COMMUNITY PAPER

VERTRAUENSWÜRDIGE KI IM GESUNDHEITSWESEN: EIN ZWEISCHNEIDIGES SCHWERT

Ein Policy Brief der Arbeitsgruppe Künstliche Intelligenz im Gesundheitswesen des Global Health Hub Germany

Hintergrund dieses Papiers

Dieser Policy Brief enthält eine Reihe aktueller, strategischer Empfehlungen zu zentralen Governance-Maßnahmen für Deutschland, um den vertrauenswürdigen Einsatz künstlicher Intelligenz (KI) im Gesundheitswesen sicherzustellen, nun da der EU AI Act in Kraft getreten ist. Der Policy Brief basiert auf der Expertise der Arbeitsgruppe 2025/2026 des Global Health Hub Germany (GHHG) zum Thema Künstliche Intelligenz in der globalen Gesundheit. Er wird durch evidenzbasierte Erkenntnisse und bewährte Verfahren aus der akademischen Forschung, Praxisberichten, regulatorischer Analyse und Konsultationen mit verschiedenen Ansprechpartnern gestützt, um strategische Entscheidungsfindung zu informieren und einen inklusiven, verantwortungsvollen und wirksamen Einsatz von künstlicher Intelligenz im Gesundheitswesen zu fördern.

Über die Autoren und Autorinnen
Verfasst von: Allison Colbert, Ertila Druga, Jana Fehr, Rachel Firestone.
Weitere Beiträge stammen von: Michael Bayerlein, Vladimir Choi, Felix Holl, Zahra Karimian, Felipe Mejia Medina, Meenu Singh, Kemna Solveig, Elias Staatz, Franziska Laporte Uribe, Sonali Wayal. Die Arbeitsgruppe wurde gemeinsam von den Hub-Communities Global Mental Health und Global Women's Health ausgerichtet. Die Unterstützung durch das Sekretariat des Global Health Hub Germany erfolgte durch Katrin Würfel und Anya Abanto Graffman.

Zusammenfassung

Künstliche Intelligenz (KI) verändert das Gesundheitswesen und die Gesundheitsversorgung rasant – von unterstützter Diagnostik über Monitoring und Gesundheitsinformationssysteme bis hin zur Arzneimittelforschung und Patientenbindung. Ihre Fähigkeit, große Datenmengen zu verarbeiten und Muster darin zu erkennen, eröffnet erhebliche Chancen.

Gleichzeitig bringt der Einsatz von KI erhebliche Risiken mit sich. Verzerrte oder intransparente Entscheidungen, algorithmische Diskriminierung und eine übermäßige Abhängigkeit von automatisierten Systemen können in klinischen und gesundheitspolitischen Kontexten direkten Schaden verursachen. **KI im Gesundheitswesen ist daher ein zweischneidiges Schwert**: Sie kann Zugang und Behandlungsergebnisse verbessern, aber auch bestehende Ungleichheiten verstärken, wenn Governance, Aufsicht und Rechenschaftspflicht unzureichend sind.

Der EU AI Act, an den Deutschland gebunden ist, schafft erstmals einen umfassenden Rechtsrahmen für KI. Er legt fest, unter welchen Bedingungen KI-Systeme rechtlich zulässig sind. Rechtliche Compliance allein garantiert jedoch noch keine vertrauenswürdige Umsetzung. Insbesondere stellt sie nicht automatisch sicher, dass KI-Systeme über unterschiedliche Bevölkerungsgruppen hinweg gerecht funktionieren oder dass Patientinnen und Patienten verständlich darüber informiert werden, wenn KI an ihrer Versorgung beteiligt ist.

Dieser Policy Brief analysiert zwei bereits eingesetzte KI-Anwendungen in klinischen Kontexten in Ländern mit hohem Einkommen: KI im Mammographie-Screening in Deutschland und KI-gestützte Triage im Bereich psychische Gesundheit im Vereinigten Königreich. Die Fallstudien zeigen, dass auch in stark regulierten Gesundheitssystemen Lücken zwischen rechtlicher Zulässigkeit und vertrauenswürdiger Umsetzung bestehen. Daraus leitet der Brief Empfehlungen ab, wie KI im Gesundheitswesen bewertet, reguliert und an konkrete Versorgungskontexte angepasst werden sollte.

Die Empfehlungen richten sich insbesondere an die Umsetzung des EU AI Act im deutschen Gesundheitswesen und an weitere gesundheitspolitische Kontexte. Im Mittelpunkt stehen Transparenz, Rückverfolgbarkeit, Evaluation und verlässliche Strukturen zur Evidenzgewinnung.

Die Empfehlungen des Beitrags konzentrieren sich auf Transparenz, Rückverfolgbarkeit, Evaluation und Evidenzinfrastruktur und stellen drei Prioritäten in den Mittelpunkt:

1. **Patientinnen, Patienten und Behandelnde in den Mittelpunkt** der KI-gestützten Versorgung stellen
2. **Eine sinnvolle menschliche Aufsicht und öffentliche Rechenschaftspflicht sicherstellen**
3. **Adaptive Governance-Strukturen** aufbauen, die Sicherheit und Leistungsfähigkeit langfristig gewährleisten

Einleitung

Künstliche Intelligenz ist zunehmend in alltägliche Entscheidungsprozesse eingebettet: beim Verfassen von Mitteilungen, beim Empfehlen von Inhalten und bei der Unterstützung alltäglicher Entscheidungen. Selbst in diesen Kontexten mit geringem Risiko erkennen wir ihre Grenzen: Ergebnisse können unvollständig oder falsch sein.

In klinischen und gesundheitspolitischen Kontexten haben diese Risiken weitaus größere Konsequenzen. Kleine Fehler können zu übersehenen Diagnosen, unangemessener Behandlung oder zum systematischen Ausschluss bestimmter Patientengruppen führen. Gleichzeitig bieten diese Systeme erhebliches Potenzial: Sie können Routineaufgaben automatisieren, Früherkennung verbessern, den Zugang zur Versorgung erweitern und überlastete Gesundheitssysteme entlasten.

Diese Spannung macht **KI im Gesundheitswesen zu einem zweischneidigen Schwert**: zu einer Technologie, die Gesundheitsergebnisse sowohl verbessern als auch untergraben kann, je nachdem, wie sie reguliert, umgesetzt und überwacht wird.

KI-Tools können den Zugang für marginalisierte Bevölkerungsgruppen erweitern, darunter Migrantinnen und Migranten, Geflüchtete und unterversorgte Gemeinschaften. Eine schwache Aufsicht kann jedoch weiterhin erhebliche Governance-Risiken schaffen – selbst in stark regulierten Systemen oder Ländern mit hohem Einkommen.

Dieser Beitrag analysiert zwei KI-Einsätze in Ländern mit hohem Einkommen, um übergreifende Governance-Herausforderungen und Umsetzungslehren zu identifizieren. Diese werden noch wichtiger, wenn KI-Tools an andere Kontexte angepasst oder in andere Kontexte übertragen werden. Die Herausforderungen sind relevant für die Umsetzung des EU AI Act im Gesundheitswesen, sowohl in Deutschland als auch im Rahmen seines internationalen Engagements. Sie machen deutlich, dass Risiken auch in stark regulierten Umgebungen fortbestehen können, wenn klare, sektorspezifische Umsetzungshinweise fehlen.

Der EU AI Act schafft eine wichtige rechtliche Grundlage. Er schreibt für Hochrisiko-Systeme Dokumentation, menschliche Aufsicht und Konformitätsbewertungen vor. Er gewährleistet jedoch nicht automatisch eine gerechte Leistung über verschiedene Bevölkerungsgruppen hinweg, unabhängige Evidenzgenerierung oder eine verständliche Information von Patientinnen und Patienten über die Beteiligung von KI an ihrer Versorgung. In beiden Fallstudien bestehen wichtige Governance-Herausforderungen fort, trotz Einhaltung regulatorischer Anforderungen.

Wenn solche Lücken in Europa bestehen bleiben, können sie auch in anderen globalen Kontexten auftreten – mit unterschiedlich stark ausgeprägter Regulierung und unterschiedlichen finanziellen Ressourcen. Deutschland kann daher nicht glaubwürdig an der internationalen Entwicklung vertrauenswürdiger Tools mitwirken oder solche Tools exportieren, ohne zugleich auf eine robuste Umsetzung im eigenen Gesundheitssystem hinzuwirken.

Zur Strukturierung der Analyse stützt sich der Beitrag auf Fallstudien aus der Frauengesundheit und der psychischen Gesundheit. Sie untersuchen den Einsatz von KI in diagnostischen Prozessen beziehungsweise in Triage-Prozessen. Die Fälle spiegeln Bereiche aktiven Engagements innerhalb der Global Health Hub Germany Community wider und verankern die Analyse in einem fortlaufenden interdisziplinären Dialog und praxisnahen Perspektiven. Als zentrale analytische Linse nutzt der Beitrag das FUTURE-AI-Framework, das im folgenden Abschnitt näher erläutert wird.

Gestützt auf die Evidenz aus den Fallstudien formuliert der Beitrag Empfehlungen zur Stärkung von Transparenz, Aufsicht, Rechenschaftspflicht und Umsetzungskapazität in der KI-Governance im Gesundheitswesen im Rahmen des EU AI Act.

Statt eine umfassende Lösung zu präsentieren, bieten die Empfehlungen ein pragmatisches Bündel von Maßnahmen, um staatliche Entscheidungsfindung mitzugestalten und zur fortlaufenden Stärkung der KI-Governance in Deutschland und in seinen globalen Gesundheitspartnerschaften beizutragen.

