

FluxLane-Studie

KI-Ethik in Krankenhaus und Unternehmen

Nutzen und Risiken, Evidenz und Verantwortung

Forschungsbericht · Für Ethikbeauftragte, pflegerische Leitung, klinische Ethikgremien und Unternehmensverantwortliche
Recherche- und Rechtsstand: 24. September 2026 · Deutschland und Europäische Union

KI-gestütztes Manuskript für externe fachliche Begutachtung. Keine eigene empirische Datenerhebung. Kein systematischer Review mit Vollständigkeitsanspruch. Keine bereits erfolgte externe Peer Review.

Dieser Forschungsbericht dient der allgemeinen Information und ersetzt keine einzelfallbezogene Rechtsberatung.

Steckbrief

Merkmal	Wert
Dokumenttyp	Strukturierte kritische Literaturstudie mit normativer und rechtsbezogener Governanceanalyse
Zielgruppe	Ethikbeauftragte, pflegerische Leitung, klinische Ethikgremien, Unternehmensverantwortliche
Recherche- und Rechtsstand	24. September 2026
Umfang	19.202 Wörter, 50 Seiten
Belege	46 Literatureinträge, 95 Quellenadressen
Gliederung	22 nummerierte Abschnitte, 14 Tabellen

Inhalt

Zusammenfassung	6
Entscheidungsrelevante Ergebnisse	6
1 Forschungsauftrag und Gegenstand	7
1.1 Erkenntnisinteresse	7
1.2 Forschungsfragen	7
1.3 Untersuchungseinheit und Abgrenzung	7
1.4 Begriffe und Ergebnisarten	8
2 Untersuchungsmethode	8
2.1 Design und methodischer Status	8
2.2 Quellenstränge	8
2.3 Recherchewege und Auswahl	9
2.4 Extraktion und Umgang mit fehlenden Angaben	9
2.5 Kritische Bewertung	9
2.6 Synthese ohne gemeinsame Effektberechnung	10
2.7 Normative Analyse und Ableitung von Empfehlungen	10
2.8 Qualitätssicherung und Grenzen der Durchführung	10
3 Aufbau und Reichweite des Evidenzbestands	11
4 Empirische Ergebnisse und kritische Bewertung	11
4.1 Erkenntnisarten im Vergleich	11
4.2 Kritische Studienprofile	12
4.3 Synthese der offenen Nachweisaufgaben	16
5 Risiken und institutionelle Wirkmechanismen	16
5.1 Überzeugende Fehler und unbemerkte Auslassungen	16
5.2 Diskriminierung ist mehr als ein unvollständiger Datensatz	16
5.3 Datenschutz und Sicherheit sind eigenständige Prüfböcke	16
5.4 Beziehung, Kompetenz und Arbeitsbedingungen können unter Druck geraten	17
5.5 Abhängigkeit, Konzentration und Ressourcenverbrauch gehören zur Ethikprüfung	17
5.6 Schwerwiegende Zukunftsrisiken weder verschweigen noch als Gewissheit verkaufen	17
6 Rechtliche Voraussetzungen und Grenzen	17
6.1 Der KI-Rechtsrahmen hat mehrere Zeitachsen	17
6.2 Datenschutz und Arbeitsrecht warten nicht auf spätere Hochrisikofristen	19
6.3 Produktkonformität, fachlicher Standard und Ethik sind nicht dasselbe	19
6.4 Zuständigkeiten nicht verwechseln	19
6.5 Einsatz und rechtliche Rollen	20
6.6 Rollenwechsel und Änderungen	20
6.7 Verträge und tatsächliche Verantwortlichkeit	21
6.8 Information Einwilligung und Beteiligung	21
6.9 Konstruierte rechtliche Grenzfälle	21
6.10 Nachweise für die Rechtsprüfung	22

7 Normative Diskussion und begründete Empfehlungen	23
7.1 Von wissenschaftlichen Befunden zu verantwortbaren Entscheidungen	23
7.2 Autonomie Nicht jede Wahlmöglichkeit schafft Selbstbestimmung	23
7.3 Nutzen und Nichtschaden Auch Nichtstun hat Folgen	24
7.4 Gerechtigkeit Verteilung statt bloßer Durchschnittsleistung	24
7.5 Beziehung und Fürsorge Der Kontakt ist nicht nur eine Prozessvariable	25
7.6 Verantwortung Befugnisse müssen zur übernommenen Aufgabe passen	25
7.7 Zielkonflikte, die ausdrücklich entschieden werden sollten	26
7.8 Empfehlungen Wissensgrundlage, Werturteil und Nachweis verbinden	26
7.9 Umsetzung Gemeinsamkeiten und Grenzen zwischen Krankenhaus und Konzern	28
8 Zuständigkeiten und institutionelle Verantwortung	28
9 Risikoadäquates Entscheidungsverfahren	29
9.1 Auftrag, Vergleich und Grenzen festlegen	30
9.2 Belege statt Vorführungen verlangen	30
9.3 Rechte, Betroffene und Zielkonflikte prüfen	30
9.4 Einen begrenzten Pilot mit vorab festgelegten Kriterien planen	30
9.5 Eine begrenzte, begründete Freigabe erteilen	31
9.6 Betrieb, Änderungen und Ausstieg organisieren	31
10 Messung und lokale Evaluation	31
10.1 Warum 95 Prozent Genauigkeit wenig sagen kann	31
10.2 Lokales Evaluationsprotokoll	32
11 Limitationen und Forschungsagenda	35
11.1 Grenzen der vorliegenden Untersuchung	35
11.2 Prioritäten weiterer Forschung	35
12 Schlussfolgerung	36
13 Informationswege und fachliche Unterstützung	36
A Arbeitsvorlagen	38
A.1 Arbeitsstruktur für die ersten 90 Tage	38
A.2 Fragen vor einer Beschaffung	38
A.3 Entscheidungsvorlage	39
A.4 Auslöser einer Neubewertung	39
A.5 Glossar	39
B Rechercheprotokoll der empirischen Vertiefung	40
B.1 Zugang zu den empirischen Originalarbeiten	40
B.2 Ergänzende Volltextsuche	40
B.3 Ergänzende Gegenrecherche	41
B.4 Auswahlentscheidungen zu fünf Kandidaten	42
B.5 Reichweite der Ergänzung	43
C Rechercheprotokoll der Rechtsvertiefung	43
C.1 Ausgeführte Suchabfragen	43
C.2 Zugriff und Grenzen	44

C.3 Abschließende Gegenprüfung	44
D Transparenz und Publikationsstatus	45
Literatur und Primärquellen	45
13 Rechtliche Hinweise	49
13.1 Keine Beratung	49
13.2 Keine Zusicherung, keine Konformitätsaussage	49
13.3 Rechtsstand	49
13.4 Haftung	49
13.5 Verweise auf fremde Inhalte	49
13.6 Marken und Produktbezeichnungen	49
13.7 Urheberrecht und Nutzung	49
13.8 Keine personenbezogenen Daten	50
13.9 Unabhängigkeit	50
13.10 Anwendbares Recht	50

Impressum und Verantwortlichkeit

Die folgenden Angaben sind wörtlich dem Impressum von fluxlane.eu entnommen, abgerufen am 24. September 2026. Sie wurden nicht ergänzt und nicht umformuliert.

Angabe	Inhalt
Angaben gemäß § 5 DDG	WebTrades GmbH Am Ölberg 2 55459 Grolsheim Deutschland
Vertreten durch	Geschäftsführerin: Laurice Bussmer
Kontakt	Telefon: 0170 890 78 00 E-Mail: info@fluxlane.eu
Registereintrag	Eintragung im Handelsregister Registergericht: Amtsgericht Mainz Registernummer: HRB 54479
Umsatzsteuer-ID	Umsatzsteuer-Identifikationsnummer gemäß § 27 a Umsatzsteuergesetz: DE461650292
Redaktionell verantwortlich	Verantwortlich für den Inhalt nach § 18 Abs. 2 MStV: Laurice Bussmer, Am Ölberg 2, 55459 Grolsheim
Verbraucherstreitbeilegung	Wir sind nicht bereit oder verpflichtet, an Streitbeilegungsverfahren vor einer Verbraucherschlichtungsstelle teilzunehmen.

Redaktioneller Hinweis des Herausgebers (wörtlich)

„fluxlane ist ein unabhängiges, redaktionelles Fachportal für Robotik, autonome physische Systeme und verkörperte KI. Der redaktionelle Geltungsbereich umfasst Industrie und Produktion, Logistik, Medizin, Pflege und Rehabilitation, Service und Haushalt sowie Forschung, Hersteller, Markt und Regulierung; künstliche Intelligenz gehört dazu, auch ohne physischen Körper. Die Inhalte dienen der allgemeinen Information und stellen keine Rechts-, Kauf- oder Anlageberatung dar. Angaben zu Preisen, Verfügbarkeit und Terminen erfolgen ohne Gewähr und können sich jederzeit ändern.“

Zitierweise

FluxLane (2026): *KI-Ethik in Krankenhaus und Unternehmen – Nutzen und Risiken, Evidenz und Verantwortung*. Fassung 1.1, Stand 24. September 2026. Herausgegeben von der WebTrades GmbH, Grolsheim, fluxlane.eu.

Herstellung dieses Dokuments

Das vorliegende PDF ist maschinell aus der Markdown-Fassung des Manuskripts gesetzt. Vor dem Satz prüft das Erzeugungsskript Belegziffern gegen das Literaturverzeichnis, die Anrede und benannte Datumsfallen; schlägt eine Prüfung fehl, entsteht kein Dokument. Das Erzeugungsskript und die Quelldatei liegen dem Herausgeber vor. Gesetzt wurde mit IBM Plex Sans und Sora.

Abbildungsnachweis

Dieses Dokument enthält keine Fotografien. Alle Abbildungen sind Diagramme, die beim Satz aus den im Text belegten Werten gezeichnet wurden; die jeweilige Fundstelle steht unter der Abbildung. Die Diagramme sind KI-gestützt erstellt.

Zusammenfassung

Hintergrund. Krankenhäuser und große Unternehmen müssen über Anwendungen künstlicher Intelligenz entscheiden, obwohl technische Leistungsfähigkeit, nachgewiesener Nutzen und institutionelle Verantwortbarkeit häufig nicht deckungsgleich sind. Ethikbeauftragte benötigen dafür nicht nur Prinzipien, sondern eine nachvollziehbare Verbindung zwischen empirischen Befunden, rechtlichen Voraussetzungen und konkreten Entscheidungen.

