

Abschlussbericht

01GP2114A

Vorhabensbezeichnung: VUKIM - Verantwortungsvoller Umgang mit KI in der Medizin –
TP-B Bioethische Aspekte und Koordination

Zuwendungsempfänger: Universität Bremen

Verfasser: PD Dr. Martin Hähnel (<https://orcid.org/0000-0002-3692-758>)
Dr. Kathi Beier (<https://orcid.org/0000-0003-1833-452X>)

Laufzeit des Vorhabens: 01.12.2021 - 28.02.2025

I. KURZE DARSTELLUNG

1. Aufgabenstellung

Die Digitalisierung wird in absehbarer Zeit weite Teile der Gesundheitsforschung und Gesundheitsversorgung nachhaltig verändern. Eine wesentliche Voraussetzung für eine sachgerechte Information und reflektierte Auseinandersetzung der Gesellschaft mit der Digitalisierung, Big Data-Anwendungen, Künstlicher Intelligenz (im Folgenden: KI) und der Medizininformatik in Forschung und Versorgung ist die sorgfältige Analyse von Chancen und Risiken für den weiteren Umgang.

Das Verbundprojekt VUKIM hat als Teil der BMBF-Fördermaßnahme „Forschung zu ethischen, rechtlichen und sozialen Aspekten (ELSA) der Digitalisierung, von Big Data und Künstlicher Intelligenz in der Gesundheitsforschung und -versorgung“ diese sorgfältige Analyse vorgenommen. Ziel der BMBF-Maßnahme ist es, die Auswirkungen der neuen Technologien auf Wissenschaft und Gesellschaft zu untersuchen und auf einen gesellschaftlich akzeptierten und verantworteten Rahmen für ihren Einsatz hinzuwirken. Spezifisches Ziel des Verbundes VUKIM war es, den verantwortungsvollen Umgang mit künstlicher Intelligenz in der Medizin aus verschiedenen Perspektiven zu untersuchen. Dazu wurden potentielle Risiken von zukünftigen KI-Anwendungen und die geltende Regulierung kritisch analysiert. Zudem wurden Beschäftigte im Gesundheitswesen und Laien eingehend zum Einsatz von KI befragt und Experten-Workshops durchgeführt.

2. Voraussetzungen, unter denen das Vorhaben durchgeführt wurde

Das Projekt war angesiedelt an der Professur für Angewandte Philosophie der Universität Bremen. Dementsprechend standen für die Durchführung des Projekts alle Ressourcen der Professur, einschließlich des Sekretariats zur Verfügung. Die beiden Mitarbeiter bekamen seitens der Universität Bremen Räumlichkeiten und Arbeitsgeräte zur Verfügung gestellt.

3. Planung und Ablauf des Vorhabens

Planung und Ablauf des Projekts entsprachen weitestgehend der Planung, wie sie im Projektantrag beschrieben worden waren. Einige Zeit nach der Bewilligung des Projekts, im März 2022, wurde der Koordinator und TP-Leiter Dr. Martin Hähnel eingestellt. Im Oktober des Jahres wurde Frau Dr. Kathi Beier als Projektmitarbeiterin eingestellt. Die Arbeit begann mit einer intensiven Sichtung der aktuellen Forschungsliteratur zur Medizinethik der KI sowie mit der Planung der verschiedenen Veranstaltungen, insbesondere mit der konkreten Vorbereitung für die Tagung *Ethische Theorien für KI / Ethical Theories for AI* im November 2022. Der weitere Ablauf des dreijährigen Projekts war sodann zum einen geprägt von der Konzipierung und Organisation von weiteren Tagungen und Workshops – allen voran die internationale Tagung zum Thema *Vertrauen und Opazität in Bezug auf KI* an der TU Dresden – und zum anderen von der Forschungs- und Publikationstätigkeit im Bremer Projektteam. Dabei waren beide Stränge der Projektarbeit eng miteinander verknüpft: Die ausführlichen Diskussionen während der Veranstaltungen sowie der Austausch mit den eingeladenen Experten und Expertinnen gab der Bremer Forschungstätigkeit einerseits maßgebliche Impulse, andererseits stellten die Mitglieder des Projektteams ihre eigenen Forschungsansätze dabei wiederholt zur Diskussion und konnten sie so kontinuierlich weiterentwickeln.