Analytischer Rahmen

Dieser Beitrag untersucht KI im Gesundheitswesen anhand von **drei sich ergänzenden Rahmenwerken** auf unterschiedlichen Analyseebenen.

Erstens stützt er sich auf die WHO-Leitlinie zu Ethik und Governance von KI für Gesundheit (WHO, 2021). Sie bildet die ethische Grundlage für vertrauenswürdige KI im Gesundheitswesen und betont Autonomie, Sicherheit, Transparenz, Rechenschaftspflicht, Gerechtigkeit und Nachhaltigkeit.

Zweitens nutzt der Beitrag das FUTURE-AI-Framework als zentrale analytische Linse zur Bewertung der Fallstudien (Lekadir et al., 2025). Die FUTURE-AI-Leitlinie wurde durch internationalen Expertinnen- und Expertenkonsens entwickelt. Sie übersetzt breite ethische Prinzipien in praktische Kriterien, um zu bewerten, ob KI-Systeme im Gesundheitswesen vertrauenswürdig und in realen Versorgungskontexten einsetzbar sind.

Das Framework bewertet KI-Systeme entlang sechs Dimensionen des Akronyms FUTURE:

- Fairness: gerechte Leistung über verschiedene Patientengruppen und Bevölkerungen hinweg
- Universalität: Fähigkeit zur Generalisierung über verschiedene Settings, Bevölkerungen und Versorgungsumgebungen hinweg
- Rückverfolgbarkeit: Transparenz, Dokumentation, Monitoring und Rechenschaftspflicht über den gesamten KI-Lebenszyklus hinweg
- Nutzbarkeit: sichere und wirksame Integration in klinische Arbeitsabläufe und Nutzerbedürfnisse
- Robustheit: zuverlässige Leistung unter variierenden realen Bedingungen
- Erklärbarkeit: Fähigkeit der Stakeholder, zu verstehen, wie Systeme funktionieren und Ergebnisse erzeugen

Drittens betrachtet die Analyse den EU AI Act als regulatorischen Umsetzungskontext. Während der Act wichtige rechtliche Anforderungen für Hochrisiko-KI-Systeme festlegt, konzentriert sich dieser Beitrag darauf, wie vertrauenswürdige KI-Governance im deutschen Gesundheitssystem und im Rahmen globaler Gesundheitsengagements praktisch umgesetzt werden kann.

Ein verantwortungsvoller Einsatz von KI im Gesundheitswesen entsteht nicht allein durch Technologie. Er erfordert ethische Grundlagen, operative Standards und regulatorische Umsetzung. Zahlreiche ethische Rahmenwerke haben versucht, diese Grundlage zu liefern (Jobin et al., 2019) (Beauchamp & Childress, 1979). Weitere Informationen zu diesen Rahmenwerken finden sich in Anhang I.

Fallstudien

Fallstudie I – KI im Mammographie-Screening in Deutschland

Das deutsche Mammographie-Screening-Programm[1] ist das größte in Europa. Es lädt rund 14 Millionen Frauen im Alter von 50 bis 75 Jahren alle zwei Jahre zum Screening ein.[2] **Jede Mammographie wird unabhängig von zwei Radiologinnen oder Radiologen beurteilt – dieses Verfahren wird als Doppelbefundung bezeichnet.** Wird eine Aufnahme von mindestens einer der beiden Personen als auffällig bewertet, wird der Fall nochmals gemeinsam beurteilt.[3]

In diese Doppelbefundung kann Vara AI integriert werden. Vara AI ist ein CE-zertifiziertes KI-basiertes Medizinprodukt der Klasse IIb, das in bestehende radiologische Arbeitsabläufe integriert wird. Mammographieaufnahmen werden automatisch an die Vara-Plattform weitergeleitet. Diese gibt eine Risikoklassifikation (unauffällig/auffällig) sowie visuelle Markierungen verdächtiger Bereiche direkt an die Workstation der Radiologin oder des Radiologen zurück. Radiologinnen und Radiologen prüfen ihren unabhängigen Befund zusammen mit der KI-Ausgabe, bevor sie die endgültige klinische Entscheidung treffen. Im Oktober 2025 erhielt Vara AI die CE-Zertifizierung als unabhängiger „Zweitbefunder".

Die Zertifizierung von Vara AI stützte sich größtenteils auf die PRAIM-Studie (2021–2023) [4], eine prospektive, multizentrische Beobachtungsstudie zur Bewertung von Leistungs- und Produktivitätseffekten (Eiseman et al., 2025). Die Studie verglich KI-unterstützte Doppelbefundung mit Standard-Doppelbefundung. Sie berichtete über höhere Erkennungsraten, kürzere Befundungszeiten bei unauffälligen Fällen und zusätzliche Karzinome, die durch Abweichungen zwischen KI und Mensch entdeckt wurden. Zugleich gab es einige Fälle, die vom KI-System übersehen wurden.

Seit Juli 2025 ist Vara AI zudem in das Hochrisiko-Screening-Programm QuaMaDi[5] in Schleswig-Holstein integriert.[6] Details zur Implementierung, einschließlich Auditmechanismen,

[1] Stand 2025, siehe BIPS – Mammographie-Screening senkt die Brustkrebsmortalität deutlich, abgerufen von: https://www.bips-institut.de

[2] Bundesamt für Strahlenschutz (BfS). (2022) Brustkrebsfrüherkennung mittels Röntgenmammographie bei Frauen ab 70 Jahren: Wissenschaftliche Bewertung des Bundesamtes für Strahlenschutz gemäß § 84 Absatz 3 Strahlenschutzgesetz https://doris.bfs.de/jspui/

[3] Institute for Quality and Efficiency in Health Care (IQWiG9) (2026) The breast cancer screening program in Germany https://www.informedhealth.org/

[4] Die Studie wurde zudem retrospektiv im Deutschen Register Klinischer Studien im März 2022 registriert, acht Monate nach ihrem Beginn (siehe Deutsches Register Klinischer Studien DRKS00027322, abgerufen am 15. Mai 2026 https://drks.de/)

[5] Bietet eine intensivere, häufig personalisierte Überwachung für asymptomatische Personen mit einem deutlich erhöhten Risiko, an Krebs zu erkranken (z. B. Personen mit BRCA-Genmutationen). Im Gegensatz dazu steht ein populationsbasiertes Screening-Programm, eine breite Maßnahme der öffentlichen Gesundheit, die systematisch Bevölkerungsgruppen mit durchschnittlichem Risiko (z. B. ausschließlich basierend auf Alter oder Geschlecht) zu standardisierten Untersuchungen einlädt. https://www.ncbi.nlm.nih.gov/books/NBK605831/

[6] QuaMaDi (Qualitätsgesicherte Mamma-Diagnostik) ist ein Diagnostikprogramm für Brustkrebs bei hohem Risiko, das sich vom standardmäßigen populationsbasierten Screening unterscheidet. Es richtet sich an Frauen, die aufgrund eines erhöhten individuellen Risikos überwiesen werden,

Patientenkommunikation und Plänen zur Leistungsüberwachung, sind jedoch weitgehend nicht öffentlich verfügbar.

Analyse der Fallstudie

F — Fairness: Derzeit liegt nur begrenzt öffentlich verfügbare Evidenz zur Leistung von Vara über Subgruppen hinweg vor, die über Alter und Brustdichte hinausgehen. Daten zu unterschiedlichen ethnischen oder sozioökonomischen Gruppen fehlen. Dadurch lässt sich schwer bewerten, ob das System über Bevölkerungsgruppen hinweg gerecht funktioniert. Unabhängige Fairness-Bewertungen sind aufgrund der proprietären Architektur von Vara AI kaum möglich.

U — Universality: Die Evidenz basiert auf europäischen Screening-Kontexten und spricht damit für eine gewisse Übertragbarkeit zwischen klinischen Standorten. Die Leistungsbewertung wird derzeit auf Mammographie-Screenings in Ägypten und Indien ausgeweitet; Details zur Implementierung oder Ergebnisse sind jedoch noch nicht öffentlich verfügbar. Zudem wurde die PRAIM-Studie vom Entwickler mitfinanziert und mitverfasst, was potenzielle Interessenkonflikte bei Interpretation und Berichterstattung mit sich bringt. Der Einsatz bleibt außerdem weitgehend auf strukturierte Screening-Umgebungen beschränkt. Unklar ist, wie gut sich die Leistung auf andere Gesundheitssysteme, Bevölkerungsrisiken, etwa Hochrisiko- gegenüber populationsbasiertem Screening, oder ressourcenärmere Settings übertragen lässt.

T — Traceability: Der Vara-AI-Anwendungsfall bietet nicht genügend öffentlich verfügbare Informationen, um Rückverfolgbarkeit über den gesamten KI-Lebenszyklus hinweg nachzuweisen. Zwar wurden klinische Leistungsdaten veröffentlicht. Zu zentralen Elementen wie Risikomanagement, technischer Dokumentation, kontinuierlicher Qualitätskontrolle, regelmäßigen Audits, Modellaktualisierungen, KI-Logging und Governance nach der Implementierung gibt es jedoch nur begrenzte Transparenz. Außerhalb der Studie erfolgt die Modellbewertung durch Vara intern und ohne unabhängige Prüfspur. Das erschwert es Gesundheitseinrichtungen und Forschenden, das Systemverhalten zu überwachen, Risiken über die Zeit zu bewerten und Rechenschaftspflicht sicherzustellen. In dem in Nature veröffentlichten Artikel zur PRAIM-Studie wird erwähnt, dass Frauen an den teilnehmenden Screening-Standorten über den Einsatz von KI informiert wurden. Informationen darüber, wie der Einsatz von Vara AI bei diagnostischen Beurteilungen gegenüber den Patientinnen kommuniziert wurde, sind nicht öffentlich verfügbar; daher konnte dieses Material weder geprüft noch bewertet werden.