Fragestellung. Unter welchen Bedingungen ist der Einsatz von KI in Versorgung und Arbeit ethisch begründbar, welche Risiken müssen geprüft werden und welche institutionellen Voraussetzungen ermöglichen wirksame Verantwortung?

Methode. Die Untersuchung verbindet eine strukturierte kritische Literaturstudie mit einer normativen und rechtsbezogenen Dokumentenanalyse. Ausgangspunkt ist ein gezielt ausgewählter Bestand empirischer Originalarbeiten zu klinischer Unterstützung, Dokumentation, Verteilungswirkungen und Wissensarbeit. Eine ergänzende Gegenrecherche prüft Versorgung außerhalb europäischer Hochressourcensysteme sowie soziale Robotik. Ausgewertet werden außerdem amtliche Rechtsquellen und institutionelle Ethik- und Risikomanagementdokumente. Die Synthese erfolgt nach Aufgaben und Endpunkten, nicht durch statistische Zusammenfassung ungleicher Interventionen. Recherchewege, Zugangsgrenzen und eigene Bewertungsentscheidungen werden offengelegt. Es handelt sich nicht um einen systematischen Review mit Vollständigkeitsanspruch.

Ergebnisse. Die ausgewertete Evidenz stützt den Nutzen bestimmter Anwendungen unter bestimmten Einsatzbedingungen. Sie trägt weder eine pauschale Überlegenheit von KI noch eine allgemeine Sicherheitsgarantie menschlicher Schlusskontrolle. Zentrale Prüfprobleme betreffen die Wahl des Zielwerts, Übertragbarkeit, Zusammenarbeit zwischen Menschen und Systemen, Arbeitsbelastung, Verteilungsfolgen und Veränderungen nach der Einführung. Rechtliche Einordnung und ethische Bewertung greifen ineinander, ersetzen einander aber nicht.

Schlussfolgerung. Empfohlen wird eine anwendungsbezogene, risikoadäquate und revidierbare Freigabe. Sie verbindet fachliche Evidenz mit Beteiligung, dokumentierter Verantwortlichkeit, wirksamer Aufsicht und fortlaufender Ergebnisprüfung. Die pflegerische Leitung ist im Krankenhaus eine eigenständige Gestaltungspartnerin. Das vorgeschlagene Governanceverfahren ist eine begründete Synthese und bedarf selbst einer empirischen Evaluation.

Schlüsselbegriffe: künstliche Intelligenz; klinische Ethik; Unternehmensethik; Pflegeleitung; menschliche Aufsicht; Evidenzbewertung; KI Governance.

Entscheidungsrelevante Ergebnisse

Die zentrale organisatorische Schlussfolgerung für die Leitung einer Einrichtung ist die Bindung einer Freigabe an den konkreten Einsatz. Ein guter Modelltest beantwortet nicht, ob Menschen mit diesem System unter den eigenen Arbeitsbedingungen bessere Entscheidungen treffen. Dafür müssen Aufgabe, Version, Daten, verfügbare Fachkompetenz und Reaktionsmöglichkeiten zusammen betrachtet werden.

Für Ethikbeauftragte ergibt sich daraus ein präzises Arbeitsfeld. Sie müssen nicht jedes technische Detail selbst prüfen. Im Rahmen ihres Mandats können sie zur Zusammenführung der Prüfungen beitragen und deren Grenzen sichtbar machen; die Leitung benennt die für das Gesamtverfahren verantwortliche Stelle. Eine Datenschutzprüfung kann beispielsweise keine klinische Wirksamkeit bescheinigen. Eine Nutzenstudie beantwortet nicht von selbst, ob eine zusätzliche Überwachung von Beschäftigten angemessen ist. Auch ein Ethikvotum übernimmt nicht die Organisationsverantwortung der Geschäftsführung.

Die Studie empfiehlt, zunächst die folgenreichen Anwendungen und ihre Verantwortlichen zu erfassen. Neue Aufgaben, neue Datenzugriffe und wesentliche Änderungen sollen nicht stillschweigend von einer alten Freigabe profitieren. Für Entscheidungen mit erheblichem Schadenspotenzial sind geeignete Vergleichsdaten, realistische Tests des Arbeitsablaufs und vorab bestimmte Reaktionen auf Warnsignale erforderlich. Bei risikoarmen, leicht korrigierbaren Anwendungen kann der Prüfungsweg entsprechend kürzer sein.

Im Krankenhaus sollten Pflegeleitung, ärztliche Verantwortung und unmittelbar Anwendende gemeinsam festlegen, wie ein System in die Versorgung eingebunden wird. Ihre Perspektiven unterscheiden sich: Eine Empfehlung kann fachlich plausibel sein und trotzdem zu spät ankommen, unklare Zuständigkeiten erzeugen oder mehr Reaktionsarbeit verlangen, als eine Schicht leisten kann. Diese Konflikte sind Gegenstand der Einführung und dürfen nicht zu persönlichen Versäumnissen einzelner Mitarbeitender umgedeutet werden.

Im Unternehmen stehen insbesondere Auswahl- und Bewertungsprozesse, Kundeninteraktionen sowie Systeme mit umfassenden Daten- oder Handlungsrechten im Vordergrund. Die Prüfung muss berücksichtigen, wer Produktivitätsgewinne erhält und wer zusätzliche Belastungen oder Fehlerfolgen trägt. Ein durchschnittlicher Zeitgewinn rechtfertigt nicht automatisch Personalabbau, Überwachung oder die Einschränkung von Überprüfungsmöglichkeiten.

Die unmittelbar umsetzbare Empfehlung ist eine dokumentierte Entscheidung mit fünf verbindlich zu klärenden Punkten: erlaubter Zweck, nachgewiesene Eignung, verantwortliche Stelle, Bedingungen der Aufsicht und Anlass für eine erneute Prüfung. Diese fünf Punkte sind ein hier vorgeschlagenes Organisationsprinzip, keine wörtliche Wiedergabe einer einheitlichen gesetzlichen Pflicht. Die rechtlichen Voraussetzungen bleiben separat zu bestimmen.

1 Forschungsauftrag und Gegenstand

1.1 Erkenntnisinteresse

Die Untersuchung richtet sich an Ethikbeauftragte, klinische Ethikgremien, pflegerische Leitungen und Verantwortliche großer Unternehmen in Deutschland. Sie soll begründete Entscheidungen über KI ermöglichen und dabei sowohl einen unnötigen Verzicht auf nützliche Technik als auch eine vorschnelle Übertragung folgenreicher Aufgaben vermeiden.

Die Forschungsfragen verbinden unterschiedliche Erkenntnisarten. Ob ein System die Dokumentationszeit verändert, ist eine empirische Frage. Ob die dabei entstehende Datenverarbeitung zulässig ist, ist eine Rechtsfrage. Ob gewonnene Zeit vorrangig zur Entlastung oder zur Leistungsausweitung genutzt werden sollte, ist eine Wert- und Organisationsentscheidung. Der Bericht behandelt diese Fragen gemeinsam, hält ihre Begründungswege aber getrennt.

1.2 Forschungsfragen

Tabelle 1 Forschungsfragen und Erkenntnisbedarf

Kennung	Forschungsfrage	Benötigte Erkenntnis
F1	Welche nachgewiesenen Vorteile und Nachteile entstehen bei den ausgewählten Anwendungen?	Vergleichende empirische Befunde mit Endpunkten und Unsicherheit.
F2	Unter welchen Bedingungen sind Befunde auf eine andere Einrichtung übertragbar?	Population, Version, Daten, Arbeitsablauf, Ressourcen und Vergleichspraxis.
F3	Welche Werte und Rechte begrenzen eine rein leistungsbezogene Entscheidung?	Normative Argumente, Betroffenenperspektiven und Rechtsprüfung.
F4	Welche Zuständigkeiten und Verfahren unterstützen eine verantwortbare Einführung?	Institutionelle Standards und begründete Organisationsvorschläge.
F5	Wie lässt sich nach der Einführung erkennen, ob Nutzen und Vertretbarkeit bestehen bleiben?	Messplan, Beschwerdewege, Änderungsprüfung und Vorfallsbearbeitung.

Die Fragen erlauben keine einzige gemeinsame Effektgröße. Sie führen zu einer differenzierten Entscheidungssystematik und zu einer Forschungsagenda. Eine universelle Antwort auf die Frage nach dem „Nutzen von KI“ wird nicht angestrebt.

1.3 Untersuchungseinheit und Abgrenzung

Untersuchungseinheit ist eine Anwendung in ihrem institutionellen Kontext. Dazu zählen das technische System, sein Zweck, seine Eingaben, die Schnittstelle, die Rolle der Anwender und die Folgen seiner Verwendung. Ein frei

zugänglicher Chatbot und derselbe Modelltyp innerhalb einer geprüften Dokumentationsumgebung sind daher nicht dieselbe Intervention.

Der klinische Teil berücksichtigt diagnostische Unterstützung, Warnsysteme und Dokumentation. Soziale Robotik wird als ethisches Anwendungsfeld und offene Evidenzlücke behandelt, nicht als quantitativ ausgewerteter Wirksamkeitsbereich. Der Unternehmensbezug umfasst Wissensarbeit, Kundenservice und Personalentscheidungen. Nicht umfassend untersucht werden autonome Waffensysteme, Finanzhandel, die vollständige Robotersicherheitsnormung oder jede branchenspezifische Zulassung. Diese Abgrenzung verhindert, dass die Schlussfolgerungen außerhalb ihrer Grundlage wie ein allgemeines Prüfsiegel verwendet werden.

Bei körpernaher Robotik müssen physische Gefährdungen und produktspezifische Wirksamkeit zusätzlich bewertet werden. Ergebnisse aus Sprachmodellstudien lassen sich dafür nicht als Ersatz verwenden. Umgekehrt belegt eine kleine soziale Roboterstudie keine sichere autonome Durchführung von Pflegehandlungen.

1.4 Begriffe und Ergebnisarten

Unter Nutzen versteht dieser Bericht eine relevante Verbesserung gegenüber einer benannten Alternative. Je nach Anwendung kann das ein Gesundheitsendpunkt, eine fachlich korrekte Bearbeitung oder eine tatsächlich eingesparte Arbeitsbelastung sein. Ein positives Nutzungserlebnis kann wichtig sein, ist jedoch nicht mit medizinischer Wirksamkeit gleichzusetzen.

Risiko bezeichnet hier die Möglichkeit einer nachteiligen Folge unter bestimmten Bedingungen. Nicht alle Folgen sind in einer gemeinsamen Geld- oder Punkteskala sinnvoll abzubilden. Ein unzulässiger Eingriff kann beispielsweise nicht allein durch eine hohe erwartete Zeitersparnis gerechtfertigt werden. Die Studie verwendet deshalb keine scheinbar objektive Gesamtnote, in der Datenschutz, Patientensicherheit und Effizienz gegeneinander verrechnet werden.