4. Wissenschaftlicher und technischer Stand, an den angeknüpft wurde

Wir befinden uns in einer Phase beschleunigter wissenschaftlicher Forschung und politischen Handlungsdrucks, die eine aktuelle ethische Neubewertung des Themas KI erforderlich macht. Insbesondere in der biomedizinischen Forschung und medizinischen Versorgung entwickelt KI derzeit eine besondere Dynamik, z. B. in Form von elektronischen Patientenakten, Telemedizin oder computergestützten Expertensystemen. Trotz ihrer großen Zukunftsaussichten sind die meisten KI-Systeme mehr oder weniger realisierbar, da sie noch mit vielen Nachteilen konfrontiert sind: Neben dem enormen Bedarf an Trainingsdaten und einer sehr aufwendigen Programmierung sind die Herstellung und Wartung von KI-Systemen sehr kostspielig. Der

Betrieb dieser Systeme kann zudem sehr schwierig oder sogar unmöglich sein. Darüber hinaus ist der Einsatz von „schwacher KI“ nur für einen kontrollierbaren Anwendungsbereich ausgelegt. Daraus folgt, dass KI-Systeme, die außerhalb ihres vorgegebenen Kontexts operieren, häufig fehleranfällig sind. Parallel zu diesen technischen Schwierigkeiten gibt es verschiedene philosophische Einwände gegen den (umfassenden) Einsatz von KI. Einige Philosophen widerlegen beispielsweise die Behauptung, dass KI-Systeme in der Lage sind, ein Selbstbewusstsein zu entwickeln (Searle 2004). Aus phänomenologischer Sicht wird sowohl die „schwache“ als auch die „starke KI“ kritisiert, da ihre Vertreter leugnen, dass KI als etwas angesehen werden kann, das ohne materielle Verkörperung existieren kann (Dreyfus 2000). Eine umfassende philosophisch-ethische Expertise zum Thema KI steht jedoch noch aus.

Obwohl KI zweifellos zu besseren Entscheidungen und Vorhersagen führen kann (z. B. Erkennung von Arzneimittelunverträglichkeiten, vorausschauende Wartung von Autos), ist die Auffassung, dass KI-Systeme als moralische Akteure behandelt werden können, die für ihre Handlungen verantwortlich sind, umstritten. Aus rein konzeptioneller Sicht haben frühere ethische Bewertungen von KI-Systemen ihre Erkenntnisse aus klassischen Darstellungen der normativen Ethik übernommen. In einer der wenigen aktuellen philosophischen Veröffentlichungen zu KI und Ethik fragt Paula Boddington (2017) nach dem am besten geeigneten normativen Rahmen für KI und verleiht dem deontologischen sowie dem konsequentialistischen oder tugendethischen Paradigma zusätzliche Plausibilität. Trotz dieser Versuche können sich alle Stakeholder noch immer nicht auf einen Ethikkodex für KI stützen, der Leitlinien und Empfehlungen für den angemessenen Umgang mit neuen KI-gestützten Umgebungen enthält.

5. Zusammenarbeit mit anderen Stellen

Mit anderen Stellen wurde insbesondere bei der Konzeption und Durchführung von Veranstaltungen zusammengearbeitet. Herr Prof. Dr. Philipp Kellmeyer (Freiburg/Mannheim) hat diesem Projekt wichtige Impulse von medizinischer Seite gegeben. So stellte er dem Team seinen partizipationsethischen Ansatz vor und half bei der Verfeinerung und Validierung des Fallbeispiels in der Abschlusspublikation. Die Zusammenarbeit mit Frau Maja Baumeyer (Helios Klinikum Erfurt) erwies sich ebenfalls als sehr hilfreich, um unsere eher theoretischen Reflexionen praxisnäher zu machen.

II. EINGEHENDE DARSTELLUNG

6. Verwendung der Zuwendung und des erzielten Ergebnisses im Einzelnen, mit Gegenüberstellung der vorgegebenen Ziele

Der Großteil der Zuwendung bestand in den Personalausgaben für die zwei wissenschaftlichen Mitarbeiter, die Veranstaltungen des Projekts konzipierten und organisierten sowie eigene Publikationen erarbeiteten und gemeinsame Publikationen betreuten (Position F0812). Darin bestehen auch die wesentlichen Ergebnisse des Projekts: einerseits in den Tagungen und Workshops, auf denen mit einer großen Zahl von Experten interdisziplinär die Leitfragen des Projekts verfolgt und unterschiedliche Antworten kontrovers diskutiert wurden; andererseits in den aus dem Projekt hervorgegangenen Publikationen, die in vielfältiger Weise mit verschiedenen Veranstaltungen verknüpft waren.