U — Usability: Vara zeigte eine hohe Nutzbarkeit in klinischen Arbeitsabläufen. Das System integriert sich in das Routine-Screening, unterstützt radiologische Entscheidungen, ohne sie zu ersetzen, und verkürzt die Befundungszeit bei unauffälligen Fällen. In der PRAIM-Studie benötigten Radiologinnen und Radiologen für von der KI als unauffällig markierte Fälle 43 Prozent weniger Zeit als bei der Standardbefundung.[7] Das Design der Entscheidungsunterstützung soll zudem Automation Bias begrenzen, indem KI-Hinweise erst nach der initialen menschlichen Interpretation angezeigt werden. Ein gewisser Einfluss kann jedoch bestehen bleiben. Diese Ergebnisse wurden bisher nicht in Settings mit geringerem Fallaufkommen oder bei weniger erfahrenen beziehungsweise stärker belasteten Radiologinnen und Radiologen repliziert, wo das Risiko für Automation Bias höher sein könnte. Die Akzeptanz bei Patientinnen und die Erfahrung der Endnutzenden wurden nicht untersucht.

einschließlich familiärer Vorbelastung mit Brustkrebs, bekannter Hochrisiko-Genvarianten (z. B. BRCA1/2) oder Befunden wie einer hohen mammographischen Brustdichte.

[7] Die durchschnittliche Befundungszeit pro Bildbefundung wurde nur in der KI-Gruppe gemessen, da die Studie angibt, dass dies in der Standard-Befundungsgruppe technisch nicht messbar war (Eisemann et al., 2025).

R — Robustness: Die PRAIM-Studie liefert reale Leistungsdaten über verschiedene Standorte, Radiologinnen und Radiologen sowie Geräte hinweg. Eine angemessene Robustheitsprüfung für klinische KI-Systeme dieser Art ist jedoch bislang nicht definiert. Es fehlen etablierte Standards, etwa zur Bewertung von Sensitivität gegenüber unterschiedlichen Eingabedaten, Scannerunterschieden oder Arbeitsabläufen. Da die Studie von der Industrie finanziert und mitverfasst wurde[8] und keine unabhängige Replikation vorliegt, bleibt das Vertrauen in die Robustheit nur eingeschränkt.

E — Explainability: Vara zeigt der Nutzerin bzw. dem Nutzer verdächtige Regionen an. Erklärbarkeit hat jedoch viele Ebenen[9], und ein KI-Tool sollte auch erklären, wie und warum es zu einer solchen Schlussfolgerung gekommen ist. Vara teilt diese Entscheidungsfindung nicht mit den Behandelnden.

Ohne diese Information können Radiologinnen und Radiologen die Grundlage einer KI-Empfehlung nicht sinnvoll bewerten, was den klinischen Nutzen der bereitgestellten Erklärung einschränkt.

Zusammenfassung:

Vara AI zeigt vielversprechende klinische Vorteile, doch wichtige Lücken bestehen weiterhin bei der unabhängigen Validierung, der Transparenz und der langfristigen Leistungsüberwachung.

Zentrale Governance-Lehren:

- **Die unabhängige Validierung bleibt begrenzt**
- **Die Transparenz gegenüber Patientinnen und Patienten bleibt schwach**
- **Rechenschaftsmechanismen nach der Implementierung bleiben unterentwickelt**

[8] Obwohl Pharmaunternehmen häufig ihre eigenen Studien zur Rechtfertigung regulatorischer Zulassungen verwenden, sollte dies nicht das anzustrebende Modell sein.

[9] Erklärbarkeit in klinischer KI sollte als mehrschichtig verstanden werden: (1) visuelle Ausgabe – zeigt das System, was es markiert hat; (2) klinische Validität – sind die markierten Regionen diagnostisch korrekt; (3) Stabilität – bleiben die Erklärungen bei Eingabevariationen konsistent; und (4) mechanistische Transparenz – vermittelt das System, warum es zu einer Schlussfolgerung gekommen ist. Derzeitige regulatorische und Evaluationsrahmen für medizinische KI adressieren weitgehend nur die erste Ebene.

Fallstudie II: KI-Einsatz in der psychischen Gesundheitsversorgung im Vereinigten Königreich

Das NHS-Talking-Therapies-Programm bietet strukturierte Behandlung für Angststörungen und Depressionen in mehr als 200 Angeboten in England. Zuweisungen sind freiwillig und können von Patientinnen und Patienten oder von Behandelnden angestoßen werden. Die Nachfrage ist stetig gestiegen: Zwischen 2012/13 und 2021/22 haben sich die Zuweisungen nahezu verdoppelt (Nuffield Trust, 2025). Gleichzeitig lagen die Zugangsraten 2021 noch 22 Prozent unter den erwarteten Werten, vor allem aufgrund von Personalengpässen und begrenzten Aufnahmekapazitäten (The King's Fund, 2024). Evidenz zeigt, dass Personen, die eine Behandlung abschließen, bessere Ergebnisse erzielen als jene, die unbehandelt bleiben (NHS Digital, 2022).

Limbic Access ist ein hybrides KI-System, das 2021 auf der Talking-Therapies-Website eingeführt wurde, um Überweisungen zu beschleunigen und Zugangslücken zu schließen (Digitalhealth, 1). Das Tool ist ein regelbasierter Chatbot, der durch ein universelles großes Sprachmodell (LLM) und eine kognitive Schichtarchitektur unterstützt wird[10], um Halluzinationen zu verhindern. Nutzerinnen und Nutzer interagieren über eine textbasierte Oberfläche, ähnlich einer Messaging-App, ohne Avatar oder virtuelle Therapeutin beziehungsweise virtuellen Therapeuten. Sie beantworten strukturierte Fragen zu ihrer psychischen Gesundheit und ihren Lebensumständen. Die Antworten werden genutzt, um Aufnahmeinformationen zu erfassen und die klinische Einordnung vor der Einbindung einer klinischen Fachkraft zu unterstützen.

Evidenz aus mehreren Beobachtungsstudien deutet auf höhere Abschlussraten bei Zuweisungen und Effizienzgewinne hin (Rollwage et al., 2023; Habicht et al., 2024; Rollwage et al., 2024). Unabhängige Evaluationsrahmen und standardisierte Transparenzmechanismen sind jedoch nicht öffentlich dokumentiert.

Analyse der Fallstudie

F — Fairness: Evidenz aus NHS-Beobachtungsstudien deutet darauf hin, dass Limbic Access mit höheren Abschlussraten bei Zuweisungen verbunden ist: bei nichtbinären Nutzenden um 179 Prozent und bei Nutzenden aus ethnischen Minderheiten um 29 Prozent im Vergleich zu den jeweiligen Referenzgruppen (Habicht et al., 2024; Rollwage et al., 2024). Qualitative Befunde deuten zudem darauf hin, dass geringere Stigmatisierung und die nicht-menschliche Schnittstelle besonders Gruppen unterstützen können, die bei der Inanspruchnahme von Hilfe auf Barrieren stoßen. Zentrale Evidenz zur Untermauerung dieser Gerechtigkeitsaussagen fehlt jedoch. Es gibt keine öffentlich verfügbaren Informationen zur Repräsentativität der Trainingsdaten und kein formales Fairness-Auditing-Framework. Zwar gibt Limbic an, eine Architektur zur Erkennung von Verzerrungen und strukturierte Minderungsverfahren zu nutzen. Diese sind jedoch proprietär und nicht öffentlich zugänglich. Offen bleibt, ob die beobachteten Unterschiede nachhaltige Verbesserungen der Gerechtigkeit oder kontextspezifische Effekte des Einsatzes widerspiegeln. Die Offline-Funktionalität adressiert zudem Konnektivitätsbarrieren, die auch in Ländern mit hohem Einkommen relevant sind, und kann so die Reichweite in ländlichen oder benachteiligten Gebieten unterstützen.

U — Universality: Evidenz aus dem groß angelegten NHS-Einsatz über mehrere Dienste hinweg liefert eine reale Validierung innerhalb eines gut ausgestatteten, englischsprachigen Gesundheitssystems. Während eine systeminterne Validierung für eine reine Nutzung im Vereinigten Königreich ausreichen mag, ist Limbic aktiv auf der Suche nach internationalen

[10] https://limbic.ai/research/limbic-layer

Märkten.[11] Dies macht das Fehlen einer kontextübergreifenden Validierung zu einem wesentlichen Anliegen, insbesondere da Sprachmodelle sensibel auf Variationen reagieren, wie psychische Belastung über kulturelle und sozioökonomische Gruppen hinweg ausgedrückt wird, und für bestimmte Bevölkerungsgruppen optimiert werden können, während andere außen vor bleiben. Diese Sensibilitäten können die Klassifikationsgenauigkeit und die Qualität der Triage beeinträchtigen und erfordern häufig Anpassungen (Desai & Chaturvedi, 2017), was die vertrauenswürdige und nachhaltige Übertragbarkeit des Tools auf andere Kontexte infrage stellt.