Wirksame menschliche Aufsicht bedeutet in dieser Untersuchung mehr als die Anwesenheit einer Person. Gemeint ist eine organisatorisch und fachlich realistische Möglichkeit, das Systemverhalten zu verstehen, relevante Fehler zu erkennen, eine Handlung zu verhindern oder zu korrigieren und bei Unsicherheit Unterstützung zu erreichen. Wie viel Aufsicht erforderlich ist, hängt vom konkreten Eingriff und dessen Folgen ab.

2 Untersuchungsmethode

2.1 Design und methodischer Status

Die Arbeit ist eine strukturierte kritische Literaturstudie mit ergänzender normativer Governanceanalyse. Sie ist eine Sekundäranalyse öffentlich zugänglicher Forschung und Dokumente. Es wurden keine Patienten untersucht, keine Beschäftigten befragt und keine Produktversuche durchgeführt. Die später dargestellten Anwendungssituationen dienen der Argumentation und sind keine erhobenen Fälle.

Die Recherche wurde in zwei Schritten durchgeführt. Der erste Schritt erzeugte einen gezielten Quellenbestand für einen Fachbeitrag. Der zweite Schritt vertiefte diesen Bestand für den vorliegenden Forschungsbericht, prüfte Zugänge und methodische Grenzen und ergänzte eine Gegenrecherche. Die im Anhang dokumentierten Abfragen des zweiten Schritts sind tatsächliche Recherchehandlungen. Sie bilden keine nachträglich erfundene Registrierung des ersten Schritts.

Die Studie beansprucht weder eine vollständige Erfassung aller einschlägigen Veröffentlichungen noch eine vorregistrierte systematische Auswahl. Entsprechend wird kein PRISMA-Flussdiagramm mit nicht dokumentierten Treffer-, Dubletten- oder Ausschlusszahlen dargestellt. PRISMA 2020 dient hier zur Abgrenzung dessen, was eine systematische Übersichtsarbeit transparent berichten müsste; seine bloße Erwähnung macht diese Untersuchung nicht zu einem solchen Review.[39]

Für die Berichtsgüte wurden die von SANRA behandelten Bereiche als Orientierung herangezogen: Relevanz und Ziel, Beschreibung der Suche, Referenzierung, wissenschaftliche Argumentation und geeignete Ergebnisdaten. Eine unabhängige SANRA-Bewertung oder ein Qualitätszertifikat wird nicht behauptet.[38]

2.2 Quellenstränge

Die Untersuchung unterscheidet vier Quellenstränge. Der empirische Strang umfasst Originaluntersuchungen zu Wirkung, Leistung und Mensch-System-Interaktion. Der institutionelle Strang erfasst Ethikkodizes, Standards und fachliche Orientierung. Der Rechtsstrang beruht auf amtlichen Normtexten und deren Änderungen; behördliche Erläuterungen werden als Interpretationshilfe gekennzeichnet. Der Methodenstrang betrifft die Bewertung und Darstellung von Evidenz.

Diese Trennung verhindert eine unzulässige Vermischung von Autorität. Eine Empfehlung einer internationalen Organisation ist kein randomisierter Wirkungsnachweis. Ein einzelnes Experiment begründet keine rechtliche Ausnahme. Ein Gesetz wiederum beantwortet nicht alle Fragen der praktischen Eignung. Im Quellenverzeichnis bleiben diese Dokumenttypen erkennbar.

2.3 Recherchewege und Auswahl

Recherchiert wurde webbasiert in Verlagsangeboten, öffentlich zugänglichen Repositorien, PubMed-Nachweisen und offiziellen institutionellen Angeboten. Für Rechtsfragen wurden insbesondere EUR-Lex und Gesetze im Internet sowie zuständige Behörden verwendet. Eine PubMed-Fundstelle wurde nicht mit einer vollständig ausgeführten systematischen PubMed-Datenbanksuche gleichgesetzt. Suchmaschinen liefern zudem dynamisch gerankte Ergebnisse; eine Wiederholung derselben Abfrage kann andere Treffer zeigen.

Die Auswahl des empirischen Kernbestands war zweckgerichtet. Aufgenommen wurden Studien, die mindestens eines der folgenden Probleme unmittelbar untersuchten: patientennahe Wirkung oder Leistungsfähigkeit, tatsächliche Arbeitsentlastung, Zusammenarbeit zwischen Fachpersonen und KI, Benachteiligung durch ein Ziel- oder Datenproblem oder Leistungsunterschiede zwischen Aufgaben. Unterschiedliche Ergebnisrichtungen waren erwünscht. Der Bestand ist deshalb ein analytischer Querschnitt, keine zufällige Stichprobe und keine Grundlage für eine Erfolgsrate aller KI-Anwendungen.

Vorrang hatten veröffentlichte Originalarbeiten mit identifizierbarem Design und benanntem Endpunkt. Frühere Arbeitspapiere wurden nicht zusätzlich als unabhängige Studie gezählt, wenn eine finale Fassung vorlag. Herstellerwerbung, reine Produktankündigungen und unbelegte Einschätzungen wurden nicht als Wirksamkeitsbeleg verwendet. Bei begrenztem Volltextzugang wurde die Quelle nicht automatisch verworfen, ihre Bewertung aber auf die tatsächlich zugänglichen Informationen beschränkt.

Der zeitliche Schwerpunkt liegt auf jüngerer Forschung; ältere Untersuchungen wurden aufgenommen, wenn sie einen relevanten Fehlmechanismus oder eine Produktentwicklung verständlich machen. So werden die ältere und neuere Sepsismodell-Validierung nebeneinander betrachtet. Daraus entsteht kein direkter randomisierter Vergleich der beiden Publikationen.

2.4 Extraktion und Umgang mit fehlenden Angaben

Für die empirischen Arbeiten wurden bibliografische Identität, Forschungsfrage, Design, Population, Vergleich, technische Intervention, Endpunkt, Effektangabe, Unsicherheit und wesentliche Begrenzungen erfasst. Finanzierungs- und Interessenkonfliktangaben werden nur berichtet, wenn sie im zugänglichen Original nachgesehen wurden. Ein nicht erhobenes Feld ist weder gleichbedeutend mit „nicht berichtet“ noch mit „nicht vorhanden“.

Die Bewertung verwendet deshalb drei klar getrennte Zustände: im zugänglichen Text belegt, dort offen beziehungsweise nicht ausreichend beurteilbar und für die jeweilige Frage nicht anwendbar. Diese Unterscheidung ist insbesondere bei Verblindung, Vorregistrierung, fehlenden Daten und Herstellerbezug wichtig. Aus einem kurzen Abstract lässt sich zum Beispiel kein vollständiges Urteil über selektive Berichterstattung ableiten.

Numerische Angaben werden zusammen mit Vergleich und Endpunkt wiedergegeben. Prozentänderungen werden nicht als Prozentpunkte umgeschrieben. Ein Konfidenzintervall wird als Ausdruck statistischer Unsicherheit behandelt, nicht als Garantie für den Effekt in einer neuen Einrichtung. Eine fehlende statistische Signifikanz beweist weder Gleichheit noch die vollständige Abwesenheit eines relevanten Effekts.

2.5 Kritische Bewertung

Die Studien wurden anhand einer gemeinsamen Fragenstruktur und designspezifischer Ergänzungen bewertet. Die Struktur fragt nach Vergleichbarkeit der Gruppen, Erfassung des Endpunkts, Einfluss der Einführungsbedingungen, Unsicherheit und Übertragbarkeit. Bei Prognosemodellen werden insbesondere zeitlicher Bezug, externe Validierung, Kalibrierung und praktische Folgen von Schwellen betrachtet. Bei experimentellen Aufgaben wird zwischen Leistungsfähigkeit in einer Versuchsanordnung und Wirkung im Arbeitsalltag unterschieden.

Diese Bewertung ersetzt keine vollständige, formal durchgeführte RoB-2-, ROBINS-I- oder QUADAS-Bewertung. Insbesondere werden keine numerischen Risikoscores und keine GRADE-Evidenzstufen ausgegeben. Eine solche Einstufung würde eine enger definierte Fragestellung, einen dafür geeigneten vollständigen Studienbestand und eine dokumentierte Anwendung des jeweiligen Instruments verlangen.

Die Formulierungen „direkter“, „indirekter“ oder „begrenzter“ Nachweis beziehen sich hier immer auf eine benannte Aussage. Eine große prognostische Studie kann beispielsweise eine belastbare Aussage zur Rangtrennung liefern und gleichzeitig keine kausale Aussage zur Sterblichkeit erlauben. Eine kleine randomisierte Simulation kann einen Mechanismus zeigen, ohne dessen Häufigkeit im Versorgungssystem zu quantifizieren.

2.6 Synthese ohne gemeinsame Effektberechnung

Die Befunde werden nach Aufgaben und Ergebnisarten geordnet. Klinische Wirkung, Prognosegüte, Dokumentationszeit und Qualität einer Managementaufgabe werden nicht in eine einzige Kennzahl umgerechnet. Auch die Anzahl statistisch positiver Studien wird nicht als Abstimmung über den Gesamtnutzen interpretiert. Das würde unterschiedliche Interventionen, Vergleichspraktiken und Messverfahren unzulässig gleichsetzen.

Für die Transparenz der narrativen Synthese ist die SWiM-Berichtsempfehlung relevant, die unter anderem die Gruppierung von Studien und die Begründung von Schlussfolgerungen behandelt. Sie wurde hier als methodische Orientierung herangezogen, ohne eine vollständige SWiM-konforme systematische Interventionsübersicht zu behaupten.[40]

Die Synthese fragt deshalb: Welche Aussage trägt der jeweilige Befund? Welcher nächste Schluss wäre bereits eine Überdehnung? Welche lokale Prüfung bleibt nötig? Wo Ergebnisse auseinandergehen, werden mögliche Unterschiede in Produkt, Aufgabe oder Kontext berücksichtigt. Ein Scheinkonsens durch Weglassen widersprechender Resultate soll vermieden werden.

2.7 Normative Analyse und Ableitung von Empfehlungen

Empirische Befunde werden mit ausdrücklich benannten ethischen Gründen verbunden. Ein beobachtetes Risiko begründet noch nicht von selbst die genaue zulässige Risikoschwelle. Diese hängt auch von Schwere und Verteilung möglicher Schäden, Alternativen, Freiwilligkeit und institutionellen Pflichten ab. Umgekehrt müssen normative Forderungen praktisch umsetzbar sein; ein Aufsichtsmodell ohne Personal oder Zeit wäre nur nominell vorhanden.

