatisch“ zu bewertenden Sachverhalts. Drohen in den Falltypen A und B sachlich überlappende Voten, kann der Falltyp C zu „Lücken“

⁵ Dies zu vertiefen sprengt den Rahmen und ist nicht unser Thema. Sicher spielen jedoch im Fall A Risiken für die Versuchspersonen sowie Kriterien für deren Auswahl und der Erweis einer nicht offenkundigen Überflüssigkeit des Versuchs eine Rolle, im Fall B sind massive Datenschutzprobleme zu prüfen (einschließlich der Zweckbindung und des methodisch hinreichend vielversprechenden Charakters der Datenerhebung), im Fall C werden Versuchspersonen, selbst wenn sie eingewilligt haben, möglicherweise nicht verantwortbaren (weil auch nicht absehbaren) politisch bedingten Gefahren exponiert, zumal die Datenerhebung mittels kommerzieller Partner erfolgt, die Daten/Informationen zur Versuchsteilnahme weitergeben könnten.

in der Begutachtung führen: Potenziell problematische Aspekte (etwa Mängel an den Bedingungen einer informierten Einwilligung oder Sicherheitsprobleme, die aufgrund der Datenverknüpfung für die Versuchspersonen entstehen) werden erst sichtbar, wenn man tatsächlich einen von vornherein „interdisziplinären“ Blickwinkel wählt (vgl. auch Beck et al. 2023).

Ethikrichtlinien und -handreichungen

Sucht man nach orientierenden Ethikrichtlinien oder Handreichungen für die Schnittstelle von Medizin und Informatik, zeigt sich, dass das einfache Bild von zwei Disziplinen, die sich in Fragen der Forschungsethik „auf Augenhöhe“ begegnen, auch in dieser Hinsicht korrekturbedürftig ist. Während die Medizin in Deutschland seit Jahrzehnten über ein flächendeckendes Netz von für klinische Forschungsprojekte zuständigen, vernetzten Ethikkommissionen sowie eine dynamische Welt von Leitlinien und Verfahrensstandards verfügt (die ihrerseits von Ethikgremien entwickelt werden), gibt es Vergleichbares in der (bzw. für die) Informatik nicht. Das Fach besitzt in Deutschland im Grunde gar keine eigene forschungsethische Tradition der Begutachtung und Qualitätssicherung.

Es ließe sich argumentieren, dass dergleichen im Sinne einer „separaten“ Säule vielleicht auch wenig hilfreich wäre, da die Belange der Anwendungsdomäne (z. B. Medizin) forschungsethisch von so großem Gewicht sind, dass die Fachkompetenz letztlich auf dieser Seite zu liegen habe. Uns sind jedoch auch keine Vorschläge etwa der *Gesellschaft für Informatik* bekannt, die auf ein Angebot der regulären Mitwirkung der Informatik in medizinischen Ethikbegutachtungen hinauslaufen würde. Etwas anders stellt sich die Lage lediglich im Feld expliziter Bindestrich-Informatiken dar. So gibt es seit 2020 immerhin einen Mustertext zur Einwilligung in Big Data-Analysen auf pseudonymisierten Patientendaten, den die *Deutsche Medizininformatik-Initiative* erarbeitet hat (Medizininformatik-Initiative 2020).⁶

Die oben skizzierten Fälle A, B und C zeigen freilich, dass es für ein Ethikvotum mit einem standardisierten Einwilligungsdokument nicht getan ist – wie überhaupt der Datenschutz nicht die einzige Bewertungsdimension ist, derer es in komplexen Fallkonstellationen bedarf. Derzeit ist also schon organisatorisch die Begutachtung von datengetriebenen und KI-Forschungsprojekten, sofern es sich nicht um rein medizinisch motivierte Datenanalysen mittels als Methode in der Medizin bereits eingeführter digitaler Analysewerkzeuge, sondern eben um Big Data- oder KI-Entwicklung und damit (auch) informatische Forschung handelt, an Forschungseinrichtungen (Universitäten, Außeruniversitären Forschungseinrichtungen, Klinika) schwierig.⁷

⁶ Im Download-Bereich zum Mustertext finden sich zahlreiche Erläuterungen. Für das Problem der Explizierbarkeit medizinischer KI-Systeme vgl. Ursin et al. (2023).

⁷ Das zeigt sich etwa auch in den aktuellen BMBF-Verbundprojekten zu KI und Medizin. Hier wiederum ist die Lage so, dass bei den ethisch ausgerichteten Projekten kaum informatische Expertise fest in die Projektgruppen eingebunden ist: Vgl. die aktuell geförderten Projekte *FRAIM – Jenseits reiner Performanz: ein ethischer Rahmen für die Nutzung von KI in der Neuromedizin* (BMBF 2023a), *ELSA-AID – Künstliche Intelligenz in der Diagnostik: Ethische, berufliche und soziale Aspekte* (BMBF 2023b), *vALID – Klinische Entscheidungsfindung durch Künstliche Intelligenz. Ethische, rechtliche und gesellschaftliche Herausfor-*

Im Entwurf wie in der abschließenden Fassung der KI-Verordnung der Europäischen Kommission wird der Bereich der medizinischen Forschung und Versorgung als eine der Domänen eingestuft, die per se als hoch riskant zu betrachten sind (EU 2021, 2024).⁸ Folgt daraus für die Forschungsethik, dass F+E-Projekte informatischer Prägung gar nicht auf medizinisch geprägten Anwendungsfällen durchgeführt werden sollten? Diese Schlussfolgerung wäre eine weitgehende und wohl nicht nur aus informatischer Sicht kontraintuitiv.⁹

Organisatorische Ansätze zur Integration informatischer Expertise in forschungsethische Prozesse

Tatsächlich ist die Informatik freilich bislang ein Fach, das – jedenfalls in Deutschland – weder eine ingenieursethische noch eine längere forschungsethische Tradition besitzt.¹⁰ Die Ethikleitlinien der *Gesellschaft für Informatik* sind alt und richten sich in recht allgemeiner Form auf Technikfolgenfragen. Der Duktus ist der einer Berufsethik, die in einem sehr allgemeinen Sinne „Verantwortung“ einfordert. Begriffe wie „Erklärbarkeit“ oder „Verstehbarkeit“ oder auch „Transparenz“ kommen in der Diskussion um Machine Learning-Verfahren seit einigen Jahren hinzu (Gehring 2023).