T — Traceability: Limbic Access ist als britisches Medizinprodukt der Klasse IIa (UKCA) zertifiziert. Dies setzt rückverfolgbare Entwicklungs- und Risikomanagementprozesse voraus. Öffentlich verfügbare Informationen zur Systemarchitektur, zu Trainingsdaten, Modellaktualisierungen oder bekannten Fehlermodi fehlen jedoch. Nach Regelwerken wie dem EU AI Act würde eine konversationelle KI, die als Zugangstor zu psychologischer Unterstützung dient, voraussichtlich als Hochrisiko-System eingestuft. Idealerweise müsste dies zu mehr Transparenzpflichten führen, als derzeit sichtbar sind. Zugleich bleiben die Anforderungen an „Informationen für Nutzerinnen und Nutzer“ im EU AI Act vage. Sie legen nicht fest, wie Rückverfolgbarkeit und Transparenz über KI-Betreiber hinaus auf weitere Stakeholder ausgeweitet werden sollten – etwa Behandelnde, Patientinnen und Patienten oder unabhängige Prüferinnen und Prüfer.

U — Usability: Die chatbot-basierte Schnittstelle von Limbic, die darauf ausgelegt ist, die direkte menschliche Beteiligung am Aufnahmeprozess zu minimieren,[12] reduziert Stigmatisierung, erhöht die Abschlussrate von Überweisungen und integriert sich reibungslos in bestehende NHS-Aufnahmeprozesse. Evidenz deutet zudem auf nachgelagerte Vorteile hin, darunter bessere Teilnahme- und Genesungsraten, wenn das Tool in Versorgungspfade eingebettet ist (Rollwage et al., 2026). Die Nutzbarkeitsgewinne werden jedoch vor allem über Effizienz- und Engagement-Kennzahlen auf Systemebene gemessen. Unterschiedliche Nutzererfahrungen vulnerabler Gruppen bleiben nur begrenzt sichtbar.

R — Robustness: Im Kontext LLM-basierter Triage in der psychischen Gesundheit bestehen erhebliche Robustheitsrisiken. Neben allgemeinen Schwachstellen wie Halluzinationen, Verzerrungen, Datenlecks und feindlichen Eingaben (Bernini et al., 2026) reduziert Limbics hybride regel- und LLM-basierte Architektur Risiken wie Sycophancy, also das Bestätigen von Nutzerüberzeugungen statt klinisch fundierter Orientierung. Sie beseitigt diese Risiken jedoch nicht vollständig. Noch dringlicher sind Fehlermodi mit direkten Sicherheitsfolgen: falsche Einstufung des Schweregrads, unangemessene Triage-Entscheidungen und unzureichende Eskalation in Krisensituationen. Schutzmaßnahmen wie kognitive Sicherheitsarchitekturen, Suizidalitätserkennung und menschliche Aufsicht könnten diese Risiken mindern. Es gibt jedoch nur begrenzte Evidenz dafür, dass solche Systeme zum Zeitpunkt des Einsatzes im gesamten NHS vorhanden waren (Rollwage et al., 2026).

E — Explainability: Erklärbarkeit stellt die grundlegendste Lücke dar. Als LLM-basiertes Konversationssystem erzeugt Limbic Access flüssige Antworten, ohne die Evidenzgrundlage, Unsicherheiten oder Entscheidungslogik hinter den Ausgaben offenzulegen. Diese Intransparenz ist besonders problematisch, weil Limbic Access am Einstiegspunkt der Zuweisung eingesetzt wird. Abbrüche oder Fehlleitungen können den Zugang zur Versorgung direkt beeinträchtigen.

[11] Wie etwa die USA

[12] Limbic verwendet eine Human-in-the-Loop-Architektur, was bedeutet, dass an zentralen Entscheidungspunkten weiterhin eine ärztliche bzw. klinische Aufsicht besteht. Die Schnittstelle minimiert die direkte menschliche Interaktion während des patientenseitigen Aufnahmeprozesses, ist jedoch nicht vollständig autonom.

Derzeit gibt es keinen veröffentlichten Mechanismus, mit dem Ergebnisse wie Abbrüche oder verzögerte Hilfesuche zuverlässig auf konkrete KI-Interaktionen zurückgeführt werden können.

Zusammenfassung:

Limbic Access zeigt vielversprechende Verbesserungen bei Effizienz und Zugang im Rahmen von NHS Talking Therapies, doch wichtige Fragen zu Transparenz, unabhängiger Evaluation und Erklärbarkeit bleiben offen.

Zentrale Governance-Lehren:

- **Verbesserungen bei Zugang und Effizienz sind vielversprechend**
- **Transparenz und Erklärbarkeit bleiben begrenzt**
- **Die Infrastruktur für unabhängige Evaluation ist unzureichend**

Synthese

In beiden Fallstudien zeigen die KI-Systeme messbare Vorteile, zugleich aber fortbestehende Governance-Lücken. Obwohl sich die Technologien in Zweck und Design deutlich unterscheiden, lassen sich drei gemeinsame Erkenntnisse ableiten:

Erkenntnis 1: Die Evidenzgenerierung bleibt stark von den Entwicklern abhängig.

In beiden Fällen wurde ein Großteil der verfügbaren Evidenz von den Systementwicklern generiert, finanziert oder mitverfasst. Obwohl solche Studien wichtige Erkenntnisse liefern, erschweren die begrenzte unabhängige Validierung und Transparenz die Bewertung von Leistung, Fairness und Sicherheit über verschiedene Bevölkerungsgruppen und reale Settings hinweg. Eine stärkere öffentliche Evidenzinfrastruktur ist erforderlich, um eine vertrauenswürdige Einführung und Aufsicht zu unterstützen.

Erkenntnis 2: Der Einsatz erfolgt schneller als der Aufbau der Governance-Infrastruktur.

Beide Systeme haben einen umfangreichen realen Einsatz erreicht, trotz erheblicher Lücken bei Transparenz, unabhängiger Evaluation und Überwachung nach der Markteinführung, einschließlich kontextspezifischer Validierung für neue Einsätze. Dies spiegelt ein breiteres Muster bei KI im Gesundheitswesen wider: Die Implementierung schreitet schneller voran als die Governance-Mechanismen, die erforderlich sind, um die langfristige Leistung zu bewerten, unbeabsichtigte Folgen zu erkennen und Rechenschaftspflicht im Zeitverlauf sicherzustellen.

Erkenntnis 3: Der EU AI Act stellt Rechtmäßigkeit her, nicht jedoch notwendigerweise Vertrauenswürdigkeit.

Der EU AI Act bietet eine bedeutsame rechtliche Grundlage für Hochrisiko-KI-Systeme durch Anforderungen an Dokumentation, Risikomanagement und menschliche Aufsicht. Die Einhaltung der derzeitigen Anforderungen des EU AI Act allein garantiert jedoch nicht eine fortlaufend gerechte Leistung, sinnvolle Transparenz oder öffentliches Vertrauen. Die folgenreichsten Governance-Herausforderungen, die in beiden Fallstudien identifiziert wurden, entstehen in diesem Spannungsfeld zwischen rechtlicher Compliance und vertrauenswürdiger Umsetzung.

Diese Lücken liefern Umsetzungslehren zu unabhängiger Evaluation, Transparenz und kontextspezifischer Validierung. Sie sind relevant für das deutsche Gesundheitssystem, aber auch für Deutschlands Rolle in der globalen Gesundheit und auf internationalen Gesundheitstechnologiemärkten. Als einer der größten staatlichen Beitragszahler der WHO und als Befürworter stärkerer multilateraler Gesundheitsgovernance kann Deutschland dazu beitragen, internationale Diskussionen zu gemeinsamen Normen und Standards für KI im Gesundheitswesen voranzubringen – insbesondere im Bereich digitale Gesundheit und Gesundheits-KI.

Deutschland ist zugleich Anwender und Exporteur KI-gestützter Gesundheitstools. Diese Tools bewegen sich in transnationalen Technologieflüssen und werden auch in ressourcenärmeren Settings eingesetzt, teils mit stärkerer, teils mit schwächerer Regulierung als in Europa. Sie betreffen zudem vulnerable Bevölkerungsgruppen innerhalb des deutschen Gesundheitssystems. Gerade wenn KI-Systeme in andere sozioökonomische und kulturelle Kontexte übertragen werden, lässt sich ihre Leistung nicht einfach voraussetzen. Testing, Kalibrierung und Anpassung an den jeweiligen Kontext sind daher unerlässlich – ebenso wie Systemtransparenz und Aufsicht über die Marktzulassung hinaus.

Warum dies jetzt wichtig ist:

- **Patientensicherheit und Gerechtigkeit:** Unzureichend gesteuerte KI-Systeme können Fehler, Verzerrungen und ungleiche Ergebnisse in großem Maßstab verstärken, insbesondere bei vulnerablen Bevölkerungsgruppen.

- **Öffentliches Vertrauen:** Vertrauen in KI im Gesundheitswesen hängt nicht nur von technischer Leistungsfähigkeit ab, sondern auch von der Transparenz, Rechenschaftspflicht und sinnvollen Einbindung von Patientinnen und Patienten, die für eine langfristig gute Leistung erforderlich sind.

- **Europäische Wettbewerbsfähigkeit:** Da KI für Innovationen im Gesundheitswesen zunehmend zentral wird, kann eine vertrauenswürdige Governance zu einem strategischen Vorteil werden, der das öffentliche Vertrauen stärkt, die Akzeptanz fördert und die Führungsrolle Deutschlands bei der verantwortungsvollen Entwicklung von KI im Gesundheitswesen – regional und international – festigt.