Die Empfehlungen werden daher nach einer nachvollziehbaren Kette dargestellt: Befund oder Problem, ethischer Grund, vorgeschlagene Handlung, verantwortliche Stelle, prüfbarer Nachweis und verbleibende Unsicherheit. Gesetzliche Voraussetzungen sind separat ausgewiesen. Weder ein Expertenkonsens noch eine Delphi-Befragung wird behauptet. Betroffene haben an dieser Ausarbeitung nicht als Studiengruppe mitgewirkt; ihre tatsächliche Beteiligung wird für die lokale Einführung empfohlen und bleibt dort durchzuführen.

2.8 Qualitätssicherung und Grenzen der Durchführung

Recherche und Ausarbeitung erfolgten KI-gestützt. Parallel erarbeitete Teilbewertungen wurden mit den Originalnachweisen abgeglichen und in einer zentralen Synthese zusammengeführt. Das ist keine unabhängige Doppelbegutachtung durch menschliche Fachpersonen. Fehler können trotz dieser Kontrollen verbleiben, insbesondere bei dynamischen Webseiten und unvollständigem Volltextzugang.

Für die Dokumentprüfung werden Zitationsverweise, Zahlenübernahmen, Inhaltsverzeichnis und Tabellen auf Konsistenz kontrolliert. Der gesetzte Bericht wird zusätzlich seitenweise auf Lesbarkeit geprüft. Diese redaktionellen Prüfungen dürfen nicht als externe wissenschaftliche Begutachtung ausgegeben werden. Vor einer institutionellen Verwendung mit Bindungswirkung sind fachliche und rechtliche Prüfung im konkreten Kontext erforderlich.

3 Aufbau und Reichweite des Evidenzbestands

Der empirische Kern umfasst zehn Originalstudien. Neun stammen aus dem Ausgangsbestand, eine weitere wurde nach der dokumentierten Gegenrecherche ergänzt. Acht Hauptartikel waren im Volltext zugänglich; MASAI wurde anhand des offiziellen Autorenabstracts und Lukac anhand indexierter Originalabschnitte bewertet. Die Arbeiten werden im nächsten Kapitel einzeln beschrieben. Sie erfüllen unterschiedliche Funktionen innerhalb der Analyse: Manche prüfen eine Intervention, andere ein Vorhersagemodell oder einen Mechanismus möglicher Fehlentscheidungen. Eine Studie ist deshalb nicht einfach „gut“ oder „schlecht“. Zu bestimmen ist, für welche Aussage sie einen geeigneten Nachweis liefert.

Tabelle 2 Erkenntnisarten und Aussagegrenzen

Erkenntnisart	Typische Frage	Was der Nachweis allein nicht leistet
Vergleichende Wirkungsstudie	Ändert der Einsatz einen relevanten Endpunkt gegenüber der Vergleichspraxis?	Universelle Übertragbarkeit auf andere Versionen und Einrichtungen.
Externe Modellvalidierung	Wie verhält sich eine Vorhersage in anderen Daten oder Standorten?	Nachweis besserer Behandlung oder kausaler Schadensvermeidung.
Experiment zu menschlichen Entscheidungen	Wie verändert ein Hinweis die Bearbeitung einer Aufgabe?	Häufigkeit entsprechender Schäden unter realen Alltagsbedingungen.
Untersuchung von Verteilungswirkungen	Welche Gruppen werden durch Zielwahl oder Anwendung benachteiligt?	Vollständige Erklärung sämtlicher Ungleichheit oder ein universelles Fairnessmaß.
Feldstudie zur Arbeit	Ändern sich bearbeitete Mengen, Zeit oder Qualität?	Gesamtwirtschaftliche Beschäftigungswirkung oder automatische Übertragbarkeit auf Pflege.

Ein Vergleich zwischen Krankenhaus und Konzern kann daher auf der Ebene der Prüfprobleme hilfreich sein, ohne die Ergebnisse zahlenmäßig zu übertragen. Die Frage nach einem passenden Zielwert stellt sich in beiden Bereichen. Die zulässige Fehlerrate, die geeignete Kontrolle und die Folgen eines Irrtums können trotzdem grundverschieden sein.

Die einzelnen Studienprofile kennzeichnen den tatsächlichen Zugang zur Quelle. Nicht vollständig beurteilbare Aspekte bleiben offen. Im weiteren Bericht beruhen Empfehlungen bei solchen Lücken gegebenenfalls auf einem Vorsorge- oder Gerechtigkeitsargument. Diese normative Begründung wird nicht als zusätzlicher empirischer Wirksamkeitsnachweis präsentiert.

4 Empirische Ergebnisse und kritische Bewertung

Recherche- und Zugriffsprüfung: 24. September 2026. Die zehn Studien tragen die Referenznummern [3]–[11] und [41] des Hauptmanuskripts. Die nachfolgenden Profile trennen berichtete Daten von unserer eigenen methodischen Bewertung. Es handelt sich um eine strukturierte, selektive Evidenzbewertung, nicht um vollständig ausgefüllte RoB-2-, QUADAS- oder GRADE-Verfahren. Nicht verfügbare Angaben werden nicht als günstiger Befund gewertet. Finanzierungs- und Interessenkonflikthinweise sind ausdrücklich ausgewählte, tatsächlich gelesene Offenlegungen, keine vollständige Wiedergabe sämtlicher Erklärungen. Anhänge und Registerhistorien wurden nicht vollständig geprüft.

4.1 Erkenntnisarten im Vergleich

Tabelle 3 Vergleichsmatrix der zehn Originalstudien

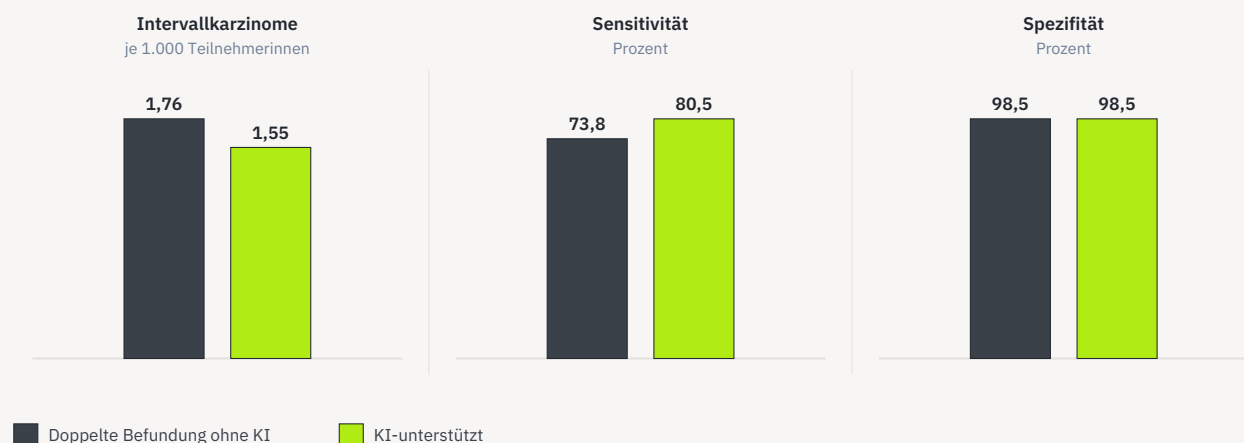
Quelle	Gegenstand	Tragende Erkenntnisart	Nicht gleichzusetzen mit
[3]	Screeningablauf	Randomisierter Nichtunterlegenheitsvergleich	Nachweis geringerer Sterblichkeit
[4]	Diagnostisches Begründen	Randomisierter Vignettentest	Wirkung auf Behandlungsergebnisse

Quelle	Gegenstand	Tragende Erkenntnisart	Nicht gleichzusetzen mit
[5]	Sepsisprognose, Version 1	Retrospektive externe Validierung	Heutiger Produktleistung
[6]	Sepsisprognose, Version 2	Prospektive multizentrische Validierung	Kausalem Versorgungsnutzen
[7]	Dokumentation	Pragmatischer randomisierter Vergleich	Gesamter eingesparter Arbeitszeit
[8]	Automationsgläubigkeit	Kontrollierte experimentelle Fehlberatung	Fehlerhäufigkeit realer KI-Produkte
[9]	Verteilungswirkung	Audit eines eingesetzten Algorithmus	Interventionsnachweis für Deutschland
[10]	Kundendienstassistentz	Quasiexperimentelle Feldanalyse	Gesamtwirtschaftlichem Beschäftigungseffekt
[11]	Wissensarbeit	Randomisiertes aufgabenbezogenes Experiment	Pauschalem Leistungsgewinn
[41]	Klinische Entscheidungshilfe	Cluster-randomisierter Versorgungsvergleich	Bewiesener Sicherheitsgleichwertigkeit

Die rechte Spalte ist unsere methodische Grenzziehung, keine zusätzliche Ergebnissammlung. Die Tabelle liefert bewusst weder einen gemeinsamen Effektwert noch ein Gesamtranking. Ein Experiment zur Fehlberatung und ein Screeningversuch beantworten verschiedene Fragen; sie können sich ergänzen, aber nicht gegenseitig ersetzen.

4.2 Kritische Studienprofile

Abbildung 1 – MASAI-Studie: die drei gemessenen Größen



Werte aus Abschnitt 4.2 dieser Studie, Beleg [3]. Jedes Paneel trägt eine eigene Skala, weil die Größen verschiedene Einheiten haben; die Balkenhöhen sind deshalb zwischen den Paneelen nicht vergleichbar. Nichtunterlegenheit bei einem Verhältnis von 0,88 (95%-KI 0,65–1,18) – kein Nachweis einer gesicherten Reduktion und keiner geringeren Sterblichkeit.

[3] Gommers et al. MASAI, 2026

Datenauszug. Schwedischer randomisierter Screeningversuch: 105.934 Frauen, 19 nach Randomisierung ausgeschlossen. Vergleich: KI-gestützte Triage/Befundung gegen doppelte menschliche Befundung. Primärendpunkt: Intervallkarzinome; Nichtunterlegenheitsgrenze 20 %. Raten: 1,55 gegenüber 1,76 je 1.000; Verhältnis 0,88 (95%-KI 0,65–1,18). Sensitivität: 80,5 gegenüber 73,8 %, Spezifität jeweils 98,5 %. Förderung: Swedish Cancer Society, Regional Cancer Centres und staatliche klinische Forschungsförderung; Interessenkonflikte nicht verifiziert. Grundlage hier: offizielles Autorenabstract, kein zugänglicher Volltext.[3]

Eigene Bewertung – fünf Domänen. Zuteilung: Randomisierung stärkt Vergleichbarkeit; Verdeckung nicht im Volltext nachgeprüft. Ausfälle: wenige Ausschlüsse, deren Gründe bleiben hier ungeprüft. Endpunkt: näher an Versorgungsrelevanz als reine Testfallgenauigkeit, aber kein Mortalitätsendpunkt. Unsicherheit: das Intervall erlaubt sowohl Vorteil als auch moderaten Nachteil; Nichtunterlegenheit ist keine bewiesene Überlegenheit. Übertragbarkeit: schwedischer Screeningablauf und konkrete Systemintegration begrenzen die Generalisierung.