Forschungsethik, wie sie seit Ende der 1970er-Jahre an Hochschulen und Forschungseinrichtungen durch Ethikkommissionen praktiziert wird, wendet daher – wo Informatikforschung Ethikvoten anfordert – transdisziplinäre Kriterien an. Betrachtet werden z. B. mögliche illegitime oder ambivalente Forschungsziele, unlautere Forschungsmethoden oder Verletzungen guter wissenschaftlicher Praxis. Unter besonderer Beobachtung steht die Forschung an und mit Menschen, insbesondere Kindern sowie anderen vulnerablen Gruppen, sowie die Forschung an Tieren und Forschungen mit personenbezogenen Daten. Forschungsethik achtet aber auch auf Fairness in der Fachkonkurrenz und hat ein Auge auf mögliches Fehlverhalten (inkorrektur Ausweis der Autorschaft, Manipulationen, Plagiate), auf Fürsorgepflichten für den wissenschaftlichen Nachwuchs sowie auf Angemessenheit und Kenntnis von Methoden (ein Musikethnologe sollte keine Teilchenbeschleunigerdaten auswerten, eine Kernphysikerin keinen psychiatrischen Heilversuch anleiten).

derungen (BMBF 2023c), *VUKIM – Verantwortungsvoller Umgang mit KI in der Medizin* (BMBF 2023d) sowie *CwiC – Coping with Certainty. Prädiktion im Zeitalter Künstlicher Intelligenz* (BMBF 2023e).

⁸ Der absolut größte Teil der Medizinprodukte wird gem. Art. 6 der Entwürfe für eine KI-Verordnung wohl als hochriskant eingestuft werden (EU 2021). Dies liegt in dem Umstand begründet, dass die Klassifizierung als Hochrisiko-KI-Systeme, die zugleich unter die Medizinprodukte-Verordnung (MDR) fallen, gem. Art. 6 der Entwürfe für eine KI-Verordnung von der Frage abhängig ist, ob die Konformitätsbewertung nach der Medizinprodukte-Verordnung (MDR) gem. Art. 52 MDR durch eine benannte Stelle durchzuführen ist. Dies ist erst ab der Risikoklasse IIa der Fall (vgl. EU 2017).

⁹ Ab der Einordnung als Hochrisiko-KI-System werden die Sicherheitsanforderungen der Art. 8–15 des Entwurfs der KI-Verordnung ausgelöst, die eine Reihe von Sicherheitsanforderungen aufstellen in Form von Datenqualitätsanforderungen, Transparenzpflichten und Anforderungen an die menschliche Aufsicht.

¹⁰ Für eine frühe Darstellung der internationalen Perspektive und der grundlegenden Probleme vgl. Floridi (2013) und Terrell (2018). Den Begriff der Computer-Ethik hat Walter Maner 1976 entwickelt, als er notierte, dass ethische Entscheidungen dann schwerer zu fällen sind, wenn Computer in der Entscheidungsfindung beteiligt sind (Maner 1996).

Historisch gehen Forschungs-Ethikkommissionen in Deutschland auf die Bereiche der klinischen Pharmaforschung und der biomedizinischen Grundlagenforschung zurück; von hier aus haben sie sich in der Medizin ganz generell etabliert.¹¹ Neben psychologischen und sozialwissenschaftlichen Ethikkommissionen haben sich namentlich an Universitäten inzwischen auch sogenannte „interdisziplinäre“ Ethikkommissionen etabliert. Rein informatische Ethikkommissionen sind uns – trotz der inzwischen stürmischen Entwicklung dieses Faches – nicht bekannt. Allerdings ist die Ethikrelevanz im Bereich der datengetriebenen Großforschung außeruniversitärer Einrichtungen erkannt und es werden Ethik-Maßnahmen etabliert. So hat im Jahr 2021 etwa das Forschungszentrum Jülich eine Ethikkommission eingerichtet, die auch datenanalytische und die Simulation umfassende Expertise umfasst, und das *Human Brain Project* scheint zwar keine Ethikkommission, aber ein Ethics Advisory Board zu besitzen (Human Brain Project 2023).¹²

Für Studien im medizinischen Bereich sieht der für Deutschland maßgebliche *Arbeitskreis medizinischer Ethikkommissionen* (AKEK) die sogenannte koordinierte Begutachtung multizentrischer Verfahren regulär vor (AKEK 2023). Ein Weg der „interdisziplinären“ Begutachtung existiert jedoch nicht, weswegen sich tatsächlich im Einzelfall Fragen der oben genannten Art auftun, also etwa die Frage einer Doppelbegutachtung stellt. Wir haben selbst in zwei Fällen eine solche koordiniert und (mit) durchgeführt: im einen Fall als zeitlich abgestimmte Befassung von zwei Kommissionen auf Basis zweier (zweiteiliger) Anträge, im anderen Fall durch Teilnahme als Gast an der Sitzung einer medizinischen Ethikkommission, die zuvor für den informatischen Part eine schriftliche Expertise zu technischen Aspekten seitens der nichtmedizinischen Ethikkommission erhalten hatte.