Politikempfehlungen

Die Fälle zum Mammographie-Screening und zur Triage psychischer Gesundheit zeigen komplexe Herausforderungen und Chancen, die über Ländergrenzen hinweg relevant sind. Die folgenden Empfehlungen folgen einem einfachen Grundsatz: Vertrauenswürdige Governance ermöglicht nachhaltige Innovation. Ziel ist nicht, die Einführung von KI im Gesundheitswesen zu verlangsamen, sondern Evidenz-, Transparenz- und Rechenschaftsmechanismen zu schaffen, die eine sichere und gerechte Skalierung auf unterschiedliche Bevölkerungsgruppen ermöglichen.

Deutschland verfügt über die regulatorischen, klinischen und wissenschaftlichen Kapazitäten, um hier eine Führungsrolle einzunehmen. Die folgenden Empfehlungen übersetzen diese Kapazitäten in konkrete Governance-Maßnahmen. Sie sollen die praktische Umsetzung des EU AI Act im Gesundheitswesen unterstützen, Vertrauen stärken und Innovation fördern.

Nationale KI-Governance im Gesundheitswesen

Empfehlung 1 – Standards für die Offenlegung von KI gegenüber Patientinnen und Patienten schaffen

Begründung: Beide Fallstudien zeigen eine gemeinsame Governance-Lücke: Patientinnen und Patienten werden selten darüber informiert, wenn KI an ihrer Versorgung beteiligt ist. Der EU AI Act legt zwar Transparenzpflichten zwischen Entwicklern und Betreibern fest, bietet aber nur begrenzte Orientierung für die Offenlegung gegenüber Patientinnen und Patienten. Das schwächt informierte Einwilligung, Patientenautonomie und öffentliches Vertrauen. Einheitliche Standards für patientenorientierte Transparenz würden Rechenschaftspflicht stärken und dazu beitragen, dass Patientinnen und Patienten aktive Beteiligte KI-gestützter Versorgung bleiben.

Umsetzungsmaßnahmen: Das BMG sollte verbindliche Transparenzstandards für die Offenlegung von KI gegenüber Patientinnen und Patienten festlegen:

- Verpflichtung zu einer klaren, zugänglichen Offenlegung gegenüber Patientinnen und Patienten, wenn KI bei ihrer Diagnose, Behandlung oder ihrem Versorgungspfad eingesetzt wird
- Standardisierung der Kommunikationsformate gegenüber Patientinnen und Patienten – von aktualisierten Einwilligungsformularen bis hin zu sinnvollen Opt-out-Optionen in zugänglichen Formaten (z. B. digitale Schnittstellen und Informationsblätter)
- Gewährleistung eines sinnvollen Rechts für Patientinnen und Patienten, eine KI-gestützte Versorgung abzulehnen, soweit klinisch vertretbar

Empfehlung 2: Ein nationales Register und eine Evidenzinfrastruktur für KI im Gesundheitswesen schaffen

Begründung: Wirksame Governance setzt voraus, zu wissen, welche KI-Systeme eingesetzt werden, wo sie genutzt werden, wie sie funktionieren und ob die Evidenz für ihren Einsatz im Zeitverlauf gültig bleibt. Derzeit gibt es in Deutschland keine umfassende Infrastruktur, um den Einsatz von KI im Gesundheitswesen systematisch zu erfassen, Ergebnisse zu überwachen oder unabhängige Evaluation zu unterstützen. Ein nationales Register und eine Evidenzinfrastruktur könnten die derzeit fragmentierte „Black Box" lokaler Einsätze in eine transparente und überprüfbare öffentliche Ressource überführen. Das würde

kontinuierliche Aufsicht, Lernen nach der Markteinführung und evidenzbasierte Entscheidungsfindung ermöglichen (Fehr et al., 2024).

Umsetzungsmaßnahmen: Das BMG ist gut positioniert, um eine Registerinfrastruktur für KI im Gesundheitswesen sowie offene Berichtspflichten einzuführen, um dies zu adressieren, unter anderem durch:

- Einrichtung eines nationalen öffentlichen Registers (in Abwesenheit eines EU-weiten Registers) für klinisch eingesetzte KI-Systeme.
- Verpflichtung der Betreiber zu standardisierter öffentlicher Berichterstattung über Systemtyp, Verwendungszweck und regulatorischen Status, Einsatzorte (z. B. Krankenhäuser, Screening-Programme, Primärversorgung), Bevölkerungsabdeckung, Einsatzumfang, Aktualisierungen und Versionsänderungen im Zeitverlauf sowie Verweise auf Evaluationsergebnisse.
- Benennung und idealerweise Finanzierung einer unabhängigen Evaluationsstelle (z. B. innerhalb bestehender HTA-Strukturen), gegebenenfalls in Zusammenarbeit mit akademischen Einrichtungen, zur Durchführung von Validierungsstudien dort, wo die Evidenz überwiegend von der Industrie generiert wird.
- Einführung von Anforderungen an die Überwachung nach der Markteinführung, um Leistung, Drift und unbeabsichtigte Folgen im Zeitverlauf zu beobachten.

Empfehlung 3 – Öffentliche Finanzierung und Erstattung an vertrauenswürdige KI-Praktiken koppeln

Begründung: Öffentliche Finanzierung und Erstattung schaffen starke Anreize, die die Entwicklung und den Einsatz von KI prägen. Derzeitige Rahmenwerke honorieren jedoch häufig den Markteintritt, ohne eine transparente Evidenzgenerierung oder eine unabhängige Bewertung der Leistung über verschiedene Bevölkerungsgruppen hinweg zu verlangen. Die Auswahl der Empfänger öffentlicher Finanzierung und Erstattung auf Basis vertrauenswürdiger KI-Praktiken würde strengere Evidenzstandards fördern und dabei auf bestehenden Bewertungs- und Erstattungsmechanismen aufbauen, statt neue regulatorische Strukturen zu schaffen.

Umsetzungsmaßnahmen: Im Zuge der Umsetzung des EU AI Act hat das BMG die einzigartige Möglichkeit, die FUTURE-AI-Prinzipien der Rückverfolgbarkeit und Erklärbarkeit in Finanzierungs-, Bewertungs- und Erstattungspfade einzubetten. Dazu könnten folgende Voraussetzungen für die Auswahl von KI-Tools für Finanzierung und Erstattung gehören:

- Verpflichtung zur prospektiven Protokollregistrierung als nicht verhandelbare Bedingung für die öffentliche Finanzierung von KI-Pilotprojekten im Gesundheitswesen und vor Beginn jeglicher Datenerhebung.
- Verpflichtende Veröffentlichung der Dokumentation der Datensätze und ihrer demografischen Merkmale.
- Verpflichtung zu unabhängigen Fairness-Analysen mit öffentlich verfügbaren Ergebnissen.
- Integration dieser Anforderungen in bestehende HTA-, Erstattungs- und Finanzierungsprozesse, aufbauend auf bestehenden Strukturen wie AMNOG, statt parallele Verfahren zu schaffen.

Internationale KI-Governance

Empfehlung 4: Internationale Qualitätsstandards für den Einsatz von KI im Gesundheitswesen voranbringen

Begründung: Evidenz aus einem Gesundheitssystem lässt sich nicht automatisch auf wesentlich andere Systeme und Strukturen übertragen. Unterschiede in Sprache, Kultur, Gesundheitskompetenz, Krankheitsprävalenz, klinischen Arbeitsabläufen, Infrastruktur und regulatorischer Kapazität können Leistung und Sicherheit von KI erheblich beeinflussen. Dennoch verlangt derzeit kein breit etablierter internationaler Mechanismus eine kontextspezifische Validierung, bevor KI-Systeme im Gesundheitswesen in neuen Settings eingesetzt werden.

Ohne solche Schutzmaßnahmen besteht das Risiko, dass KI-Tools bei der Übertragung auf andere Bevölkerungsgruppen und Gesundheitssysteme bestehende Ungleichheiten verstärken, weniger wirksam sind oder neue Schäden verursachen. Durch kontextspezifische Validierungsstandards und Qualitätslabels als Anreiz im Markt könnte Deutschland dazu beitragen, dass KI-Systeme in internationalen Partnerschaften und Exporten sicher, wirksam und gerecht bleiben. Hohe Qualitätsanforderungen auf deutschen und europäischen Märkten können zugleich vergleichbare Standards in internationalen Kontexten fördern.

Umsetzungsmaßnahmen: Das BMG und das BMZ könnten zusammenarbeiten, um inklusive, fairnessbasierte Standards für den internationalen Einsatz von KI im Gesundheitswesen zu schaffen, insbesondere in weniger regulierten oder ressourcenärmeren Settings, unter anderem durch:

- Veröffentlichte, nach demografischen Subgruppen aufgeschlüsselte Leistungsdaten als Bedingung für die internationale Anerkennung von Evidenz im Rahmen der WHO, bilateraler Partnerschaften und Entwicklungszusammenarbeitsabkommen festlegen
- Entwicklung und verbindliche Vorgabe eines spezifischen Sicherheitsrahmens, der sprachliche Robustheit, Halluzinationen in klinischen Kontexten und die kulturelle Kalibrierung des Symptomausdrucks adressiert, bevor LLM-basierte Triage- bzw. Weiterleitungstools im Gesundheitswesen weltweit gefördert werden.

Empfehlung 5: Einen verantwortungsvollen Exportrahmen für KI im Gesundheitswesen schaffen

Begründung: In Europa entwickelte und validierte KI-Systeme im Gesundheitswesen werden zunehmend international eingesetzt, häufig in Settings mit anderen Patientenpopulationen, Gesundheitssystemen, Sprachen und regulatorischen Kapazitäten. Eine in einem Kontext nachgewiesene Leistung lässt sich nicht automatisch auf einen anderen Kontext übertragen. Ein verantwortungsvoller Exportrahmen, der internationale Expertise von Beginn an systematisch einbezieht – einschließlich Perspektiven aus Ländern mit niedrigem und mittlerem Einkommen (LMICs) – würde dazu beitragen, dass KI-Systeme auch im internationalen Einsatz sicher, wirksam und gerecht bleiben. Zugleich würde er Deutschlands übergeordnete globale Gesundheitsziele unterstützen.