Zulässige Schlussfolgerung/offene Prüfung. Der untersuchte Ablauf hält die festgelegte Nichtunterlegenheitsgrenze ein. Vor lokaler Übernahme sind Volltext, Protokollabweichungen, Ausschlussgründe,

Langzeitfolgen und populationsbezogene Passung nachzuprüfen. „Zwölf Prozent weniger Krebs“ wäre als gesicherte Wirkung unzulässig.

[4] Goh et al. diagnostisches Begründen, 2024

Datenauszug. Randomisierte Studie mit 50 US-Ärzten; GPT-4 plus übliche Informationsquellen gegen übliche Quellen allein. Sechs schriftliche Fälle standen zur Verfügung, 244 Fallbearbeitungen wurden bewertet. Primärendpunkt war ein diagnostischer Begründungsscore, verblindet beurteilt. Adjustierte Differenz: +2 Prozentpunkte (95%-KI -4 bis +8; p=0,60). Förderung: Gordon and Betty Moore Foundation; ein Mitautor war dort beschäftigt, weitere finanzielle Beziehungen wurden offengelegt.[4]

Eigene Bewertung – fünf Domänen. Zuteilung: randomisierter Vergleich ist eine Stärke. Messung: verblindete Beurteilung reduziert Bewertungsbias, ein Score bleibt jedoch ein Ersatzmaß. Präzision: die Stichprobe schließt moderate Vorteile oder Nachteile nicht aus. Umsetzung: freier Chatbotzugang prüft eine bestimmte Nutzungssituation, kein optimiertes Assistenztraining. Übertragbarkeit: ausgewählte Vignetten erfassen weder reale Anamneseinteraktion noch Versorgung unter Schichtdruck.

Zulässige Schlussfolgerung/offene Prüfung. Diese Assistenzform verbesserte den Score nicht statistisch nachweisbar. Das beweist weder allgemeine Nutzlosigkeit noch klinische Austauschbarkeit von Ärzten und KI. Für eine Einführung wären ein integrationsnaher Vergleich, tatsächliche Entscheidungsfehler, Folgemaßnahmen und patientenrelevante Endpunkte erforderlich.

[5] Wong et al. Epic-Sepsismodell Version 1, 2021

Datenauszug. Retrospektive externe Validierung an Michigan Medicine: 27.697 Erwachsene, 38.455 Aufenthalte. Modellvorhersagen wurden mit einer zusammengesetzten Sepsisdefinition abgeglichen; kein randomisierter Vergleich. AUROC auf Aufenthaltsebene: 0,63 (95%-KI 0,62–0,64). Bei Schwelle ≥ 6 blieben 67 % der Sepsisfälle unerkannt; 18 % der Aufenthalte hätten einen Alarm ausgelöst. Autorenförderung: NIH/NHLBI; ein Autor erklärte redaktionelle Vergütungen außerhalb der Studie. Der zugängliche Artikel enthält eine Korrektur von 2021 zur Zahl notwendiger Überprüfungen.[5]

Eigene Bewertung – fünf Domänen. Auswahl: ein großes, aber einzelnes Zentrum. Referenzstandard: die zusammengesetzte Falldefinition ist nicht unehlbar. Zeitbezug: Vorhersagen vor und nach Krankheitsbeginn müssen getrennt bleiben. Schwellenwahl: Sensitivität und Alarmbelastung sind gemeinsam relevant. Aktualität: historische Modellversion und damalige Versorgung, keine Aussage über jede spätere Generation.

Zulässige Schlussfolgerung/offene Prüfung. Der Befund begründet lokale externe Prüfungen und die Skepsis gegenüber isolierten Herstellerkennzahlen. Er misst keine tatsächlich verursachten Patientenschäden. Vor einer heutigen Entscheidung sind aktuelle Version, Kalibrierung, Vorwarnzeit, Alarmmanagement und klinischer Zusatznutzen neu zu untersuchen.

[6] Wong et al. Epic-Sepsismodell Version 2, 2026

Datenauszug. Prospektive Validierung an vier US-Gesundheitssystemen: 227.091 Aufenthalte. Vergleich mit Sepsis-3-Kriterien und klinischen Erkennungszeitpunkten. Alle Standorte hatten lokal nachtrainiert. Aufenthaltsbezogene AUROC: 0,82 (95%-KI 0,81–0,83) bis 0,92 (0,92–0,93). Bei lokalen Schwellen für 60 % Sensitivität: positiver Vorhersagewert 13 % (13–14 %) bis 26 % (25–27 %). Universitäre/öffentliche Autorenförderung; unter anderem Google-Beratungshonorare eines Autors außerhalb der Studie offengelegt.[6]

Eigene Bewertung – fünf Domänen. Auswahl: mehrere Versorgungstypen erhöhen die Breite. Referenzstandard: behandlungsabhängige Sepsisdefinition kann übersehene unbehandelte Fälle verfehlen. Rückkopplung: bereits aktive Alarmer können die zur Ergebnisdefinition verwendeten Anordnungen beeinflussen. Modelltransport: lokale Anpassung verhindert Aussagen über das unveränderte Basismodell. Endpunkt: prognostische Güte ist kein nachgewiesener Behandlungseffekt.

Zulässige Schlussfolgerung/offene Prüfung. Die aktualisierten lokalen Modelle erreichen bessere Diskrimination; dennoch bleibt erheblicher Fehlalarmaufwand. Kliniken müssen passende Zeithorizonte, Kalibrierung, Untergruppen, Reaktionskapazität und patientenrelevante Folgen prüfen. Hohe AUROC allein rechtfertigt keine Freigabe; der

Vergleich mit 2021 ist zudem kein randomisierter Generationenvergleich.

[7] Lukac et al. KI-Dokumentationsassistenten, 2025

Datenauszug. Pragmatische Dreiarmsstudie: 238 ambulante Ärzte, 14 Fachrichtungen, zwei Monate; DAX 79, Nabla 79, Standarddokumentation 80. Kovariatenbalancierte Randomisierung. Primär: Veränderung der Dokumentationsfensterzeit. Gegen Kontrolle: Nabla -9,5 % (95%-KI -17,2 bis -1,8 %), DAX -1,7 % (-9,4 bis +5,9 %). Förderung durch UCLA-Einheiten und NIH-Autorenförderung aus indextierten Originalabschnitten gelesen; separate Interessenkonfliktformulare nicht geprüft. Volltext nur auszugsweise über den Suchindex zugänglich.[7]

Eigene Bewertung – fünf Domänen. Zuteilung: angemessener aktiver Produktvergleich mit zeitgleicher Kontrolle. Messung: Dokumentationsfensterzeit erfasst nicht gesamte Arbeit oder Beziehungsqualität. Umsetzung: freiwillige Nutzung ist praxisnah, erlaubt aber keine Gleichsetzung von Zugang und regelmäßiger Anwendung. Sicherheitsdaten: subjektiv berichtete Ungenauigkeiten ersetzen kein unabhängiges Fehleraudit. Übertragbarkeit: kurzer ambulanter Einsatz, kein Nachweis für deutsche Stationen oder Pflege.

Zulässige Schlussfolgerung/offene Prüfung. Ein Produkt zeigte begrenzte Entlastung beim gemessenen Zeitmaß; universelle Effizienz- oder Sicherheitsbehauptungen sind nicht gedeckt. Offen bleiben Langzeitwirkung, Gesamtkorrekturaufwand, relevante Auslassungen, Verständlichkeit für Patienten sowie Register- und Protokollabgleich.

[8] Dratsch et al. Automationsgläubigkeit, 2023

Datenauszug. Experiment mit 27 Radiologen und 50 Mammographien; ein Fall wurde aus der Auswertung ausgeschlossen. Nach zehn korrekten Einstiegsempfehlungen enthielt der zweite Block zwölf absichtlich falsche BI-RADS-Vorschläge eines vermeintlichen KI-Systems. Bei den fünf sehr erfahrenen Befundern waren 82,3 % der Einstufungen bei korrekten gegenüber 45,5 % bei falschen Vorschlägen richtig (SD 4,2 beziehungsweise 9,1; $p=0,003$). Förderung durch das Bundesgesundheitsministerium; unter anderem Industriebonorare, Beratungsbeziehungen und ein Patent offengelegt.[8]

Eigene Bewertung – fünf Domänen. Manipulation: gezielt falsche Empfehlungen prüfen Fehleranfälligkeit, nicht Produktgüte. Vergleich: keine unbeeinflusste Kontrollbefundung derselben Fälle. Fallauswahl: angereicherte Untersuchungsfälle statt Screeningprävalenz. Präzision: besonders die erfahrene Untergruppe ist klein; SD ist kein Konfidenzintervall. Kontext: vorbereiteter Vertrauensaufbau und künstliche Oberfläche begrenzen Alltagstransfer.

Zulässige Schlussfolgerung/offene Prüfung. Erfahrung allein garantiert keine wirksame Korrektur falscher Empfehlungen. Für die Beschaffung ist ein Mensch-KI-Test mit dem tatsächlichen Produkt sinnvoll: realistische Fehler, unabhängige Erstbewertung, Arbeitslast und sichere Übersteuerung. Aus diesem Experiment lässt sich keine allgemeine klinische Schadensrate berechnen.

[9] Obermeyer et al. verzerrte Bedarfssteuerung, 2019

Datenauszug. Beobachtendes Audit eines eingesetzten US-Versorgungsalgorithmus: 43.539 weiße und 6.079 Schwarze Patienten. Erwartete Ausgaben fungierten als Zielgröße. Bei der Schwelle zur automatischen Programmauswahl hatten Schwarze Patienten durchschnittlich 4,8 gegenüber 3,8 chronische Erkrankungen bei weißen Patienten, ein Unterschied von 26,3 % ($p<0,001$; hier kein verifiziertes Konfidenzintervall). Verglichen wurden Bedarf und Risikowerte zwischen Gruppen, nicht randomisierte Behandlungen. Förderung: National Institute for Health Care Management Foundation. Die Autoren erklärten Analyseunabhängigkeit und keine Vergütung oder wirtschaftlichen Interessen am Hersteller.[9]

Eigene Bewertung – fünf Domänen. Zielvalidität: Kosten sind nicht mit Bedarf identisch. Vergleichbarkeit: ein Audit kann Mechanismen plausibilisieren, aber nicht alle Zugangsunterschiede kausal isolieren. Bedarfsmaß: dokumentierte Erkrankungen sind ebenfalls versorgungsabhängig. Reproduzierbarkeit: geschützte Originaldaten verhindern freien personenbezogenen Nachvollzug. Übertragbarkeit: ein US-Kontext begründet keine deutsche Effektgröße.