Die Ergebnisse des Verbundes sind in eine Abschlusspublikation eingegangen, die in Kürze erscheinen wird. In diesem Leitfaden wurden potentielle Risiken von zukünftigen KI-Anwendungen und die geltende Regulierung kritisch analysiert. Zudem wurden Beschäftigte im Gesundheitswesen und Laien eingehend zum Einsatz von KI befragt und Experten-Workshops durchgeführt. Antworten auf folgende Fragen wurden zusammengetragen: Wie verändert KI unser (Arbeits-)Leben? Können KI-basierte Technologien unser Wertegefühl beeinflussen? Wie gehen Beschäftigte in Gesundheitsberufen mit KI-basierten Empfehlungsvorschlägen um, die der eigenen klinischen Erfahrung widersprechen? Wie kann KI aufgebaut sein, damit ein verantwortungsvoller Umgang mit ihr sichergestellt und vereinfacht wird? Sie analysieren bzw. bewerten diese Fragen nach den Methoden und Forschungszugängen der jeweiligen Disziplin.

Das Bremer Teilprojekt, vertreten durch PD Dr. Martin Hähnel und Dr. Kathi Beier, erfasste aus ethisch-normativer Perspektive die verschiedenen Anwendungsmöglichkeiten von Künstlicher Intelligenz im Gesundheitsbereich und versuchte, sie zu bewerten. Vorrangiges Ziel dabei war es, im Rahmen des Leitfadens Grundzüge für eine kontextsensitive und in normativer Hinsicht multikriterial zu gewichtende KI-Medizinethik zu entwickeln.

Der Leitfaden bietet in erster Linie eine interaktive Synopse und transformative Integration der Forschungsergebnisse der verschiedenen Teilprojekte, insofern anhand des fiktiven, aber nahe an der klinischen Realität angesiedelten Falles von Herrn Müller gezeigt wird, an welchen Stellen seines Patientenweges epistemologische, medizinethische, soziale und rechtliche Aspekte zur Sprache kommen, die einer detaillierten Problemanalyse und normativen Bewertung zugeführt werden müssen. Am Ende jeder Station seiner Patientenreise, die Herr Müller in der berechtigten Hoffnung zurücklegt, mit neuen KI-Methoden schneller und nachhaltiger eine Heilung seiner Krankheit zu erfahren, werden die wichtigsten Befunde und Ergebnisse zusammengefasst und in Form von „Key take-aways“ als Handlungsempfehlungen formuliert.

Damit hat der Verbund versucht, eine Gratwanderung zu meistern: Auf der einen Seite sollten die fachlichen Perspektiven und Erkenntnisse zur Geltung kommen, auf der anderen Seite die jeweiligen Ergebnisse auch für die anderen Disziplinen anschlussfähig bleiben und darüber hinaus ein breiteres wissenschaftliches Publikum ansprechen. Dieser Balanceakt zwischen Tiefgang und Verständlichkeit ist generell eine zentrale Herausforderung interdisziplinärer Zusammenarbeit, insbesondere bei einem Thema, das sowohl hochkomplex als auch gesellschaftlich relevant ist. Gerade die immense Komplexität und Dynamik des Forschungsfeldes selbst war und ist ein Umstand, der dazu geführt hat, dass permanent neue Forschungsfragen aufgeworfen werden und das eigene Verständnis kontinuierlich hinterfragt und weiterentwickelt werden musste. KI in der Medizin ist ein dynamisches, sich rasant entwickelndes Feld, in dem die Halbwertszeit von Wissen oft erschreckend kurz ist und in dem wichtige Impulse aus ganz unterschiedlichen Richtungen und Disziplinen kommen können. Die Herausforderungen der Technik, der sozialen und ethischen Implikationen, der rechtlichen Rahmenbedingungen und der praktischen Anwendbarkeit sind so vielschichtig, dass sie von einer einzelnen Disziplin allein nicht durchdrungen werden können.

Im Rahmen der Evaluation des medizinethischen Teilprojekts durch die anderen Disziplinen – Erkenntnistheorie, Soziologie und Jura – lassen sich mehrere Schwerpunkte und Ergebnisse herausarbeiten.