Umsetzungsmaßnahmen: Deutschland ist gut positioniert, um die Zusammenarbeit zwischen dem BMG, dem BMZ und dem Bundesministerium für Forschung, Technologie und Raumfahrt (BMFTR) zu nutzen und einen internationalen Governance- und verantwortungsvollen Exportrahmen für das Gesundheitswesen zu entwickeln, um sicherzustellen, dass der vertrauenswürdige Einsatz von KI im Gesundheitswesen in internationalen Kontexten mit der eigenen innenpolitischen Praxis übereinstimmt. Dies könnte unter anderem folgende Maßnahmen umfassen:

- Verpflichtung zu kontextspezifischer Validierung vor internationalen Einsätzen von in Deutschland entwickelten KI-Systemen im Gesundheitswesen
- Einbettung strukturierter Komponenten zum Kapazitätsaufbau und zu technischen Partnerschaften in internationale Einsatzvereinbarungen.
- Verpflichtung zu dokumentierter lokaler Anpassung, wenn sich Sprache, Bevölkerung oder Versorgungskontext wesentlich unterscheiden.

Zweck dieses Papers
Dieses Policy Paper wurde von der Jahresthema-Arbeitsgruppe 2025/2026 des GHHG entwickelt, einer interdisziplinären Gruppe von Fachleuten aus Wissenschaft, Technologieentwicklung, klinischer Versorgung und öffentlicher Gesundheit, die KI im Gesundheitswesen im Rahmen ihrer Arbeit betrachten, bewerten, entwickeln oder nutzen.

Dieses Papier untersucht die duale Natur von KI im Gesundheitswesen und in der öffentlichen Gesundheit und bietet umsetzbare, fundierte Empfehlungen für Politikgestalterinnen und -gestalter sowie Regulierungsbehörden zur Sicherstellung eines vertrauenswürdigen Einsatzes von KI im Gesundheitswesen. Es steht dabei in explizitem Dialog mit aktuellen und sich entwickelnden regulatorischen Entwicklungen in Europa und darüber hinaus und identifiziert zentrale Bausteine einer evidenzbasierten Governance-Infrastruktur für KI im Gesundheitswesen – eine, die gerecht, rechenschaftspflichtig und der Komplexität globaler Gesundheitssysteme angemessen ist.

Bitte beachten: In einer früheren Version dieses Artikels (Juli 2026) wurde die Häufigkeit der Einholung der Patienteneinwilligung während der PRAIM-Studie falsch dargestellt.

Über den Global Health Hub Germany
Der Global Health Hub Germany bietet allen Einzelpersonen und Institutionen, die im Bereich globale Gesundheit aktiv sind, die Möglichkeit, sich in einem unabhängigen Netzwerk über acht verschiedene Stakeholdergruppen hinweg zu vernetzen: internationale Organisationen, Jugend, Politik, Stiftungen, Think Tanks, Wirtschaft, Wissenschaft und Zivilgesellschaft. Die Mitglieder des Hubs arbeiten gemeinsam an aktuellen Themen der globalen Gesundheit. Der interdisziplinäre Austausch erzeugt Themen, Fragestellungen und Lösungen, die der Hub an Politikgestalterinnen und -gestalter heranträgt, um eine informierte Politikgestaltung zu unterstützen und Fortschritte in der globalen Gesundheit zu erzielen. Der 2019 gegründete Hub hat heute rund 2.500 Mitglieder. Weitere Informationen finden Sie unter www.globalhealthhub.de.

Über die Hub-Communities
Die Hub-Communities sind Arbeitsgruppen, die von den Mitgliedern des Global Health Hub Germany selbst geleitet werden. Sie treffen sich regelmäßig, um Ideen auszutauschen, Fachwissen zu teilen und gemeinsam an Themen der globalen Gesundheit zu arbeiten. Wenn Sie einer Hub-Community beitreten oder mehr über ihre Arbeit erfahren möchten, wenden Sie sich an Katrin Lea Würfel, Head of Community Management: katrin.wuerfel@globalhealthhub.de.

Herausgegeben von: Global Health Hub Germany

c/o Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH
Köthener Str. 2-3, 10963 Berlin
Telefon: +49 30 59 00 20 210
info@globalhealthhub.de
www.globalhealthhub.de

Version: August 2026

Supported by:
Federal Ministry of Health
on the basis of a decision by the German Bundestag

Literaturverzeichnis

Ethischer Rahmen

- Beauchamp, T. L., & Childress, J. F. (1979). Principles of Biomedical Ethics. Oxford: Oxford University Press.
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2
- John, S., & Wu, J. (2022). "First, Do No Harm"? Non-Maleficence, Population Health, and the Ethics of Risk. *Social Theory and Practice*. 48(3), 525–551; https://aihcp.net/2024/09/10/understanding-non-maleficence-in-health-care-ethics/
- Lekadir, K., Frangi, A. F., Porras, A. R., Glocker, B., Cintas, C., Langlotz, C. P., Weicken, E., Asselbergs, F. W., Prior, F., Collins, G. S., Kaissis, G., Tsakou, G., Buvat, I., Kalpathy-Cramer, J., Mongan, J., Schnabel, J. A., Kushibar, K., Riklund, K., Marias, K., … Starmans, M. P. A. (2025). FUTURE-AI: International consensus guideline for trustworthy and deployable artificial intelligence in healthcare. *BMJ*, *388*, e081554. https://doi.org/10.1136/bmj-2024-081554
- World Health Organization (2021). Ethics and governance of artificial intelligence for health: WHO guidance. World Health Organization. https://iris.who.int/handle/10665/341996. Lizenz: CC BY-NC-SA 3.0 IGO

Fallstudie Mammographie

- Bundesamt für Strahlenschutz. (2025, Juli). Mammography screening considerably reduces breast cancer mortality. Abgerufen von https://www.bfs.de/SharedDocs/Pressemitteilungen/BfS/EN/2025/010.html
- Buschmann, L., Bonberg, D. S. N., Karch, A., Brenner, H., Harth, V., Heise, J. K., et al. (2025). Participation in the German mammography screening program: an analysis of data from the NAKO health study. *Deutsches Ärzteblatt International*, 122(24), 655. https://doi.org/10.3238/arztebl.m2025.0156
- Cuocolo, R., Bernardini, D., Pinto dos Santos, D., Klontzas, M. E., Akinci D'Antonoli, T., Semedo, L. C., ... & Williams, M. C. (2025). AI medical device post-market surveillance regulations: consensus recommendations by the European Society of Radiology. *Insights into Imaging*, *16*(1), 275.
- Eisemann, N., et al. (2025). Prospective multicenter observational study of an integrated AI system with live monitoring in mammography screening. *Nature Medicine*, 31, 188–196. https://doi.org/10.1038/s41591-024-03408-6
- Eisemann, N., & Katalinic, A. (2025). Research briefing. *Nature Medicine*, 31, 1422–1423. https://doi.org/10.1038/s41591-025-03714-7
- European Commission. (2025). European Health Data Space Regulation (EHDS). https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space-regulation-ehds_en
- Johner Institute. (2025). EU AI Act and medical devices. https://blog.johner-institute.com/iec-62304-medical-software/ai-act-eu-ai-regulation/
- Morley, J., Machado, C. C., Burr, C., Cowls, J., Joshi, I., Taddeo, M., & Floridi, L. (2020). The ethics of AI in health care: a mapping review. *Social Science & Medicine*, 260, 113172. https://doi.org/10.1016/j.socscimed.2020.113172
- Pressemitteilungen: Ägypten, 2026 (LinkedIn); Indien, 2024
- Reed Smith. (2025, Juni). The EU AI Act and medical devices: Navigating high-risk compliance. https://www.reedsmith.com/our-insights/blogs/viewpoints/102kq35/the-eu-ai-act-and-medical-devices-navigating-high-risk-compliance/

- Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Official Journal of the European Union*, L 2024/1689.
- Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices (Medical Device Regulation). *Official Journal of the European Union*, L 117/1.
- Regulation (EU) 2025/327 of the European Parliament and of the Council of 11 February 2025 on the European Health Data Space (EHDS). *Official Journal of the European Union*, L 2025/327
- Vara. (2025). PRAIM Study and Company Publications. https://www.vara.ai/
- World Health Organization. (2021). *Ethics and Governance of Artificial Intelligence for Health: WHO Guidance*.