Zulässige Schlussfolgerung/offene Prüfung. Ein technisch gut vorhergesagter Stellvertreter kann eine ethisch ungeeignete Steuerungsgröße sein. Lokale Prüfungen müssen Zieldefinition, gruppenspezifische Messqualität und

tatsächliche Versorgungsfolgen untersuchen; eine simulierte Umverteilung wäre noch kein nachgewiesener Gesundheitsgewinn.

[10] Brynjolfsson, Li und Raymond Kundendienst, 2025

Datenauszug. Feldanalyse der gestaffelten Einführung eines GPT-3-basierten Assistenten bei 5.172 Beschäftigten eines Unternehmens. Difference-in-Differences-Vergleich mit noch nicht unterstützten Beschäftigten; bevorzugtes Modell mit Personen-, Zeit- und Erfahrungskontrollen. Endpunkt: gelöste Anliegen pro Stunde. Effekt +0,301 (robuster Standardfehler 0,0329), entsprechend rund 15,2 % gegenüber dem Ausgangswert. Förderung: Stanford Digital Economy Lab. Eine gesonderte Interessenkonflikterklärung wurde hier nicht verifiziert.[10]

Eigene Bewertung – fünf Domänen. Zuteilung: der Haupteffekt stammt nicht aus vollständig randomisierter Einführung. Identifikation: parallele Trends und fehlende unbeobachtete zeitgleiche Veränderungen bleiben Annahmen. Messung: gelöste Anliegen berücksichtigen Output, aber nicht sämtliche Qualitätsdimensionen. Heterogenität: Mittelwerte sind keine individuelle Nutzenprognose. Übertragbarkeit: einzelnes Unternehmen, konkrete Aufgaben und ältere Modellgeneration.

Zulässige Schlussfolgerung/offene Prüfung. Plausible reale Produktivitätsgewinne rechtfertigen aufgabenbezogene Erprobung, keine allgemeine Personalabbauprognose. Offen bleiben lokale Qualitätsverteilung, Arbeitsintensität, Kompetenzentwicklung und wirtschaftlicher Nettonutzen einschließlich Prüfaufwand. Der Standardfehler bezeichnet Schätzunsicherheit und darf nicht als Streuung individueller Produktivitätsgewinne gelesen werden.

[11] Dell’Acqua et al. Aufgabenabhängigkeit, 2026

Datenauszug. Präregistriertes Experiment mit 758 Unternehmensberatern: kein KI-Zugang, GPT-4 oder GPT-4 mit Einführung. Bei geeigneten Aufgaben: 12,2 % mehr erledigte Teilaufgaben; 25,1 % kürzere Zeit bis Frage 18, nicht für den gesamten Aufgabenblock. In einer ungeeigneten Managementaufgabe: richtige Lösungen 84,5 % ohne KI, 60 % beziehungsweise 70,6 % mit KI. Unadjustierte Effekte –24,5 beziehungsweise –13,9 Prozentpunkte; robuste Standardfehler 5,4 beziehungsweise 5,3. Harvard-Forschungsförderung; drei Mitautoren waren beim untersuchten Beratungsunternehmen tätig.[11]

Eigene Bewertung – fünf Domänen. Zuteilung: Randomisierung stärkt interne Vergleichbarkeit. Aufgabenselektion: bewusst innerhalb/außerhalb der damaligen Fähigkeiten konstruierte Fälle sind keine repräsentative Tätigkeitsstichprobe. Messung: Richtigkeit und Geschwindigkeit dürfen nicht verrechnet werden. Unsicherheit: Standardfehler gelten für definierte Aufgaben, nicht alle Wissensarbeit. Übertragbarkeit: selektierte Beraterpopulation und konkrete Modellversion.

Zulässige Schlussfolgerung/offene Prüfung. Unterstützung kann abhängig von der Aufgabe helfen oder schaden. Freigaben sollten deshalb Tätigkeiten und Fehlertoleranzen benennen. Lokale Prüfung muss insbesondere scheinbar plausible Fehlantworten sowie veränderte Aufgaben und Modellversionen einschließen.

[41] Agweyu et al. Entscheidungsassistenz in Kenia, 2026

Datenauszug. Cluster-randomisierte Studie: 103 Clinical Officers, nichtärztliche klinische Fachkräfte, an 16 kenianischen Einrichtungen; 9.347 Behandlungskontakte in der Hauptanalyse. Elektronische Patientenakte mit GPT-4o-Assistenz gegen dieselbe Akte ohne Assistenz. Primärer kombinierter 14-Tage-Endpunkt Behandlungsversagen, einschließlich Wiedervorstellung wegen fortbestehender Beschwerden, ungeplanter Eskalation oder unerwünschter Ereignisse: 102/4.693 (2,2 %) gegenüber 94/4.654 (2,0 %); adjustierte Odds Ratio 0,77 (95%-KI 0,55–1,08; $p=0,13$). Sekundär verbesserte Dokumentationsqualität. OpenAI gewährte Rechengutschriften und technische Produktentwicklungsberatung, laut Autoren ohne Beteiligung an Studiendesign, Analyse oder Veröffentlichung. Robert Korom und Sarah Kiptinness hielten Penda-Aktionsoptionen.[41]

Eigene Bewertung – fünf Domänen. Zuteilung: Randomisierung auf Behandlerenebene, mögliche Kontamination zwischen Kollegen. Messung: verblindete Endpunktbeurteilung stärkt Aussagekraft. Ausfälle: Ausschlüsse und Umgang mit Protokollabweichungen bedürfen weiterer Klärung. Schätzung: rohe Raten und adjustierte Odds Ratio sind unterschiedliche Größen. Übertragbarkeit: einzelnes privates Kliniknetz und kurze Nachbeobachtung begrenzen den Transfer.

Zulässige Schlussfolgerung/offene Prüfung. Kein statistisch gesicherter Nutzen beim Primärendpunkt; bessere Dokumentation ist kein Ersatznachweis. Seltene Schäden sind nicht ausgeschlossen. Vor Übernahme sind Analysepopulation, lokale Versorgung und längerfristige Sicherheit zu prüfen.

4.3 Synthese der offenen Nachweisaufgaben

Unsere Synthese unterscheidet vier Ebenen, die in einer Beschaffungsakte getrennt bleiben sollten. Erstens: Kann das Modell die definierte Aufgabe in der Zielpopulation bewältigen? Zweitens: Können Beschäftigte mit dem konkreten System besser handeln als mit der bestehenden Praxis? Drittens: Verbessert der gesamte Ablauf ein relevantes Ergebnis, ohne Belastungen unbemerkt zu verschieben? Viertens: Bleibt das Verhältnis zwischen Nutzen, Risiken und Rechten bei Updates, Schichtwechseln und anderen Bedingungen tragfähig?

Diese Reihenfolge ist ein eigener Prüfraum, kein validiertes Messinstrument. Ein positives Ergebnis auf einer Ebene beantwortet nicht automatisch die nächste. Umgekehrt macht fehlende Mortalitäts-evidenz einen Schreibassistenten nicht grundsätzlich unbrauchbar: Die erforderliche Evidenz muss zur behaupteten Wirkung und zum Schadenspotenzial passen. Entscheidend ist, dass die Einrichtung ihre Freigabe nicht auf einen weiterreichenden Nutzen stützt, als tatsächlich untersucht wurde.

5 Risiken und institutionelle Wirkmechanismen

5.1 Überzeugende Fehler und unbemerkte Auslassungen

Generative Systeme können sachlich falsche, unbelegte oder widersprüchliche Antworten formulieren. Ein sprachlich überzeugender Text ist kein Wahrheitsnachweis. Auch ausgelassene Informationen können entscheidend sein: etwa eine Kontraindikation in einer Zusammenfassung oder eine Ausnahme in einer internen Regel. Die WHO behandelt solche Risiken ausdrücklich in ihrer Orientierung zu großen multimodalen Modellen.[2]

Unsere Empfehlung lautet: In folgenreichen Abläufen müssen kritische Aussagen anhand geeigneter Originalinformationen überprüfbar sein. Eine Quellenangabe allein genügt nicht; die Quelle muss existieren, zur Aussage passen und den jeweiligen Fall tragen. Ein System mit Dokumentensuche, sogenanntem Retrieval-Augmented Generation oder RAG, kann die Wissensgrundlage verbessern. Es beseitigt Fehler aber nicht automatisch. Die Datenschutzkonferenz behandelt zudem spezifische Zugriffs- und Datenschutzfragen solcher Architekturen.[19]

5.2 Diskriminierung ist mehr als ein unvollständiger Datensatz

Prüfgegenstand sind Daten, Zielvariable, Entscheidungsschwelle und Folgen. Gleiche durchschnittliche Genauigkeit schließt unterschiedliche Fehlerraten oder Belastungen nicht aus. Besonders schwer sichtbar bleiben kleine oder in den Daten kaum erfasste Gruppen. Die STANDING-Together-Empfehlungen richten den Blick auf Datenzusammensetzung, Grenzen der Repräsentation und die Prüfung von Benachteiligungen.[30]

Als Umsetzung schlagen wir vor: Relevante Gruppen vorab begründen, Leistungs- und Ergebnisunterschiede mit Unsicherheiten ausweisen und Betroffene an der Auswahl der Kriterien beteiligen. Fehlende Daten sind dabei eine Erkenntnisgrenze, kein Beweis gleicher Behandlung. Die Erhebung sensibler Merkmale zur Fairnessprüfung benötigt ihrerseits eine rechtlich tragfähige Gestaltung. Gerechtigkeit lässt sich außerdem nicht auf eine universell richtige Kennzahl reduzieren; Konflikte zwischen Kriterien müssen begründet entschieden werden.