Der häufigste Weg, den medizinische Ethikkommissionen wählen, um der Schnittstelle zu informatischen Problemstellungen gerecht zu werden, dürfte jedoch nicht die Befassung einer zweiten Kommission, sondern die Integration von Personen mit medizinischer oder sogar „KI“-Expertise in eine medizinische Ethikkommission sein. Wir greifen als Beispiel die Ethikkommission des Fachbereichs Medizin der Philipps-Universität Marburg heraus, die ein (im konkreten Fall: humanbiologisch ausgebildetes) Mitglied für „KI-Technologie“ kooptiert hat; der Bereich „KI-Technologie“ wird hier ähnlich wie die Felder „Pflege“ oder aber „Recht“ gleichsam als Nachbargesamt aufgefasst und integriert.¹³ Eine informatische bzw. technische Ethikkommission besitzt die Philipps-Universität (wie die meisten anderen Universitäten) nicht. Psychologische oder sozialwissenschaftliche bzw. erziehungswissenschaftliche Ethikkommissionen fassen Digitalfragen in der Regel noch enger als Frage nach „Methoden“, so dass hier zumeist datenanalytische Metho-

¹¹ Erste deutsche Forschungsethik-Kommissionen entstanden auf Anregung der *DFG*. Vorreiter der Entwicklung waren die USA, Großbritannien und die skandinavischen Staaten. Vgl. zur Entwicklung der Ethikberatung in der Medizin (Frewer et al. 2012; von Unger und Simon 2016; Doppelfeld und Hasford 2019; Buchner et al. 2019).

¹² Als beratende Instanz implementieren im Übrigen Forschungsnetzwerke rund um KI-Forschung Ethik-Expertise gerne in ihren Boards, so etwa auch „Europas größtes KI-Forschungskonsortium“ – das *Cyber Valley* (2023).

¹³ Vgl. die Auflistung der Mitglieder der Ethikkommission Marburg (2023).

denfachleute der Fächer, nicht aber Informatik im Sinne von „Technologie“ ins Gremium integriert ist.

Die Alternative einer explizit „interdisziplinär“ zusammengesetzten (und zuständigen) Ethikkommission wird beispielsweise an der Technischen Universität Darmstadt praktiziert. Hier sieht die Satzung u. a. zwei Mitglieder aus den Ingenieurwissenschaften verpflichtend vor, in der Praxis befindet sich unter diesen stets eine Person aus der forschenden Informatik. Digitale Methodenkompetenz aus den Sozialwissenschaften kommt (neben naturwissenschaftlicher, juristischer und philosophischer (Ethik-)Expertise) als weitere, ebenfalls obligatorische Fachlichkeit hinzu.¹⁴ Diese Institutionalisierungsform von Forschungsethik stellt auch den Hintergrund unserer eigenen Erfahrungen dar.

Die Vorteile einer von vornherein interdisziplinär zusammengesetzten Kommission gegenüber einem engeren Zuständigkeitsbereich, durch welchen Kommissionswissen im Zweifel ergänzt oder kombiniert werden muss, liegen immer dann auf der Hand, wenn Informatikforschung mit andersfachlicher, in der Kommission aber ebenfalls repräsentierter Forschung verzahnt werden muss. Allerdings ist die interdisziplinäre Kommission mit zumeist nur einem medizinisch ausgebildeten sachverständigen Mitglied dezidiert keine medizinische Ethikkommission, weswegen die TU Darmstadt medizinische Forschungsvorhaben regelmäßig an eine medizinische Ethikkommission, die dem AKEK angehört, weiterverweist. Die Integration von Informatik in Ethikbegutachtungen kann in interdisziplinären Kommissionen also gelingen. Die Problemstellung, um die es in diesem Aufsatz geht, die Integration von Informatik und Medizin in die Begutachtung von Projekten, in welchen auch die informatische Seite aktiv forscht, wird aber ebenfalls nicht gelöst.

Fassen wir die Möglichkeiten zusammen, so lauten sie im Falle einer Verfahrensführung durch eine medizinische Forschungsethikkommission: Konsultation und/oder Doppelbegutachtung durch eine (auch) „technische“ Kommission – oder aber Erweiterung im Einzelfall durch externe Gutachtende, Gäste oder andere externe Expertise. Ebenso ist die Abgabe des Falles an eine technische oder interdisziplinäre Kommission denkbar, die dann sicher aber ihrerseits medizinethische Zuarbeiten oder „Amtshilfe“ erbitten wird.

Die Alternativen lauten ansonsten, eine dauerhafte interdisziplinäre Erweiterung der „eigenen“ medizinischen Ethikkommission vorzusehen. Die vielleicht naheliegende Forderung, die Ethikkommissions-„Landschaft“ in Deutschland solle durch eine gesonderte Säule (rein) informatischer Forschungsethikkommissionen ergänzt werden, halten wir aus verschiedenen Gründen nicht für angeraten. Erstens fehlt der Informatik hierzu die (der klassischen Ingenieursverantwortung vergleichbare) Tradition mit den dazugehörigen (etwa „Prüfung“ und Haftung) umfassenden Erfahrungswerten; sie sollte von daher wohlmöglich besser im Wege der Integration in „interdisziplinäre“ Gremien von den Erfahrungen anderer Fachkulturen profitieren.

¹⁴ Um ausgehend von fachlicher Expertise in der Beurteilung kritischer Forschung zu Urteilen über die ethische Vertretbarkeit zu kommen, sind Qualifikationen förderlich. So kann ein geteiltes Wissen, z. B. um generelle ethische Problemstellungen oder datenethische Fragen, zu einer gemeinsamen Arbeitssprache beitragen und etwa auch Laien oder studentischen Mitgliedern die Mitarbeit in einer Kommission erleichtern.

Und zweitens sind eben Forschungsprojekte gerade im Bereich Big Data und KI eher selten „rein“ informatisch. Fast immer sind die Konstellationen interdisziplinär, und also Datendomänen und damit auch Forschungsfragen anderer Fächer involviert.