- (1) Die erkenntnistheoretische Diskussion konzentrierte sich stark auf die Frage, ob man KI-Systemen „vertrauen“ könne. Während die Erkenntnistheorie dies für möglich hält und sich dabei auf die Vielschichtigkeit des Vertrauensbegriffs stützt, beharrt die Medizinethik darauf, dass Vertrauen genuin interpersonal sei und die Rede von Vertrauen in KI daher einem Kategorienfehler gleichkomme. Dieser Dissens beruht nicht auf einer unterschiedlichen Bewertung technischer Zuverlässigkeit, sondern auf divergierenden Begriffsdeutungen. Trotz des Widerspruchs erkennen beide Seiten die jeweilige Position als legitime Perspektive an, die weiter untersucht werden sollte. Für das Bremer medizinethische Teilprojekt ergibt sich daraus die Notwendigkeit, zwischen „Vertrauen“, „Verlassen“ und „Zuversicht“ klar zu unterscheiden und ihre jeweilige Relevanz im klinischen Alltag zu bestimmen.
- (2) Aus soziologischer Sicht wurde insbesondere die Stärke des Projekts hervorgehoben, eine tiefe Reflexion über Autonomie, Gerechtigkeit und Verantwortung zu leisten. Besonders positiv bewertet wurde der Ansatz des „embedded ethics“, also einer engen Verzahnung ethischer Reflexion mit Entwicklungs- und Implementationsprozessen. Gleichzeitig wurde betont, dass Technikentwicklung frühzeitig an reale Nutzungskontexte rückgekoppelt werden müsse. Dabei stehen nicht nur partizipative Ansätze im Vordergrund, sondern auch

die Frage, welche Formen ethischer Reflexion unter den Bedingungen knapper Ressourcen und hoher Arbeitsdichte in der Klinik überhaupt praktikabel sind. Es bleibt offen, wer die Verantwortung für die Übersetzung ethischer Prinzipien in klinische Routinen übernimmt. Deutlich wurde auch, dass sich durch KI die Entscheidungs- und Verantwortungsverhältnisse zwischen Patient:innen und Ärzt:innen verschieben. Die informierte Einwilligung wird so zum Prüfstein: Einerseits kann KI durch „objektivierte“ Daten die Transparenz erhöhen, andererseits erschwert ihre Komplexität die Nachvollziehbarkeit. Die Soziologie plädiert deshalb für neue Aufklärungs- und Entscheidungsformate, die gemeinsam mit Kommunikationswissenschaft und Psychologie entwickelt werden sollten, um Autonomie nicht nur formell, sondern auch praktisch realisieren zu können. Für das Bremer medizinethische Teilprojekt ergibt sich daraus die Aufgabe, gemeinsam mit den Nachbardisziplinen kliniktaugliche Formate zu entwerfen, die ethische Prinzipien in den Alltag der Versorgung einbetten.

- (3) Das juristische Teilprojekt stellte heraus, dass Ethik und Recht beide auf normative Leitlinien abzielen, die jedoch in unterschiedlicher Form zur Geltung kommen. Das **das** Bremer medizinethische Teilprojekt betont die Notwendigkeit bereichsspezifischer Vorgaben, und auch im Recht zeigt sich, dass branchenspezifische Anpassungen, etwa im Rahmen der DSGVO, die Akzeptanz fördern. Zugleich wird kritisch angemerkt, dass die geplante KI-Verordnung der EU zu breit angelegt sei und für die komplexen medizinischen Kontexte eine stärkere Kontextualisierung erforderlich wäre. Besonders problematisch ist der Rückzug der KI-Haftungsrichtlinie im Februar 2025, da er die zentrale Frage der Verantwortungszuschreibung und Rückverfolgbarkeit offenlässt – genau jene Dimension, die das medizinethische Projekt als notwendig herausstellt. Hervorgehoben wurde zudem, dass viele ethische Forderungen des Bremer medizinethischen Teilprojektes bereits Eingang in die europäische Regulierung gefunden haben, etwa über die Prinzipien der HLEG-AI. Allerdings warnte das juristische Teilprojekt davor, allgemeine Moralprinzipien ungebrochen ins Recht zu übertragen: Rechtsnormen müssen zweck- und grundrechtlich fundiert sein. Für das Bremer medizinethische Teilprojekt folgt daraus, dass es enge Kooperation mit dem juristischen Projekt braucht, um gemeinsam Dokumentations- und Nachverfolgbarkeitsstandards zu entwickeln, die Haftungsfragen klären und zugleich ethischen Anforderungen genügen.

Disziplinübergreifend lassen sich also mehrere Ergebnisse festhalten. Als gemeinsamer Nenner erwies sich der Ansatz des „embedded ethics“, der eine frühe und prozessorientierte

Integration ethischer Reflexion fordert. Alle Projekte betonten zudem, dass Kontextualisierung wichtiger ist als abstrakte Regulierung: Die konkreten Bedingungen des Klinikalltags sind ausschlaggebend für die Wirksamkeit ethischer Vorgaben. Deutlich wurde auch, dass sich durch den Einsatz von KI Autonomie und Einwilligung im medizinischen Entscheidungsprozess verändern, klassische Gesprächsformate stoßen hier an Grenzen. Eng verknüpft damit ist die Frage der Transparenz und Erklärbarkeit, die nicht nur ethisch, sondern auch rechtlich von großer Bedeutung ist, insbesondere solange Haftungs- und Verantwortlichkeitsfragen ungelöst bleiben. Schließlich bleibt der Begriff des Vertrauens ein strittiger Punkt, dessen Klärung für die praktische Handhabung unverzichtbar ist.