5.3 Datenschutz und Sicherheit sind eigenständige Prüfböcke

Ein nützliches System kann unzulässig auf Daten zugreifen. Ein zulässiger Datenfluss kann technisch unsicher sein. Bei generativen Assistenzsystemen kommen Risiken durch manipulierte Eingaben oder Dokumente hinzu: Fremde Inhalte können als Anweisungen missverstanden werden. Besonders folgenreich wird dies, wenn ein Assistent zusätzlich Nachrichten versenden, Daten ändern oder externe Werkzeuge bedienen darf.[17,34]

Unsere Empfehlungen sind abgestufte Rechte, eine Trennung von Informationszugriff und Handlungsvollmacht, überprüfbare Datenflüsse und zusätzliche Freigaben bei folgenreichen Aktionen. Vertrauliche Daten gehören nicht allein deshalb in ein System, weil dessen Oberfläche bequem erreichbar ist. Umgekehrt macht ein Betrieb im eigenen Rechenzentrum eine Verarbeitung nicht automatisch rechtmäßig. Zweck, Berechtigungen und Rechtsgrundlage

bleiben zu prüfen.[18,19,22]

5.4 Beziehung, Kompetenz und Arbeitsbedingungen können unter Druck geraten

Der ICN-Ethikkodex fordert für KI und entsprechende Geräte ausdrücklich personenzentrierte Pflege und den Erhalt menschlicher Beziehungen.[12] Daraus folgt nicht, dass jede technische Interaktion entmenslicht. Eine gut gestaltete Assistenz kann Zeit für Gespräche eröffnen oder Selbstständigkeit stärken. Ob das tatsächlich geschieht, hängt von Einsatz und Organisation ab.

Als prüfbare Umsetzung empfehlen wir, erreichbare menschliche Ansprechpartner, erlebte Selbstbestimmung und unerfüllten Unterstützungsbedarf zu erfassen. Ein Gesprächsroboter sollte nicht verschleiern, dass er ein technisches System ist. Bei verletzlichen Personen sind Verständnis, Freiwilligkeit und Ausweichmöglichkeiten besonders sorgfältig zu betrachten. Befürchtungen zu Kompetenzverlust oder dauerhafter Beziehungssubstitution sind ernst zu nehmen; die hier ausgewählten Studien quantifizieren deren langfristiges Ausmaß jedoch nicht.

Für körpernah eingesetzte Robotik empfehlen wir zusätzlich eine eigenständige Prüfung physischer Gefährdungen und der sicheren Übergabe an Menschen: etwa bei Kollision, unerwarteter Bewegung, Ausfall oder einer nicht verstandenen Aufforderung. Je nach Aufgabe gehören Hygiene, Zugänglichkeit und die Möglichkeit, sich einer Interaktion zu entziehen, dazu. Die in diesem Beitrag ausgewählten Softwarestudien liefern keinen Wirksamkeits- oder Sicherheitsnachweis für einen Pflegeroboter. Seine Bewertung benötigt zusätzliche produktspezifische Evidenz und Fachkompetenz.

5.5 Abhängigkeit, Konzentration und Ressourcenverbrauch gehören zur Ethikprüfung

Die Bindung an wenige Anbieter kann Wechselkosten und institutionelle Abhängigkeit schaffen. Ein späterer Modellwechsel kann Leistung, Datenverarbeitung und Verfügbarkeit verändern. Auch Energie- und Infrastrukturbedarf sind relevant. Die IEA untersucht hierfür die Entwicklung von Rechenzentren und KI; Prognosen sind jedoch keine bereits eingetretenen Verbräuche, und Rechenzentrumsstrom darf nicht vollständig der KI zugerechnet werden.[33]

Unsere Empfehlung ist eine aufgabenbezogene Ressourcenbilanz: Welcher Nutzen rechtfertigt welchen Rechen-, Prüf- und Personalaufwand? Ist ein kleineres Modell oder eine konventionelle Lösung ausreichend? Anbieterwechsel und Ausfall sollten vor der Beschaffung praktisch durchgespielt werden. Dabei sind wirtschaftliche Effizienz, Versorgungssicherheit und Nachhaltigkeit gemeinsam zu betrachten, nicht gegeneinander auszuspielen.

5.6 Schwerwiegende Zukunftsrisiken weder verschweigen noch als Gewissheit verkaufen

Der International AI Safety Report 2026 unterscheidet Risiken durch schädliche Nutzung, Fehlfunktionen und systemische Entwicklungen. Er behandelt auch mögliche Kontrollverluste bei leistungsfähigeren allgemeinen Systemen. Solche Szenarien sind nicht mit der heute belegten Fehlerquote eines klinischen Assistenten gleichzusetzen. Eintrittswahrscheinlichkeiten und Übertragbarkeit experimenteller Befunde bleiben vielfach unsicher.[32]

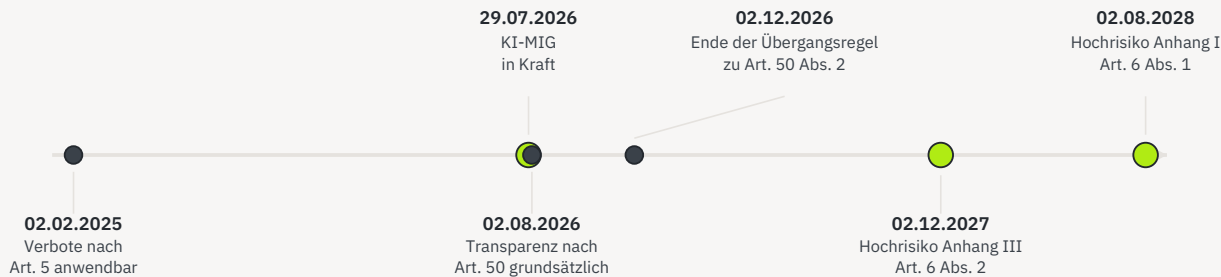
Für Einrichtungen empfehlen wir eine proportionale Konsequenz: Mit wachsender Autonomie, Systemanbindung und möglichem Schadensumfang steigen die Anforderungen an Begrenzung, unabhängige Prüfung und Rückholbarkeit. Für Gesellschaft und Politik sind gemeinsame Sicherheitsforschung, Vorfallauswertung und grenzüberschreitende Regeln sinnvoll. Eine einzelne Ethikbeauftragte kann diese Aufgaben nicht stellvertretend für Staaten oder Hersteller erfüllen.

6 Rechtliche Voraussetzungen und Grenzen

Dieser Abschnitt betrifft Deutschland und die EU. Er ist ein Einstieg in die Prüfung, keine vollständige Behandlung aller Haftungs-, Berufs-, Sozial-, Forschungs- und sektoralen Vorschriften. Für Konzernstandorte außerhalb der EU sind weitere Rechtsordnungen zu berücksichtigen.

6.1 Der KI-Rechtsrahmen hat mehrere Zeitachsen

Abbildung 2 – Geltungstermine des KI-Rechtsrahmens



Termine aus Abschnitt 6.1 und 6.4 dieser Studie. Die Marken sitzen im Verhältnis ihrer tatsächlichen Abstände; wo Termine dicht beieinanderliegen, zeigt eine Führungslinie auf den wahren Ort. Rechtsgrundlagen: Verordnung (EU) 2024/1689 in der durch Verordnung (EU) 2026/1744 geänderten Fassung, Art. 113 Buchst. c; KI-MIG vom 22.07.2026.

Der AI Act, Verordnung (EU) 2024/1689, wurde durch die Verordnung (EU) 2026/1744 geändert. Dieser „AI-Omnibus“ ist seit 27. Juli 2026 in Kraft. Ein älterer Fristenplan darf deshalb nicht unverändert verwendet werden.[20,21]

Tabelle 4 Anwendungsfristen im KI Rechtsrahmen

Regelungsbereich	Stand am 24. September 2026	Konsequenz für die Einrichtung
Bereits anwendbare Verbote nach Art. 5	Die seit Februar 2025 anwendbaren Verbote gelten weiter; neu eingefügte Verbote besitzen teilweise eigene spätere Termine.	Unzulässige Funktionen nicht als gewöhnliches Pilotprojekt behandeln.
KI-Kompetenz, Art. 4	Die geänderte Vorschrift verpflichtet Anbieter und Betreiber zu Maßnahmen zur Förderung der Kompetenzentwicklung, nicht zur Garantie eines bestimmten individuellen Kompetenzniveaus.	Auf Aufgaben und Risiken bezogene Qualifizierung organisieren; kein gesetzlich vorgeschriebenes allgemeines „KI-Führerschein“-Zertifikat behaupten.
Transparenz, Art. 50	Grundsätzlich seit 2. August 2026 anwendbar. Für Anbieter bestimmter vor diesem Datum vermarkteter generativer Systeme besteht eine begrenzte Übergangsregel zu Art. 50 Abs. 2 bis 2. Dezember 2026.	Nicht sämtliche Transparenzpflichten für verschoben halten; Adressat und konkrete Funktion prüfen.
Hochrisiko nach Art. 6 Abs. 2 / Anhang III	Kapitel III Abschnitte 1–3, mit Ausnahme von Art. 6 Abs. 5, ab 2. Dezember 2027.	Einstufung und Umsetzungsplanung auf den aktuellen Rechtsakt stützen.
Produktbezogenes Hochrisiko nach Art. 6 Abs. 1 / Anhang I	Die entsprechenden Vorschriften ab 2. August 2028.	Medizinprodukte- und andere Produktpflichten gelten unabhängig davon weiter.