Zumindest abstrakt scheint es uns klar auf der Hand zu liegen, dass auch Big Data und KI-Forschung sich hier den Zumutungen zu stellen hat, die Forschenden im Bereich der Medizin wie auch des klassischen Engineering längst vertraut sind. Jedwede Kooperationsform zur forschungsethischen Begutachtung datengetriebener oder mittels KI-Verfahren durchgeführter Forschung braucht, nicht nur im Bereich von Medizin und Pharmazie, ein Set von Kriterien, durch das eine Beurteilung zum einen angeleitet wird und das zugleich der Transparenz und Verbindlichkeit von Urteilen über den Einzelfall hinaus dient. Zudem können Kriterienkataloge bei der Konsolidierung forschungsethischer Herausforderungen helfen, insofern sie dem Forschungsstand entsprechend sowohl etablierte Kriterien wie auch offene oder umstrittene Gesichtspunkte erörtern.

Zu vermerken ist, dass inzwischen beispielsweise die DFG tatsächlich verstärkt Ethik-Voten auch für informatisch geprägte Forschungsvorhaben einfordert. Dies entspricht den Vorgehensweisen auch in anderen westlichen Ländern. Allerdings wird augenscheinlich erwartet, dass Ethikkommissionen bei der DFG gestellte Anträge (und nicht eigens die detaillierteren, auf die Belange einer Forschungsethikkommission zugeschnittene Anträge) „ethisch“ prüft.¹⁵

Kriterien für die forschungsethische Begutachtung datengetriebener Projekte

Mit der stürmischen Entwicklung namentlich von allein auf hinsichtlich möglicher Zwecke noch ungerichteter Mustererkennung angelegter Big Data-Forschung sowie von KI-Experimenten stellen sich neben der Herausforderung, organisatorisch eine angemessene Begutachtungspraxis sicherzustellen, auch Fragen nach – für die Forschung geeigneten – Kriterien einer Ethikbegutachtung von anwendungsnaher (also beispielsweise Echtdaten nutzender) Big Data- und KI-Forschung. Ganz kurz versuchen wir uns auch hierzu an einem praxisnahen Lagebild.

Im Bereich „KI“ wird über die Marktzulassung von KI-Produkten öffentlich breit diskutiert. Reizworte lauten „Risikostufen“, „Diskriminierung“, „Erklärbarkeit“ etc.

¹⁵ Ein solches Verfahren würde den Unterschied ignorieren zwischen einem DFG-Förderantrag (der vor allem wissenschaftliche Zielstellungen etc. zum Gegenstand hat) und einem Antrag auf ein Ethikvotum, der nicht „Forschungen“ über Jahre hinweg charakterisiert, sondern die genaue Beschreibung eines konkreten Versuchsablaufs/experimentellen Settings – einschließlich der Software und der Datenflüsse – enthalten muss. – Forderte die DFG im Bereich der forschenden Informatik Ethikvoten zu herkömmlich geschnittenen Förderanträgen, so liefe dies darauf hinaus, nur eine sehr pauschale (Vorab-)Begutachtung von informatisch geprägten Projekten für ausreichend zu halten – oder sogar einen anderen Begutachtungspfad zu eröffnen als denjenigen, der an den wissenschaftlichen Einrichtungen vor Ort vorgesehen ist. Aus der Sicht einer Kommissionskultur, für welche die medizinische Forschungsethik die Maßstäbe gesetzt hat, wäre die geschilderte Entwicklung bedenklich. Die Informatik sollte sich aus unserer Sicht auch in der Art der Begutachtungsgrundlage an dem orientieren, was medizinische Ethikkommissionen oder auch andere Forschungsethik-Kommissionen als Standard kennen. Auch die Bewertung von KI-Forschung kann nicht mit einem gesonderten Maß gemessen werden, was die Begutachtungsgrundlagen anbetrifft.

Interessanterweise strahlt diese Diskussion jedoch nicht wirklich auf den Bereich der Forschungsethik ab.¹⁶ Ob man (und wie man) mittels KI sowie zu KI-Lösungen forschen sollte, bleibt damit bisher eine Frage, die sich nur anhand von allgemeinen forschungsethischen Kriterien beantworten lässt.

Hierbei kommt zum Tragen, dass KI-Forschung in der Begutachtung insbesondere solche Aspekte von Forschungsethik (oder auch guter wissenschaftlicher Praxis) tangiert, die man als klassische Qualitätssicherung *von Forschung* verstehen kann. Ein Beispiel ist die Vermeidung von – im Rahmen von KI-Verfahren bekanntlich schwer auszuschließenden – Biases bzw. Diskriminierungen. Zu Forschungszwecken kann ein Bias durchaus hingenommen werden, vielleicht sogar geboten sein (etwa, wenn man über Biases forscht). Umso wichtiger sind aber Gütekriterien der Forschungsarrangements selbst: Ethik muss hier Verfahren der Dokumentation von Datengewinnung, Datenpräparation, Auswertungsdurchläufen, Speicherung und auch der Minimierung unnötiger Forschung an kritischen (z. B. personenbeziehbaren) Daten einfordern.¹⁷

Ein Beispiel für eine Richtlinie mit Fokus auf die Qualitätssicherung wissenschaftlicher Forschungsprozesse vor dem Hintergrund datengetriebener Möglichkeiten in der Informatik oder aber unter massiver Einbeziehung nicht trivialerer algorithmischer Werkzeuge ist die seit 2022 vorliegende *Handreichung zur forschungsethischen Begutachtung von KI-Forschungsprojekten*, die das hessische *Zentrum verantwortungsbewusste Digitalisierung* entwickelt hat (ZEVEDI 2022). Wir präsentieren die *Handreichung* hier kurz in ihren Grundzügen, weil sie das erste Dokument ist, das forschungsethische Kriterien für die Forschung an und mit KI spezifisch für die Begutachtungsprozesse von Ethikkommissionen zusammenstellt. Für Aushandlungsprozesse innerhalb einzelner Ethikkommissionen (gleich welcher der im vorigen Abschnitt diskutierten Zusammensetzungen) bietet sie eine Grundlage, trägt aber auch darüber hinaus im Sinne eines ersten Schrittes (oder Vorschlags) zu einer Konsolidierung forschungsethischer Standards für Deutschland bei.