Insgesamt bestätigen die anderen Disziplinen den Mehrwert des medizinethischen Teilprojekts, insbesondere durch den Fokus auf Autonomie und die Idee einer eingebetteten Ethik. Zugleich markieren sie zwei zentrale Baustellen: Zum einen müssen Begriffe wie Vertrauen, Einwilligung und Verantwortung so präzisiert werden, dass sie im Klinikalltag handlungsleitend wirken können; zum anderen gilt es, Nachverfolgbarkeit als Brücke zwischen ethischen Anforderungen und rechtlicher Durchsetzbarkeit zu etablieren.

PD Dr. Martin Hähnel hat während des Projektes intensiv an der Schnittstelle von Philosophie, Medizinethik und Künstlicher Intelligenz (KI) gearbeitet. Seine Arbeiten der letzten drei Jahre zeichneten sich durch eine konsequent interdisziplinäre Herangehensweise aus, die Perspektiven aus der Philosophie, der Medizin, der Informatik und den Sozialwissenschaften miteinander verband. Ein Schwerpunkt lag auf der Frage, wie ein verantwortungsvoller und praxisnaher Umgang mit KI in klinischen Kontexten gestaltet werden kann, was er etwa in einem fallbasierten Leitfaden zum Einsatz von KI in der Medizin im Rahmen des Verbundes herausgearbeitet hat. Darüber hinaus untersucht er grundlegende ethische und epistemische Probleme wie Vertrauen und Opazität in der KI, die Dual-Use-Problematik oder die Rolle der Tugendethik in digitalen Kontexten. Ein weiterer Fokus seiner Forschung lag auf den sozialen und organisationalen Implikationen von KI, insbesondere in Gesundheitseinrichtungen, wo er Verschiebungen von Macht und Verantwortung analysiert. Neben diesen angewandten Fragestellungen widmet er sich auch philosophischen Grundlagenfragen, etwa nach dem Wesen, den Möglichkeiten und der normativen Bewertung von Künstlicher Intelligenz.

Wissenschaftliche Vorträge

- | | |
|------------|---|
| 18.09.2025 | Verantwortungsvoller Umgang mit KI in der Medizin, Philosophische Gesellschaft Bremerhaven |
| 09.04.2025 | Responsible Use of AI in Medicine – An Interdisciplinary Inquiry (Symposium Responsible AI: Promises, Pitfalls, and Practices, München) |

- 21.02.2025 Die Rückkehr teleonomischen Denkens in der aktuellen KI-Ethik-Diskussion (Tagung Künstliche Intelligenz und nicht-menschliche Entitäten: Synergien aus unterschiedlichen philosophischen Debatten, Universität Heidelberg)
- 20.06.2024 A new age of dual-use technologies? Military medical ethics as a touchstone for evaluating AI-induced risks and opportunities (Military Medical Ethics Workshop, Genf)
- 02.05.2024 Janusgesichtige Chatbots? Dual-Use-Bedenken bei generativer KI und großen Sprachmodellen (Probenvortrag W3 Ethik der Digitalisierung, KU Eichstätt-Ingolstadt)
- 08.01.2024 Das Janusgesicht der Künstlichen Intelligenz – Medizin- und medienethische Überlegungen zum Dual-Use-Potential digitaler Technologien (Probenvortrag im Rahmen der Besetzung der W3-Professur für Angewandte Ethik an der Hochschule für Philosophie in München)
- 30.11.2023 Dual-use of AI-generated content – A question of trustworthiness? (Tagung „SAS23: Reliability or Trustworthiness?“, Universität Stuttgart)
- 20.10.2023 Dual-Use und AI (Philosophisches Kolloquium, Universität Bonn)
- 02.10.2023 Verantwortungsvoller Umgang mit Künstlicher Intelligenz in der Medizin – Ethische Fallbeispiele und -stricke (Probenvortrag Professur für Health Care Ethics, Universität Graz)
- 28.09.2023 Das AI Alignment-Problem und der moralischer Exemplarismus (Tagung für Praktische Philosophie Salzburg)
- 03.08.2023 Writing research proposals with chatGPT: new Opportunities, unknown risks (Workshop “Machine Discovery and Creation”, Leibniz Universität Hannover)
- 15.06.2023 What is AI readiness? (BMBF-Projekt PROKIP, Universität Bremen)
- 16.03.2023 AI Alignment and Moral Exemplarism (Value alignment workshop, KNAW Amsterdam)
- 17.11.2022 Ethical Theories for AI – Introductory Comments (Workshop “Ethical Theories for AI”, Universität Bremen)
- 20.10.2022 What is AI Ethics? (Probenvortrag W2 Ethische, soziale und rechtliche Aspekte der KI, Universität Lübeck)
- 10.10.2022 Building Algorithms Around Beauchamp and Childress` Principles: Scopes and Challenges (Beauchamp and Childress-Kolloquium, Universität Tübingen)