Fallstudie Triage psychische Gesundheit

- AI.gov.uk Knowledge Hub. (2025) Limbic Access: Streamlining Access to Therapy through a chatbot. https://ai.gov.uk/knowledge-hub/tools/limbic-access/
- Aymen Dia Eddine Berini, Norziana Jamil, Ala-Eddine Benrazek, Abderrahmane Lakas, Leila Ismail, Mohamed Amine Ferrag, Kwok-Yan Lam. Security and privacy in LLMs: A comprehensive survey of threats and mitigation strategies, Information Fusion, Volume 132, 2026, 104241, ISSN 1566-2535, https://doi.org/10.1016/j.inffus.2026.104241. (https://www.sciencedirect.com/science/article/pii/S156625352600120X)
- Desai G, Chaturvedi SK. Idioms of Distress. J Neurosci Rural Pract. 2017 Aug;8(Suppl 1):S94-S97. https://doi.org/10.4103/jnrp.jnrp_235_17. PMID: 28936079; PMCID: PMC5602270.
- Digital Health London. (2022) Limbic. https://digitalhealth.london/innovation-directory/profile/limbic
- Habicht, J., Viswanathan, S., Carrington, B., Hauser, T. U., Harper, R., & Rollwage, M. (2024). Closing the accessibility gap to mental health treatment with a personalized self-referral chatbot. *Nature medicine*, *30*(2), 595–602. https://doi.org/10.1038/s41591-023-02766-x
- Limbic. (2026). The Most Proven AI in Mental Healthcare. https://limbic.ai/
- Limbic. (2026). Using AI to Make Care More Human Not Less: Meet Limbic Layer. https://limbic.ai/layer
- Naddaf M. AI chatbots are sycophants - researchers say it's harming science. Nature. 2025 Nov;647(8088):13-14. https://doi.org/10.1038/d41586-025-03390-0. PMID: 41136779.
- Rollwage M, Habicht J, Juechems K, Carrington B, Viswanathan S, Stylianou M, Hauser TU, Harper R. Using Conversational AI to Facilitate Mental Health Assessments and Improve Clinical Efficiency Within Psychotherapy Services: Real-World Observational Study. JMIR AI. 2023 Dec 13;2:e44358. Https://doi.org/10.2196/44358. PMID: 38875569; PMCID: PMC11041479.
- Rollwage, M., McFadyen, J., Juchems, K., Balogh, A., Pisupati, S., Mircea, M. T., Hauser, T. U., Prichard, G., & Harper, R. (2026). A cognitive layer architecture to support large-language model performance in psychotherapy interactions. *Nature medicine*, 10.1038/s41591-026-04278-w. Advance online publication. https://doi.org/10.1038/s41591-026-04278-w
- Rollwage, M. (2024). Our Commitment to Safety. *Resources: Blog / News,* 18. Juli 2024. https://limbic.ai/blog/our-commitment-to-safety. Abgerufen am 05. Mai 2026

Synthese & Empfehlungen

- Fehr J, Citro B, Malpani R, Lippert C, Madai VI. A trustworthy AI reality-check: agar transparency of artificial intelligence products in healthcare. Front Digit Health. 2024 Feb 20;6:1267290. Https://doi.org/10.3389/fdgth.2024.1267290. PMID: 38455991; PMCID: PMC10919164.
- https://pmc.ncbi.nlm.nih.gov/articles/PMC3240751/ - Automation bias: Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. J

Am Med Inform Assoc. 2012 Jan-Feb;19(1):121-7. Https://doi.org/10.1136/amiajnl-2011-000089 Epub 2011 Jun 16. PMID: 21685142; PMCID: PMC3240751.

- Parliamentary Assembly of the Council of Europe. Committee on Migration, Refugees and Displaced Persons. Artificial intelligence and migration: report (Doc. 15952). Strasbourg: Council of Europe; 2025
- Lundh A, Lexchin J, Mintzes B, Schroll JB, Bero L. Industry sponsorship and research outcome. Cochrane Database Syst Rev. 2017 Feb 16;2(2):MR000033. Https://doi.org/10.1002/14651858.MR000033.pub3. PMID: 28207928; PMCID: PMC8132492.
- Governance and structure of HTA bodies: https://toolbox.eupati.eu/resources/governance-and-structure-of-hta-bodies/#:~:text=In%20general%2C%20HTA%20bodies%20are%20established%20in,to%20administrative%20structures%20for%20a%20health%20system.
- https://www.mckinsey.com/industries/life-sciences/our-insights/generative-ai-in-the-pharmaceutical-industry-moving-from-hype-to-reality
- https://artificialintelligenceact.eu/high-level-summary/
- https://artificialintelligenceact.eu/national-implementation-plans/
- https://www.gleisslutz.com/en/know-how/federal-government-draft-bill-implement-eu-artificial-intelligence-act
- https://artificialintelligenceact.eu/article/50/
- https://www.technologyslegaledge.com/2025/11/state-of-the-act-eu-ai-act-implementation-in-key-member-states/
- https://go.pharmazie.com/en/pharma-pricing-germany-lt/
- https://www.imda.gov.sg/-/media/imda/files/about/emerging-tech-and-research/artificial-intelligence/mgf-for-agentic-ai.pdf
- https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for-organisation/ai/sgmodelaigovframework2.pdf
- https://gdpr-info.eu/art-9-gdpr/
- https://healthai.agency/app/uploads/2025/05/UNITE-HealthAIPositionPaper_March2025.pdf
- https://healthai.agency/app/uploads/2025/05/HealthAI_GlobalLandscapeReport_Oct.2024.pdf

Anhang I: Analytische Rahmenwerke

Tabelle 1: Analytische Rahmenwerke, die dem Policy Brief zugrunde liegen

Rahmenwerk/ Instrument	Kernfokus	Zentrale Prinzipien/Merkmale	Rolle im Policy Brief
FUTURE-AI International Consensus Guideline (Lekadir et al., 2025)	Operative Leitlinie für vertrauenswürdige KI im Gesundheitswesen	Nutzbarkeit, Robustheit, Fairness, Universalität, Rückverfolgbarkeit und Erklärbarkeit	Fungiert als operative analytische Linse für die Fallstudienanalyse und prägt die Empfehlungen
World Health Organization Guidance on AI for Health (2021)	Ethische Governance-Grundlage für KI in der globalen Gesundheit	Sechs ethische Prinzipien zu Autonomie, Rechenschaftspflicht, Gerechtigkeit, Transparenz, Menschenrechten und der Governance von KI im Gesundheitswesen	Dient als ethische Governance-Leitlinie im Kontext der globalen Gesundheit, als eine angesehene und offen gestaltete Ressource für ein Publikum internationaler Ministerien, Regulierungsbehörden, Unternehmen und Gesundheitseinrichtungen, von denen viele in ressourcenarmen Gesundheitssettings tätig sind und diese steuern
Principles of Biomedical Ethics (Beauchamp & Childress, 1979)	Normativer Rahmen für die klinische Ethik	Autonomie, Beneficizienz (Fürsorge), Nicht-Schädigung und Gerechtigkeit	Definiert übergreifende ethische Verpflichtungen, die Gesundheitssysteme gegenüber Patientinnen und Patienten haben, und verankert ethische KI-Debatten – einschließlich des FUTURE-AI-Leitliniendokuments – in einer langen Tradition der biomedizinischen Ethik

Anhang II: Darstellung der FUTURE-AI- und WHO-Prinzipien

Zwei internationale ethische Leitlinien für vertrauenswürdige/verantwortungsvolle KI im Gesundheitswesen:

Die WHO-Leitlinie legt fest, was im Gesundheitswesen in Bezug auf KI gewahrt werden muss, während das FUTURE-AI-Framework festlegt, wie KI-Systeme entwickelt und verwaltet werden müssen, um von Prinzipien zur Praxis zu gelangen.

FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare[13]

Sechs Kernprinzipien

1. **Fairness**

Fairness bedeutet, dass KI-Tools im Gesundheitswesen für alle Personen und Gruppen gerecht funktionieren sollten, einschließlich unterrepräsentierter und benachteiligter Bevölkerungsgruppen, und ohne systematische Verzerrung. Sie verlangt, Verzerrungen in Daten und Modelldesign zu erkennen, zu messen und zu mindern, damit eine durch KI informierte Versorgung bestehende gesundheitliche Ungleichheiten nicht verstärkt. Die Gewährleistung von Fairness stärkt das Vertrauen und unterstützt einen ethischen, diskriminierungsfreien Zugang zu KI-gestützter Gesundheitsversorgung.

2. **2Universalität**

Universalität betont, dass KI im Gesundheitswesen über verschiedene klinische Settings, Patientenpopulationen und Versorgungsumgebungen hinweg generalisierbar und interoperabel sein sollte. Dieses Prinzip erfordert, dass Entwickler die vorgesehenen Einsatzkontexte frühzeitig definieren und Tools mit externen und vielfältigen Datensätzen validieren, um eine breite Anwendbarkeit sicherzustellen. Durch die Unterstützung von Anpassungsfähigkeit und Übertragbarkeit trägt Universalität dazu bei, dass der Nutzen von KI nicht auf enge klinische Szenarien beschränkt bleibt.

3. **Rückverfolgbarkeit**

Rückverfolgbarkeit umfasst die Aufrechterhaltung einer robusten Dokumentation, Überwachung und Aufsicht über den gesamten KI-Lebenszyklus hinweg, sodass Entwicklung, Einsatz, Ergebnisse und mögliche Fehler nachvollziehbar und überprüfbar sind. Sie unterstützt die Rechenschaftspflicht, indem sie Rollen und Verantwortlichkeiten von Entwicklern, Behandelnden und Institutionen festlegt und klare Verfahren ermöglicht, um Probleme oder Schäden im Zusammenhang mit dem KI-Einsatz zu untersuchen und zu beheben. Eine wirksame Rückverfolgbarkeit fördert das Vertrauen in KI-Systeme und gewährleistet die Einhaltung ethischer und regulatorischer Standards.