Die *Handreichung* führt zunächst knapp in das Themenfeld der Forschungsethik ein und benennt klassische Prüfkriterien der Ethikbewertung von Forschung (also etwa das Verbot körperlicher oder psychischer Schädigung, die Einhaltung der

¹⁶ Ein Beispiel hierfür sind die von der EU im Forschungsprogramm *Horizon Europe* geförderten „Testing and Experimentation Facilities“, von denen es seit 2023 vier Stück gibt und die als Reallabore die Integration von Forschung und Marktzulassung befördern sollen. Neben TEFs zu Agri-Food, Manufacturing und Smart Cities & Communities gibt es auch eines zum Bereich Healthcare: *TEF-Health*. Das Ziel ist hier die Entwicklung von Standards für die Zertifizierung und Qualitätskontrolle vertrauenswürdiger KI-Anwendungen im Medizinbereich, so dass Marktzulassungsbedingungen, *Medical Device Regulation* und Forschung aufeinander abgestimmt werden können. Im Zentrum steht dabei einerseits die Definition von Qualitätskriterien und Referenzmetriken, um die von KI-Systemen genutzten Daten und Algorithmen bewerten zu können, und andererseits die Definition von Standards für agile Zertifizierungsprozesse (TEF-Health 2023). Forschungsethik scheint hier – wenn überhaupt, dann nur implizit – in die Überlegungen zum gegenseitigen Austausch von Expertise einzufließen.

¹⁷ Eine Verhinderung von Verzerrungen sehen auch die Entwürfe für eine KI-Verordnung vor (jeweils Art. 10 Abs. 2 lit. f in den Entwürfen von Kommission, Rat und Parlament). Übereinstimmend wollen die Entwürfe die nach dem Unionsrecht verbotenen Diskriminierungen (vgl. Erwägungsgrund 44; Art. 10 Abs. 2 lit. f) und andere sicherheitsrelevante Verzerrungen verhindern. Konkrete Kriterien und Verfahren, „wie“ die Zulässigkeit ermittelt bzw. die Verhinderung von Verzerrungen erreicht werden kann, ermangelt es jedoch bislang. So werden technische Normen hier wohl eine große Rolle spielen.

DSGVO oder der Standards guter wissenschaftlicher Praxis). Diese Prüfkriterien gelten auch für KI-Forschungsprojekte. Weil aber „Künstliche Intelligenz“ als Begutachtungsgegenstand einer Ethikbewertung mit einigen Besonderheiten aufwartet, werden zusätzlich spezifische Hinsichten für die Bewertung von KI-Forschungsprojekten aufgeführt:

- die soziale Eingriffstiefe von Forschungen im Bereich von KI
- die operativen Besonderheiten von KI als Gegenstand im Forschungsprozess (Dokumentation, Angaben zu Datenbeschaffenheit und Algorithmik, Umgang mit fehlender Reproduzierbarkeit, Nachnutzbarkeit und Forschungsdatenmanagement)
- die Verwendung proprietärer KI-Tools („Black-Boxing“) sowie die gebotenen Voraussetzungen für eine Verwendung von KI-Tools in nichtinformatischen Disziplinen
- EU-Konformität im Bereich der Produktentwicklung
- Dual Use

Die *Handreichung* wird durch einige weitere Hinweise für Antragsstellende (zu KI-Forschung als Grundlagenforschung, zu Forschungsplanung und den Zeitschienen von forschungsethischen Begutachtungsverfahren, zur Zuständigkeit von Kommissionen an den Einrichtungen der Forschenden und zur Möglichkeit der Vernetzung über das Zentrum verantwortungsbewusste Digitalisierung) abgerundet. In seinem Kern will das Dokument die Kommissionsarbeit durch das Aufzeigen von 23 klar unterschiedenen KI-bezogenen Aspekten, geordnet nach den genannten fünf Hinsichten, unterstützen. Für diese werden jeweils spezifische Kriterien dafür angeführt, wann ein Bescheid Nachbesserungen von Seiten der Antragstellenden fordern sollte.¹⁸

Wir gehen hier in illustrierender Absicht nur auf die erste und auf die letzte der oben genannten Dimensionen kurz näher ein, um die Funktionsweise und die Spannweite der *Handreichung* zu verdeutlichen.

Mit dem Stichwort der „sozialen Eingriffstiefe“ von Forschungen im Bereich von KI wird in der *Handreichung* eine Art Gesamtbetrachtung des (insbesondere den Themenkomplex „Biases“ sowie überhaupt die Verschiebung sozialer Normen betreffenden) „impact“ von KI-Forschung versucht. Dieser hat in den letzten Jahren enorm zugenommen, weil einerseits immer mehr Trainingsdaten zur Verfügung stehen und diese mit anspruchsvoller Algorithmik und wachsender Rechenleistung verarbeitet werden können. KI-Systeme als trainierte Algorithmen können für eine breite Palette an Produkten und Anwendungen entwickelt werden oder auch nachträglich bzw. beiläufig in diese einfließen. Der Übergang von Grundlagenforschung zu Anwendungen und Produkten liegt im Bereich von KI nicht in der Verarbeitung der Daten selbst vor Augen, sondern bedarf einer soziotechnischen Analyse.

¹⁸ Der RatSWD hat die *Handreichung* inzwischen in die Zusammenstellung von generellen und spezifischen Standards für Begutachtungen der guten wissenschaftlichen Praxis aufgenommen. Siehe den Abschnitt „Methodenspezifisches“ (RatSWD 2023). Eine englische Fassung der *Handreichung* wurde im Februar 2023 ebenfalls publiziert. Sowohl die deutsche wie auch die englische Fassung der *Handreichung* finden sich zum Download auf der Webseite des Zentrums verantwortungsbewusste Digitalisierung.