- 22.09.2022 On the Digital Transformation and the Shift of Structural Power Relations in Work Organizations – Sociological Analysis and Ethical Evaluation at the Example of AI in Healthcare (Konferenz “AI and the Future of Work”, TU Delft)
- 13.07.2022 Verantwortungsvoller Umgang mit KI in Medizin und Gesundheitsversorgung („Künstliche Intelligenz und Datenschutz: Was heißt hier Selbstbestimmung?“, Der Landesbeauftragte für den Datenschutz und die Informationsfreiheit Baden-Württemberg, Stuttgart)
- Public Engagement*
- 29.10.2025 „Verantwortungsvoller Umgang mit KI in der Medizin“, Symposium „Zukunftsperspektiven für Krankenhäuser“
- 18.09.2025 „Verantwortungsvoller Umgang mit KI in der Medizin“, Podiumsdiskussion mit Prof. Dr. Klaus Eickel, Hochschule Bremerhaven
- 24.06.2025 "KI in der Medizin - die Zukunft?", Junior Science Café, Podiumsdiskussion am Thomaeum
- 13.07.2022 „Wer braucht noch Ärzte?“ Podiumsdiskussion mit Prof. Dr. Wilhelm Stork (Direktor FZI Forschungszentrum Informatik und Professor am KIT im Institut für Technik der Informationsverarbeitung (ITIV)) und Dr. Keno März (Deutsches Krebsforschungszentrum Heidelberg, Forschungsgruppenleiter in der „Division of Intelligent Medical Surgery“, Abteilung Intelligente medizinische Chirurgie; Gruppenleiter Intelligente Systeme in der klinischen Onkologie), Veranstaltung „Künstliche Intelligenz und Datenschutz: Was heißt hier Selbstbestimmung?“, Der Landesbeauftragte für den Datenschutz und die Informationsfreiheit Baden-Württemberg, Stuttgart)
- 30.3.2023 "Maschinen können menschliche Akteure nicht ersetzen", *Weserkurier*
- 5.11.2023 "Die Verantwortung bleibt bei den Ärzten", *Buten un binnen*, <https://www.butenunbinnen.de/videos/kuenstliche-intelligenz-medizin-ki-philosophie-verantwortung-martin-haehnel-100.html>

Dr. Kathi Beier hat ebenfalls zur Schnittstelle von Ethik und KI geforscht, insbesondere zu den theoretisch-philosophischen Begründungen und den praktisch-anwendungsbezogenen Problemen im Bereich von Medizin und Pflege. Ihr Schwerpunkt lag einerseits bei der Analyse der Stärken und Schwächen eines tugendethischen Zugangs zu den Herausforderungen durch KI. In dieser Hinsicht hat sie die Heterogenität solcher Zugänge herausgearbeitet und die These vertreten, dass eine einheitliche KI-Tugendethik nicht zu erkennen und wohl auch

nicht möglich ist. Andererseits ist sie der Frage nachgegangen, wie KI-Systeme selbst so gestaltet werden können, dass sie „ethisch“, gar „tugendhaft“ handeln. Im Zusammenhang mit dieser Frage hat sie sich vor allem mit der Rolle sozialer Roboter und empathischer Chatbots und ihrem Einsatz sowohl in Pflegeeinrichtungen als auch in privaten Kontexten auseinandergesetzt. Ergebnis dieser Forschung sind die Thesen, dass ethische Beurteilungen hier jeweils immer mit genauem Blick auf die Technologien und den Einsatzbereich solcher „artificial companions“ zu erfolgen haben und dass die Simulation sozialer Fähigkeiten durch KI-Systeme die menschlichen Bedürfnisse, für deren Befriedigung sie entwickelt werden, tendenziell unterlaufen und damit Frustration und Enttäuschung hervorrufen.