4. **Nutzbarkeit**

[13] Lekadir K, Frangi A F, Porras A R, Glocker B, Cintas C, Langlotz C P et al. FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare *BMJ* 2025; 388 :e081554 doi:10.1136/bmj-2024-081554

Nutzbarkeit bedeutet, dass KI-Tools mit Blick auf die tatsächlichen Nutzerinnen und Nutzer entwickelt werden müssen – Behandelnde, Patientinnen und Patienten sowie andere Stakeholder –, um sicherzustellen, dass sie sicher und wirksam in klinische Arbeitsabläufe integriert werden können. Dazu gehört, Endnutzende frühzeitig einzubeziehen, um Bedürfnisse zu verstehen, intuitive Schnittstellen zu gestalten und Tools in realen Settings zu evaluieren. Eine hohe Nutzbarkeit fördert die Akzeptanz, verringert Fehler und verbessert den klinischen Nutzen von KI-Technologien.

5. Robustheit

Robustheit bezieht sich auf die Zuverlässigkeit, Stabilität und Widerstandsfähigkeit von KI-Systemen unter variierenden realen Bedingungen, einschließlich Schwankungen bei Daten, Geräten und klinischen Praktiken. Ein robustes KI-Tool sollte seine Leistung über verschiedene Kontexte hinweg aufrechterhalten und rigoros getestet werden, um Schwachstellen oder Fehlermodi zu erkennen. Dieses Prinzip trägt dazu bei, dass KI-Systeme im Zeitverlauf und über verschiedene Versorgungsumgebungen hinweg sicher, verlässlich und widerstandsfähig bleiben.

6. Erklärbarkeit

Erklärbarkeit verlangt, dass KI-Systeme für Entwickler, Behandelnde, Regulierungsbehörden und mitunter auch Patientinnen und Patienten transparent und nachvollziehbar sind, sodass ihre Funktionsweise und Entscheidungen verstanden und hinterfragt werden können. Dazu gehört eine klare Dokumentation von Datenquellen, Modelllogik, Annahmen und Einschränkungen sowie deren zielgruppengerechte Kommunikation. Erklärbarkeit unterstützt Vertrauen, informierten Einsatz und eine wirksame Aufsicht über KI im Gesundheitswesen.

Das FUTURE-AI-Framework liefert zudem allgemeine Empfehlungen:

- **Stakeholder kontinuierlich einbeziehen –** KI-Entwickler sollten vielfältige Stakeholder über den gesamten KI-Lebenszyklus hinweg einbeziehen, um Bedürfnisse zu antizipieren, Risiken zu erkennen und Akzeptanz sowie eine verantwortungsvolle Einführung zu unterstützen.
- **Datenschutz gewährleisten –** Über den gesamten KI-Lebenszyklus hinweg müssen starke Maßnahmen zum Schutz der Privatsphäre, zur Sicherheit, Governance und Cybersicherheit angewendet werden, um Gesundheitsdaten zu schützen und Missbrauch oder Re-Identifizierung zu verhindern.
- **Maßnahmen zur Bewältigung von KI-Risiken umsetzen –** Entwickler sollten Risiken wie Verzerrungen, mangelnde Robustheit und unzureichende Generalisierbarkeit proaktiv durch geeignete Modellierungs- und Validierungsstrategien erkennen und mindern.
- **Einen angemessenen KI-Evaluationsplan festlegen –** KI-Tools sollten anhand unabhängiger Testdaten, geeigneter Kennzahlen und eines Vergleichs mit Standards oder bestehenden Lösungen evaluiert werden, um eine vertrauenswürdige Leistung sicherzustellen.
- **KI-Vorschriften einhalten –** Geltende KI-Gesetze und regulatorische Anforderungen sollten frühzeitig identifiziert und während der gesamten Entwicklung und des Einsatzes eingehalten werden, um rechtliche und ethische Compliance sicherzustellen.
- **Anwendungsspezifische ethische Fragestellungen untersuchen –** Entwickler sollten systematisch ethische, soziale und gesellschaftliche Fragestellungen

identifizieren und adressieren, die für jede einzelne KI-Anwendung im Gesundheitswesen spezifisch sind.

- **Soziale und ökologische Fragestellungen untersuchen –** Die breiteren sozialen, arbeitsmarktbezogenen und ökologischen Auswirkungen von KI im Gesundheitswesen, einschließlich Nachhaltigkeit und CO2-Fußabdruck, sollten bewertet und gemindert werden, um positive gesellschaftliche Ergebnisse sicherzustellen.

WHO-Leitlinie – Ethik und Governance von KI für die Gesundheit[14]

Sechs zentrale ethische Prinzipien zur Orientierung bei der Entwicklung und Nutzung von KI-Technologie für die Gesundheit. Während ethische Prinzipien universell sind, kann ihre Umsetzung je nach kulturellem, religiösem und anderem gesellschaftlichem Kontext unterschiedlich ausfallen.

- **Autonomie schützen:** Dieses Prinzip verlangt, dass KI-Systeme die menschliche Entscheidungsfindung im Gesundheitswesen unterstützen, statt sie zu ersetzen, und sicherstellen, dass Behandelnde und Patientinnen und Patienten die Kontrolle über medizinische Entscheidungen behalten. Eine sinnvolle menschliche Aufsicht muss stets gegeben sein, sodass KI-Entscheidungen überwacht, hinterfragt, außer Kraft gesetzt oder bei Bedarf rückgängig gemacht werden können. Der Schutz der Autonomie umfasst zudem die Wahrung von Privatsphäre, Vertraulichkeit und informierter Einwilligung, wobei KI niemals zur Manipulation oder zu Experimenten an Personen ohne gültige Einwilligung oder zur Einschränkung des Zugangs zu essenziellen Gesundheitsleistungen eingesetzt werden darf.
- **Menschliches Wohlergehen, menschliche Sicherheit und das öffentliche Interesse fördern:** KI-Technologien im Gesundheitswesen dürfen keinen Schaden verursachen und sollten sowohl vor als auch nach dem Einsatz strenge Standards für Sicherheit, Genauigkeit und Wirksamkeit erfüllen. Entwickler, Geldgeber und Nutzer tragen eine fortlaufende Verantwortung, die Leistung zu überwachen und unbeabsichtigte physische, psychische oder soziale Schäden zu erkennen. Der Einsatz von KI sollte das Wohl der Patientinnen und Patienten sowie das öffentliche Interesse priorisieren, insbesondere indem Diskriminierung, Stigmatisierung oder Schäden durch Diagnosen oder Informationen verhindert werden, auf die Patientinnen und Patienten vernünftigerweise nicht reagieren können.
- **Transparenz, Erklärbarkeit und Verständlichkeit gewährleisten:** KI-Systeme sollten für Entwickler, Nutzerinnen und Nutzer, Regulierungsbehörden und betroffene Personen durch Transparenz hinsichtlich ihres Designs, ihrer Daten, Annahmen und Einschränkungen verständlich sein. Erklärbarkeit sollte auf unterschiedliche Zielgruppen zugeschnitten sein, damit Menschen KI-gestützte Entscheidungen sinnvoll verstehen und hinterfragen können, selbst wenn technische Komplexität Herausforderungen schafft. Transparente Testung, Evaluation und unabhängige Aufsicht sind unerlässlich, um Sicherheit, Fairness, Rechenschaftspflicht und Vertrauen in realen Versorgungssettings zu gewährleisten.
- **Verantwortung und Rechenschaftspflicht fördern:** Obwohl KI-Systeme Aufgaben übernehmen, liegt die Verantwortung für ihren Einsatz und ihre Ergebnisse stets bei

[14] World Health Organization (2021). Ethics and governance of artificial intelligence for health: WHO guidance. World Health Organization. https://iris.who.int/handle/10665/341996. Lizenz: CC BY-NC-SA 3.0 IGO

menschlichen Akteuren und Institutionen. Es müssen klare Rechenschaftsmechanismen bestehen, um Verantwortung zuzuweisen, Abhilfe zu schaffen und Wiedergutmachung zu gewährleisten, wenn KI-Systeme Schaden verursachen. Um eine Verantwortungsdiffusion zu vermeiden, sollten alle an der Entwicklung, dem Einsatz und der Nutzung von KI beteiligten Stakeholder eine gemeinsame Verantwortung dafür tragen, Schaden zu minimieren und ethische Standards zu wahren.

- **Inklusion und Gerechtigkeit gewährleisten:** KI im Gesundheitswesen sollte so gestaltet und eingesetzt werden, dass sie einen gerechten Zugang und Nutzen für alle Bevölkerungsgruppen fördert, unabhängig von Alter, Geschlecht, Einkommen, Fähigkeiten oder geografischer Lage. Entwickler müssen Verzerrungen in Daten und Algorithmen aktiv erkennen, verhindern und mindern, um eine Verstärkung gesundheitlicher Ungleichheiten oder Diskriminierung zu vermeiden. Inklusive Beteiligung, vielfältige Datensätze und kontinuierliches Monitoring sind notwendig, um sicherzustellen, dass KI-Technologien allen zugutekommen, insbesondere marginalisierten und vulnerablen Gruppen.
- **Reaktionsfähige und nachhaltige künstliche Intelligenz fördern:** KI-Systeme müssen kontinuierlich evaluiert werden, um sicherzustellen, dass sie angemessen auf reale Gesundheitsbedürfnisse reagieren und in ihrem jeweiligen Einsatzkontext wie vorgesehen funktionieren. Reaktionsfähigkeit umfasst die Möglichkeit, KI-Technologien zu modifizieren, zu verbessern oder einzustellen, wenn sie unwirksam, schädlich oder nicht nachhaltig sind. Nachhaltigkeit erfordert eine Abstimmung mit der langfristigen Kapazität des Gesundheitssystems, der Anpassung der Arbeitskräfte und der ökologischen Verantwortung, sodass KI Gesundheitssysteme im Zeitverlauf stärkt, statt sie zu belasten.