Dies ist vor allem in Feldern angezeigt, bei denen bestimmte Ergebnisse der Informationsverarbeitung weitreichende Folgen für die individuelle Existenz oder das gesellschaftliche Zusammenleben haben, also z. B. im Sicherheits-, Finanz- oder eben im Medizinbereich. Die vielfach genutzte Metapher vom Lebenszyklus (*life cycle*) von KI-Systemen erlaubt es, die Spezifik des hier umrissenen Problems als eines des Übergangs von spielerischen Versuchen in geschützten Umgebungen (wie Laboren oder *sandboxes*) hin zu einem Einsatz als Expertenmodelle in auf Entscheidung drängenden Situationen zu erfassen. Für eine forschungsethische Beurteilung der sozialen Eingriffstiefe sind Kriterien wie Datenschutz, Wirksamkeit der Einwilligung, Lieferketten, Diskriminierungsfreiheit oder Nachhaltigkeit relevant.

Die letzte der fünf Hinsichten einer KI-spezifischen Forschungsethik betrifft KI und die sogenannten Dual Use-Konstellationen, die es durchaus auch im Medizinbereich gibt.¹⁹ Gemeint sind nicht nur explizit der Kriegsführung dienende Forschungsprojekte, sondern auch militärischen Szenarien benachbarte Pfade der Forschungsförderung, etwa die Katastrophenabwehr oder die Kriminalitätsbekämpfung oder auch Forschungen rund um das Thema Belastbarkeit/Stress, da hierbei auch Nutzungsoptionen für unfriedliche Zwecke entwickelt werden können. Forschungsethische Begutachtung arbeitet hier – und zwar jenseits des Risikos für Versuchspersonen – in einem hochgradig sensiblen Feld, das von manchen Einrichtungen durch Zivilklauseln vereindeutigt wird. Dass wiederum seit dem russischen Angriffskrieg auf die Ukraine die Diskussion über Zivilklauseln neu entbrannt ist (vgl. Stark-Watzinger 2023), berührt die Aufgabe, der sich Forschungsethik-Kommissionen stellen müssen, zweifellos. Die DFG hat die verantwortungsbewusste Auswahl von Kooperationspartnern unlängst ausdrücklich als Teil der „ethischen Verantwortung der Forschenden“ (DFG 2023, S. 2) bezeichnet und nicht nur die „Kommissionen für Ethik sicherheitsrelevanter Forschung“, sondern auch allgemein „Gutachtende, Fachkollegien und Gremien“ (DFG 2023, S. 4) und damit alle Begutachtungsvorgänge zur Sicherung eines guten Umgangs mit Dual Use-Risiken in die Pflicht genommen.

Auch die Frage nach einer etwaigen militärischen Nutzung und vor allem nach einer unerwünschten militärischen Nutzung von Forschungsergebnissen sowie *Knowing How* aus der Forschung stellt sich rund um KI insofern dringlich, als die Fachgemeinschaft der Informatik auf die Problemstellung einer Dual Use-Kontrolle für digitale Artefakte bisher kaum eigene Antworten besitzt. Forschungspolitische Instrumente wie klassische Exportbeschränkungen greifen im Bereich datengetriebener und algorithmischer Lösungen jedenfalls nur schlecht.

Schlussfolgerungen

Abschließend fassen wir die zentralen Punkte unserer Überlegungen zusammen. Es verwundert nicht, bedarf aber größerer Aufmerksamkeit aller Beteiligten, dass vor dem Hintergrund des rasch wachsenden Gewichts datenanalytischer Verfahren und

¹⁹ Als einschlägiges und viel diskutiertes Beispiel vgl. Urbina et al. (2022), die beschreiben, wie ein Algorithmus durch die Veränderung weniger Parameter statt möglichen Medikamenten hochtoxische Moleküle ermittelt.

von KI-Tools die forschungsethische Begutachtung im Einzelfall wie auch grundlegend vor Herausforderungen gestellt ist. Für eine umfassende Ethikbegutachtung wird die Expertise mehrerer Disziplinen benötigt, wofür fachkulturell insbesondere auf Seiten der Informatik weder hinsichtlich einschlägiger Kriterien noch organisatorisch bereits klare Voraussetzungen gegeben sind. Für die medizinische Forschung haben wir Risikobereiche dargestellt, bei denen es durch Big Data- und KI-Verfahren einen Aufwuchs an ethischem Klärungsbedarf gibt, den Ethikkommissionen in ihren Begutachtungen zusätzlich berücksichtigen müssen. Als mögliche Wege medizinische, datenwissenschaftliche und informatische Expertise in forschungsethische Begutachtungsprozesse einzubringen, haben wir die Integration von Perspektiven mittels Doppelbegutachtungen, durch eine Erweiterung medizinischer Ethikkommissionen oder durch die Etablierung von vornherein interdisziplinär zusammengesetzter Kommissionen vorgestellt. Für alle Formen zeigen wir auf, wo sie jeweils auch mit Nachteilen einhergehen. Vor diesem Hintergrund der Frage nach geeigneten Verfahren heben wir des Weiteren die Bedeutung von Kriterien hervor, die zu einer Konsolidierung des forschungsethischen Diskurses hinsichtlich der Herausforderungen durch datengetriebene Forschung beitragen können. Schließlich haben wir mit der sozialen Eingriffstiefe von KI-Forschung und der wachsenden Bedeutung von Dual Use-Konstellationen noch zwei weiterreichende Hinsichten vorgestellt, auf welche in der Arbeit von Ethikkommissionen, die KI-Forschungsprojekte begutachten, mit dem Ziel einer Herausbildung von Standards geachtet werden muss. Gerade weil die forschungsethische Begutachtung von KI-Projekten über etablierte Standards noch nicht verfügt, stellt die Auseinandersetzung mit den zuletzt genannten Aspekten in der Medizin und auch im Umfeld medizinischer Fragen im engeren Sinne eine besondere Herausforderung dar.