Wissenschaftliche Vorträge

- | | |
|-------------|--|
| 15.06. 2023 | Online-Impuls mit Dr. Martin Hähnel über „Ethische Grundlagen und Implikationen von AI Readiness“ im Rahmen des Projekts ProKIP (= Projektentwicklung und -begleitung von KI-Einsatz in der Pflege) an der Universität Bremen. |
| 07.09. 2023 | Vortrag/Impuls mit Prof. Dr. Dagmar Borchers zum Thema „Erkenne dich selbst: Welche ethischen Überzeugungen leiten mein Handeln und Entscheiden?“ auf dem Netzwerktreffen KI in der Pflege (Urania Berlin). |
| 19.09. 2024 | Vortrag „Great pretenders? Social robots and the virtues of truth“ auf der internationalen Tagung <i>Virtue ethics and Technology</i> am Institut für Philosophie der KU Leuven, Belgien |
| 27.02. 2025 | Täuschung und Selbsttäuschung? Wie KI-Freunde unsere Haltungen zur Wahrheit herausfordern“, Philosophische Gesellschaft Bremen |

Darüber hinaus hat Kathi Beier am 14. Mai 2025 an der Universität Bremen im Masterseminar „Ethik und Versorgungsgerechtigkeit“ von Janissa Altona am Institut für Public Health eine Sitzung übernommen und darin mit den Studierenden über „Ethik und KI in Medizin und Pflege“ gesprochen.

7. Notwendigkeit und Angemessenheit der geleisteten Arbeit

Die Beschäftigung des Koordinators und Teilprojektleiters und der wissenschaftlichen Mitarbeiterin als zentralen wissenschaftlichen Akteuren des Projekts war unabdingbar, wenn auch die Beschäftigung in einem Umfang zu je 50% für die zu leistende Arbeit recht knapp bemessen war.

Die Durchführung der Tagungen und Workshops mit den zahlreichen eingeladenen Experten waren notwendig, um vor allem den medizinischen Input für die Projektarbeit zu gewinnen und um eine intensive diskursive Auseinandersetzung mit den verschiedenen Theorien und Argumenten zu ermöglichen, um so schließlich fundierte und differenzierte Antworten auf die Forschungsfragen zu geben. Die Einladung von Gästen aus Übersee bei der Tagung im November 2023 zu Trust and Opacity in AI an der TU Dresden war angemessen, um das Projekt international zu vernetzen und die anvisierte Synthese bislang getrennter Diskursstränge auf hohem Niveau zu ermöglichen.

8. Voraussichtlicher Nutzen, insbesondere der Verwertbarkeit des Ergebnisses im Sinne des fortgeschriebenen Verwertungsplans

- Nicht anwendbar -

9. Der während der Durchführung des Vorhabens dem ZE bekannt gewordene Fortschritt auf dem Gebiet des Vorhabens bei anderen Stellen

Über weite Strecken des Projekts gab es im Feld der KI-Entwicklung und der damit verbundenen Forschungslage zahlreiche neue aber letztlich innerhalb des Feldes teils inkrementelle Entwicklungen. Das hat sich mit Veröffentlichung von ChatGPT und in Folge weiterer sog. Large Language Models (LLMs) dramatisch verändert. Damit hat sich nicht nur die öffentliche Wahrnehmung von KI deutlich erweitert und sensibilisiert, auch im Untersuchungsfeld Medizin vervielfachen sich damit die potenziellen Einsatzgebiete. Diese weiten die bisherige Dominanz des KI-Einsatzes in der bildgebenden Diagnostik aus und eröffnen dadurch gänzlich neue Anwendungsmöglichkeiten von KI. Das andauernde Monitoring des Forschungsstands konnte einen Teil dieser neuen Entwicklungen auffangen. Allerdings ergaben sich daraus einige neue Anforderungen, denen wir in den letzten Projektmonaten und mittels der zu beantragenden Verlängerung nachgehen wollen. Vor allem lässt sich aus ethischer Sicht eine Schwerpunktverlagerung in der normativen Bewertung feststellen, weg von Ansätzen, die rechtliche und ethische Richtlinien für KI in praktischer Unabhängigkeit zur technologischen Entwicklung entwerfen und implementieren wollen und hin zu Ansätzen, welche von Anfang das ethische Design von KI-Systemen in den Blick nehmen und versuchen daran mitzuwirken.

Unser Verbundprojekt hat sich von Anfang an stark interdisziplinär konzipiert. Der kontinuierliche Austausch der Teilprojekte stellt schon zum jetzigen Zeitpunkt einen unverzichtbaren Mehrwert für unsere Forschungsergebnisse dar. Um den Wissens- und

Ergebnistransfer der von uns im Verbund gelebten Interdisziplinarität auch der Forschungsgemeinschaft und Öffentlichkeit zugänglich zu machen haben wir uns im Verbund dazu entschlossen, unsere Ergebnisse in einer gemeinsamen und durchgängig interdisziplinär strukturierten Buchpublikation zu fassen, in die ebenfalls die neueren Erkenntnisse zu LLM noch einzuarbeiten sind und auch dafür soll die Verlängerung genutzt werden.