Funding Open Access funding enabled and organized by Projekt DEAL.

Einhaltung ethischer Richtlinien

Interessenkonflikt A. Brenneis, P. Gehring und A. Lamadé geben an, dass kein Interessenkonflikt besteht.

Ethische Standards Für diesen Beitrag wurden von den Autor/-innen keine Studien an Menschen oder Tieren durchgeführt. Für die aufgeführten Studien gelten die jeweils dort angegebenen ethischen Richtlinien.

Open Access Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

Literatur

- Acatech – Deutsche Akademie der Technikwissenschaften (2020) Machine Learning in der Medizintechnik. Analyse und Handlungsempfehlungen. München. <https://www.acatech.de/publikation/machine-learning-in-der-medizintechnik/>. Zugegriffen: 26. Nov. 2023
- Arbeitskreis Medizinischer Ethik-Kommissionen (AKEK) (2023) Sonstige Studien. <https://www.akek.de/sonstige-studien/>. Zugegriffen: 26. Nov. 2023
- Beck S, Faber M, Gerndt S (2023) Rechtliche Aspekte des Einsatzes von KI und Robotik in Medizin und Pflege. *Ethik Med* 35(2):247–263
- Buchner B, Hase F, Borchers D, Pigeot I (2019) Aufgaben, Regularien und Arbeitsweise von Ethikkommissionen. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* 62:690–696
- Budde K (2020) KI in der Medizin und Pflege aus der Perspektive Betroffener. Tagungsbericht zum Runden Tisch mit Patientenvertretungen aus der Plattform Lernende Systeme. https://www.plattform-lernende-systeme.de/files/Downloads/Publikationen/AG6_Whitepaper_Medizin_Pflege_Tagungsbericht.pdf. Zugegriffen: 26. Nov. 2023
- Bundesministerium für Bildung und Forschung (BMBF) (2023a) FRAIM – Jenseits reiner Performanz: ein ethischer Rahmen für die Nutzung von KI in der Neuromedizin. <https://www.gesundheitsforschung-bmbf.de/de/fracim-jenseits-reiner-performanz-ein-ethischer-rahmen-fur-die-nutzung-von-ki-in-der-14555.php>. Zugegriffen: 26. Nov. 2023
- Bundesministerium für Bildung und Forschung (BMBF) (2023b) ELSA-AID – Künstliche Intelligenz in der Diagnostik: Ethische, berufliche und soziale Aspekte. <https://www.gesundheitsforschung-bmbf.de/de/elsa-aid-kunstliche-intelligenz-in-der-diagnostik-ethische-berufliche-und-soziale-aspekte-11030.php>. Zugegriffen: 26. Nov. 2023
- Bundesministerium für Bildung und Forschung (BMBF) (2023c) vALID – Klinische Entscheidungsfindung durch Künstliche Intelligenz. Ethische, rechtliche und gesellschaftliche Herausforderungen. <https://www.gesundheitsforschung-bmbf.de/de/valid-klinische-entscheidungsfindung-durch-kunstliche-intelligenz-ethische-rechtliche-und-10430.php>. Zugegriffen: 26. Nov. 2023
- Bundesministerium für Bildung und Forschung (BMBF) (2023d) VUKIM – Verantwortungsvoller Umgang mit KI in der Medizin. <https://www.gesundheitsforschung-bmbf.de/de/vukim-verantwortungsvoller-umgang-mit-ki-in-der-medizin-14560.php>. Zugegriffen: 26. Nov. 2023
- Bundesministerium für Bildung und Forschung (BMBF) (2023e) CwiC – Coping with Certainty. Prädiktion im Zeitalter Künstlicher Intelligenz. <https://www.gesundheitsforschung-bmbf.de/de/cwic-coping-with-certainty-pradiktion-im-zeitalter-kunstlicher-intelligenz-11191.php>. Zugegriffen: 26. Nov. 2023
- Burrell J (2016) How the machine ‘thinks’: Understanding opacity in machine learning algorithms. *Big Data Soc*: 1–12
- Caliebe A, Scherag A, Strech D, Mansmann U (2019) Wissenschaftliche und ethische Bewertung von Projekten in der datengetriebenen Medizin. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* 62:765–772. <https://doi.org/10.1007/s00103-019-02958-2>
- Challen R, Denny J, Pitt M, Gompels L, Edwards T, Tsaneva-Atanasova K (2019) Artificial intelligence, bias and clinical safety. *BMJ Qual Saf* 28(3):231–237
- Cyber Valley (2023) Public advisory board. <https://www.cyber-valley.de/public-advisory-board>. Zugegriffen: 26. Nov. 2023
- Datenethikkommission (2019) Gutachten der Datenethikkommission der Bundesregierung. <https://www.bmi.bund.de/SharedDocs/downloads/DE/publikationen/themen/it-digitalpolitik/gutachten-datenethikkommission.html>. Zugegriffen: 17. Febr. 2024
- Deutsche Forschungsgemeinschaft (DFG) (2023) Umgang mit Risiken in internationalen Kooperationen. Empfehlungen der Deutschen Forschungsgemeinschaft (DFG), September 2023. https://www.dfg.de/download/pdf/dfg_im_profil/geschaeftsstelle/publikationen/stellungnahmen_papiere/2023/risiken_int_kooperationen_de.pdf. Zugegriffen: 26. Nov. 2023
- Deutscher Ethikrat (2017) Stellungnahme: Big Data und Gesundheit: Datensouveränität als informationelle Freiheitsgestaltung. <https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-big-data-und-gesundheit.pdf>. Zugegriffen: 26. Nov. 2023
- Deutscher Ethikrat (2023) Stellungnahme: Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz. <https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-mensch-und-maschine.pdf>. Zugegriffen: 26. Nov. 2023