10. Erfolgte oder geplante Veröffentlichungen des Ergebnisses nach Nr. 6.

Aus dem Projekt sind zahlreiche Publikationen hervorgegangen, die teilweise schon erschienen sind und teilweise noch erscheinen werden. Einerseits spiegeln sie die durchgeführten Veranstaltungen wider und andererseits stellen sie die eigenen Arbeiten des Bremer Projektteams dar. Einige besonders erwähnenswerte Publikationen seien im Folgenden genannt. Im Mittelpunkt stand die Arbeit am fallbasierten Leitfaden, d.h. das Schreiben der Kommentare zu den Stationen 1 bis 5, die Diskussion aller Kommentare mit den VUKIM-Mitgliedern, das Verfassen des Querkommentars zu den Betrachtungen der anderen beteiligten Disziplinen, die Ausarbeitung der Glossar-Einträge und schließlich die Erstellung des Manuskripts, inklusive zweier Korrekturrunden. Die Abschlusspublikation *Verantwortungsvoller Umgang mit KI in der Medizin aus interdisziplinärer Perspektive: Ein fallbasierter Leitfaden* ist im Verlag Nomos erschienen (doi.org/10.5771/9783748954392). Weitere zu erwähnende Publikationen bzw. Publikationsvorhaben im Projektteam sind:

PD Dr. Martin Hähnel

2025 Trust and Opacity in Artificial Intelligence: Mapping the Discourse (mit Rico Hauswald). *Philosophy and Technology* 38, 115 (2025). <https://doi.org/10.1007/s13347-025-00947-9> (peer-reviewed)

Blackwell Companion to Applied Philosophy of AI (mit Regina Müller), Wiley: Hoboken 2025.

Ethical challenges and solutions in AI-driven medical data management: a focus on distributed machine learning. *Discover Artificial Intelligence* 5, 53 (2025). <https://doi.org/10.1007/s44163-025-00266-0> (peer-reviewed)

2024 *Was ist, was kann, was soll Künstliche Intelligenz? Ein philosophisches Gespräch* (mit Vincent C. Müller), Meiner: Hamburg 2024. <https://meiner.de/was-ist-was-kann-was-soll-ki.html?>

Conceptualizing Dual-use: A Multidimensional Approach, in: *Research Ethics* (2024), <https://doi.org/10.1177/17470161241261466>. (peer-reviewed)

Striking the Balance: Ethical Challenges and Social Implications of AI-Induced Power Shifts in Healthcare Organizations (mit Sabine Pfeiffer and Stephan Grassmann), in: *AI and Society* (2024), <https://doi.org/10.1007/s00146-024-02043-6>. (peer-reviewed)

Der tugendethische Beitrag (mit Kathi Beier, Dagmar Borchers, Hans-Henrik Dassow, Björn Haferkamp, Antonia Kempkens, Regina Müller), in: Grimm, P. / Trost, K. E. / Zöllner, O. (Hg.), *Digitale Ethik. Handbuch für Wissenschaft und Praxis*, Baden-Baden: Nomos 2024, 49-62. <https://doi.org/10.5771/9783748942399-49>.

2023 Der Patientenwille, seine Identität und die Bewertung der Absicht, ihn mittels eines KI-gestützten *recommender system* zu ersetzen, in: Fuchs, M. / Hähnel, M. / Simmermacher, D. (eds.), *Der Patientenwille und seine (Re-) Konstruktion*, Springer VS: Wiesbaden 2023, 233-248. https://doi.org/10.1007/978-3-658-40192-4_11.

Dr. Kathi Beier

2025 Virtue Ethics and AI, in: *A Blackwell Companion to Applied Philosophy of AI*, ed. by Regina Müller, Martin Hähnel, Hoboken: Wiley-Blackwell 2025, 178-191.

2024 Der tugendethische Beitrag (mit Dagmar Borchers, Hans-Henrik Dassow, Martin Hähnel, Björn Haferkamp, Antonia Kempkens, Regina Müller), in: Grimm, P. / Trost, K. E. / Zöllner, O. (Hg.), *Digitale Ethik. Handbuch für Wissenschaft und Praxis*, Baden-Baden: Nomos 2024, 49-62. <https://doi.org/10.5771/9783748942399-49>.