- Doppelfeld E, Hasford J (2019) Medizinische Ethikkommissionen in der Bundesrepublik Deutschland: Entstehung und Einbindung in die medizinische Forschung. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* 62:682–689
- Ethikkommission Marburg (2023) Ethikkommission. <https://www.uni-marburg.de/de/fb20/fachbereich/gr/emien/ethik/ek-marburg>. Zugegriffen: 26. Okt. 2023
- EU (2017) Medical Device Regulation (MDR). <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32017R0745>. Zugegriffen: 26. Nov. 2023
- EU (2021) Proposal for Artificial Intelligence Act. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206>. Zugegriffen: 26. Nov. 2023
- EU (2024) Artificial Intelligence Act. https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf. Zugegriffen: 16. Apr. 2024
- Floridi L (2013) *The ethics of information*. Oxford University Press, Oxford
- Frewer A, Bruns F, May A (Hrsg) (2012) *Ethikberatung in der Medizin*. Springer, Heidelberg, Berlin
- Gehring P (2023) Transparenz, Erklärbarkeit, Interpretierbarkeit, Vertrauen: Digitalethische Doppelgänger des Verantwortungsbegriffs. *Dtsch Z Philos* 71(5):629–645
- Human Brain Project (2023) *Ethics and society*. <https://www.humanbrainproject.eu/en/science-development/ethics-and-society/>. Zugegriffen: 26. Nov. 2023
- Katzenmeier C (2021) KI in der Medizin – Haftungsfragen. *MedR* 39(10):859–867
- Koeppel D (2020) Towards guidelines for medical professionals to ensure cybersecurity in digital health. In: Christen M et al (Hrsg) *The ethics of cybersecurity*. The International Library of Ethics, Law and Technology 21. Springer, Cham, S 331–345
- Latour B (1994) Der Biologe als wilder Kapitalist. *Karrierestrategien im internationalen Wissenschaftsbetrieb*. *Lettre Int* 27(4):77–83
- Latour B (1996) Portrait eines Biologen als wilder Kapitalist. In: Latour B (Hrsg) *Der Berliner Schlüssel. Erkundungen eines Liebhabers des Wissens*. Akademie, Berlin, S 113–144
- Maner W (1996) Unique ethical problems in information technology. *Sci Eng Ethics* 2:137–154
- Medizininformatik-Initiative (2020) Mustertext zur Patienteneinwilligung. <https://www.medizininformatik-k-initiative.de/de/mustertext-zur-patienteneinwilligung>. Zugegriffen: 26. Nov. 2023
- Plattform Lernende Systeme (2019) *Lernende Systeme im Gesundheitswesen – Bericht der Arbeitsgruppe Gesundheit, Medizintechnik, Pflege*. München. https://www.plattform-lernende-systeme.de/files/Downloads/Publikationen/AG6_Lernende_Systeme_im_Gesundheitswesen_web_final.pdf. Zugegriffen: 26. Nov. 2023
- RatSWD (2023) Standards zur Begutachtung. Generelle und spezifische Standards für Begutachtungen der guten wissenschaftlichen Praxis. <https://www.konsortswd.de/themen/best-practice-forschungsethik/ethikkommissionen/standards/>. Zugegriffen: 26. Nov. 2023
- Stark-Watzinger B (2023) Wir müssen unsere Forschung besser vor China schützen. In: FAZ vom 20.8.2023. <https://www.faz.net/aktuell/politik/inland/stark-watzinger-wir-muessen-unsere-forschung-g-vor-china-schuetzen-19116350.html>. Zugegriffen: 26. Nov. 2023
- Strech D (2018) Normative Governance der Big Data Forschung. *Forschung: Polit Strateg Manag* 11:53–60
- TEF-Health (2023) Testing and experimentation facility for health AI and robotics. <https://tefhealth.eu/home>. Zugegriffen: 14. Apr. 2024
- Terrell B (2018ed) Computer and information ethics. In: Zalta E (Hrsg) *The Stanford encyclopedia of philosophy*. Metaphysics Research Lab, Stanford University, Stanford (<https://plato.stanford.edu/archives/sum2018/entries/ethics-computer/>). Zugegriffen: 26. Nov. 2023
- von Unger H, Simon D (2016) Wissensregulierung durch Ethikkommissionen? In: Simon D et al (Hrsg) *Handbuch Wissenschaftspolitik*. Springer, Wiesbaden, S 395–410
- Urbina F, Lentzos F, Invernizzi C, Ekins S (2022) Dual use of artificial-intelligence-powered drug discovery. *Nat Mach Intell* 4:189–191
- Ursin F, Lindner F, Ropinski T, Salloch S, Timmermann C (2023) Levels of explicability for medical artificial intelligence: What do we normatively need and what can we technically reach? *Ethik Med* 35:173–199
- World Health Organization (WHO) (2021) *Ethics and governance of artificial intelligence for health*. Geneva. <https://www.who.int/publications/i/item/9789240029200>. Zugegriffen: 26. Nov. 2023
- Zentrale Ethikkommission bei der Bundesärztekammer (ZEKO) (2021) *Stellungnahme: Entscheidungsunterstützung ärztlicher Tätigkeit durch Künstliche Intelligenz*. *Deutsches Ärzteblatt* 118: A1–A13. <https://www.zentrale-ethikkommission.de/stellungnahmen/kuenstliche-intelligenz-2021>. Zugegriffen: 26. Nov. 2023
- Zentrum verantwortungsbewusste Digitalisierung (ZEVEDI) (2022) *Zur forschungsethischen Begutachtung von KI-Forschungsprojekten*. Handreichung zur Unterstützung der Arbeit von Ethikkommissionen

nen an Hochschulen. Darmstadt. https://zevedi.de/wp-content/uploads/2022/11/ZEVEDI_Handreichung-KI-Forschungsethik_2022.pdf. Zugegriffen: 26. Nov. 2023

Hinweis des Verlags Der Verlag bleibt in Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutsadressen neutral.