Ein Krankenhaus ist nicht als Ganzes „Hochrisiko-KI“. Zu prüfen sind Zweck und Funktion des einzelnen Systems sowie die Voraussetzungen und Ausnahmen der jeweiligen Einstufung. Bei produktbezogener KI ist insbesondere die Verbindung mit den einschlägigen Produktvorschriften und einer erforderlichen Drittstellen-Konformitätsbewertung relevant. Der bloße Einsatz in einer Klinik macht allgemeine Software noch nicht zum Medizinprodukt.[20,21,23]

Vor der Bestimmung der Pflichten steht außerdem die Rollenprüfung: Eine Einrichtung kann Systeme nur betreiben, aber auch selbst entwickeln oder so verändern, dass zusätzliche Anbieterpflichten relevant werden. Eigenentwicklungen, Integration und wesentliche Veränderungen sind deshalb ausdrücklich zu erfassen; insbesondere Art. 25 ist im einschlägigen Anwendungsbereich und zum jeweiligen Anwendungszeitpunkt zu berücksichtigen.[20,21]

Für einschlägige Hochrisikosysteme verlangt Art. 26 nach dem jeweiligen Anwendungsbeginn unter anderem bestimmungsgemäße Nutzung, geeignete Aufsicht und Überwachung. Aufsichtspersonen benötigen Kompetenz, Schulung, Befugnisse und Unterstützung. Die organisatorische Empfehlung dieses Beitrags lautet, entsprechende Voraussetzungen bereits bei heutigen folgenreichen Einsätzen herzustellen. Damit wird die spätere AI-Act-Pflicht

nicht fälschlich als heute schon generell anwendbar ausgegeben.[20,21]

Die nach dem einschlägigen Anwendungsbeginn zu erfüllende Pflicht zur Grundrechte-Folgenabschätzung nach Art. 27 trifft nicht unterschiedslos jeden Konzern oder jedes medizinische KI-Produkt. Sie betrifft bestimmte Betreiber bestimmter Anhang-III-Systeme, unter anderem öffentlich-rechtliche Stellen und private Erbringer öffentlicher Dienste; gesetzliche Ausnahmen und besondere Finanzanwendungen sind zu beachten. Sie ist von der Datenschutz-Folgenabschätzung zu unterscheiden, auch wenn passende Bestandteile übernommen oder miteinander verknüpft werden können.[20,21]

6.2 Datenschutz und Arbeitsrecht warten nicht auf spätere Hochrisikofristen

Die DSGVO verlangt eine tragfähige Grundlage für personenbezogene Verarbeitung. Bei Gesundheitsdaten ist zusätzlich eine einschlägige Ausnahme nach Art. 9 erforderlich. Ein Auftragsverarbeitungsvertrag ist nicht selbst die Rechtsgrundlage. Bei voraussichtlich hohem Risiko ist nach Art. 35 eine Datenschutz-Folgenabschätzung erforderlich; nicht jede Verwendung des Etiketts „KI“ löst sie automatisch aus.[22]

Art. 22 betrifft ausschließlich automatisierte Entscheidungen mit rechtlicher oder ähnlich erheblicher Wirkung, vorbehaltlich seiner Ausnahmen und Schutzvorkehrungen. Nicht jede Assistenz fällt darunter. Eine nur formale menschliche Bestätigung darf jedoch nicht als substantielle Prüfung ausgegeben werden.[18,22]

Beschäftigtenvertretungen sind frühzeitig einzubeziehen, wo die einschlägigen Beteiligungsrechte greifen. § 90 BetrVG erfasst ausdrücklich die Planung von Arbeitsverfahren einschließlich KI; § 87 Abs. 1 Nr. 6 betrifft technische Einrichtungen zur Verhaltens- oder Leistungsüberwachung. Öffentliche und kirchliche Krankenhausträger unterliegen teilweise anderen Vertretungsregeln. Daneben bleibt das Antidiskriminierungsrecht relevant.[24,25]

Eine besonders deutliche rote Linie ist die grundsätzlich verbotene KI-gestützte Ableitung von Emotionen auf Grundlage biometrischer Daten am Arbeitsplatz nach Art. 5 Abs. 1 Buchst. f AI Act. Die vorgesehenen medizinischen oder sicherheitsbezogenen Ausnahmen sind zweckbezogen; der Standort „Krankenhaus“ schafft keine pauschale Ausnahme für Personalbewertung.[20]

6.3 Produktkonformität, fachlicher Standard und Ethik sind nicht dasselbe

Bei medizinischer Software sind Zweckbestimmung und mögliche Folgen wesentlich. Regel 11 der MDR differenziert nach Funktion und Risiko; „medizinische KI ist immer Klasse IIa“ wäre falsch. Konformität im jeweiligen Geltungsbereich ersetzt nicht die Prüfung, ob das System für die konkrete lokale Verwendung geeignet ist. Diese zweite Aussage ist eine fachlich-organisatorische Folgerung, keine Behauptung, eine CE-Kennzeichnung sei bedeutungslos.[23,36]

§ 630a Abs. 2 BGB knüpft medizinische Behandlung grundsätzlich an die allgemein anerkannten fachlichen Standards zum Behandlungszeitpunkt an. Eine Ethikempfehlung wird dadurch nicht automatisch zur verbindlichen Rechtsnorm. Die Frage, ob und wie sie für einen fachlichen Standard relevant ist, muss gesondert betrachtet werden.[27]

Auch WHO-, AEM- und NIST-Dokumente sind keine deutschen Gesetze. ISO/IEC 42001 betrifft ein KI-Managementsystem. Ein entsprechendes Zertifikat belegt nicht automatisch die klinische Wirksamkeit eines Produkts oder die Rechtmäßigkeit jedes Einsatzes. Dafür sind Zertifikatsumfang und anwendungsspezifische Nachweise zu prüfen.[31]

6.4 Zuständigkeiten nicht verwechseln

Das deutsche Gesetz zur Marktüberwachung und Innovationsförderung von künstlicher Intelligenz (KI-MIG) gilt seit 29. Juli 2026. Die Bundesnetzagentur besitzt zentrale und auffangende Aufgaben; daneben bleiben sektorale und länderspezifische Zuständigkeiten relevant. § 6 regelt eine zentrale Anlaufstelle. Daraus folgt weder, dass die Bundesnetzagentur jedes klinische System allein beaufsichtigt, noch dass jede Einrichtung dort eine verbindliche kostenlose Einzelfallfreigabe erhält.[26]

Für Medizinprodukte bietet das BfArM Informationen zu Qualifikation, Klassifizierung und Vorkommnismeldungen. Datenschutzfragen gehören zur jeweils zuständigen Datenschutzaufsicht. Eine Beratung durch das klinische

Ethikkomitee ersetzt keinen dieser Wege.[18,26,35]

6.5 Einsatz und rechtliche Rollen

„Wir verwenden Produkt X“ genügt als Beschreibung für die Rechtsabteilung nicht. Dieselbe Software kann Dokumente formatieren, Patientenrisiken bewerten oder Beschäftigte vorsortieren. Deshalb sollte jede Freigabe eine bestimmte Kombination aus Zweck, Funktion, Datenfluss, Version, Einrichtung und betroffener Personengruppe betreffen. Das ist eine Governance-Empfehlung zur nachvollziehbaren Umsetzung verschiedener Rechtspflichten, keine gesetzlich vorgeschriebene Einheitsvorlage.

Der erste Prüfschritt lautet: Welche Entscheidung wird vorbereitet oder getroffen, mit welcher tatsächlichen Wirkung? Der zweite: Welche juristische Person bestimmt den Einsatz und welche Unternehmen entwickeln, betreiben oder verändern das System? Der dritte: Welche Norm ist bereits anwendbar, welche erst künftig und welche sektorale Sonderregel greift? Ein Ethikvotum beantwortet diese Fragen nicht stellvertretend. Ebenso wenig ersetzt eine CE-Kennzeichnung die Prüfung des konkreten Datenschutzes, der Beschäftigtenrechte oder der bestimmungsgemäßen Verwendung. [20–24,27,36]

Die Rollenprüfung muss mindestens drei getrennte Ebenen umfassen:

Tabelle 5 Getrennte rechtliche Rollen

Ebene	Maßgebliche Frage	Typischer Fehler
KI-Verordnung	Wer entwickelt beziehungsweise lässt entwickeln und stellt unter eigenem Namen bereit; wer verwendet das System in eigener Verantwortung?	Ein Krankenhaus oder Konzern sei immer nur Betreiber.
Medizinproduktrecht	Welche medizinische Zweckbestimmung hat die Software; wer stellt sie her oder verändert ihren Zweck?	Jede Krankenhaussoftware sei ein Medizinprodukt – oder jede Eigenentwicklung sei ausgenommen.
Datenschutz	Wer entscheidet tatsächlich über Zwecke und wesentliche Mittel jeder Verarbeitung?	„Auftragsverarbeiter“ im Vertrag entscheide die Rolle abschließend.

Diese Rollen sind nicht deckungsgleich. Ein externer KI-Anbieter kann für die Kundendatenverarbeitung Auftragsverarbeiter sein. Für eine eigene Nutzung derselben Daten zu Trainingszwecken ist seine Rolle jedoch gesondert zu prüfen. Vertragsbezeichnungen können tatsächliche Verantwortlichkeit nicht beliebig umverteilen. Die EDSA-Leitlinien 07/2020 erläutern diese funktionale Zuordnung; sie sind behördliche Auslegungshilfe, kein eigenständiges Gesetz. [20,22,23,42]

6.6 Rollenwechsel und Änderungen

Ein Unternehmen kann schon nach Art. 3 KI-Verordnung Anbieter sein, wenn es ein System entwickeln lässt und unter eigenem Namen in Betrieb nimmt. Für Hochrisikosysteme beschreibt Art. 25 zusätzliche Rollenwechsel: insbesondere eigene Kennzeichnung, wesentliche Veränderung eines weiterhin hochriskanten Systems oder eine Zweckänderung, durch die ein bislang nicht hochriskantes System hochriskant wird. Die einschlägigen Hochrisikopflichten sind dabei nach dem jeweiligen Anwendungsdatum zu unterscheiden. Nicht jede neue Eingabe, Promptänderung oder lokale Konfiguration ist automatisch eine wesentliche Veränderung. Entscheidend sind die gesetzlichen Kriterien, die ursprünglich bewertete Zweckbestimmung und die konkrete Auswirkung. [20,21]

Praktisch sollte der Änderungsantrag deshalb beantworten: Bleibt die Aufgabe gleich? Kommen neue Betroffenengruppen hinzu? Werden Ausgaben erstmals unmittelbar handlungsleitend? Ändern sich Trainingsdaten, Modell, Schnittstellen oder menschliche Kontrolle? Die Empfehlung lautet, eine erneute Prüfung an solche Änderungen zu knüpfen, nicht ausschließlich an große Versionsnummern. Eine dokumentierte Ablehnung der Neueinstufung ist aussagekräftiger als die ungeprüfte Herstellerformel „nur ein Update“.

Bei Medizinprodukten kommt eine eigene Bewertung hinzu. Insbesondere die In-house-Regel des Art. 5 Abs. 5 MDR setzt mehrere Bedingungen voraus; „nur für den Eigenbedarf“ reicht nicht. Dazu gehören ein geeignetes Qualitätsmanagement, die dokumentierte spezifische Versorgungslücke gegenüber gleichwertigen Marktprodukten,

Dokumentation und öffentliche Erklärung sowie die Begrenzung auf dieselbe rechtlich eigenständige Einrichtung. Eine konzerninterne Weitergabe an eine andere Gesellschaft ist deshalb nicht ohne Weiteres privilegiert. Die einschlägige MDCG-Leitlinie ist als unverbindliche Auslegungshilfe zu kennzeichnen. [23,36,45]

6.7 Verträge und tatsächliche Verantwortlichkeit

Liegt Auftragsverarbeitung vor, muss sie durch einen Vertrag oder ein anderes bindendes Rechtsinstrument nach Art. 28 Abs. 3 DSGVO geregelt sein. Diese Regelung ersetzt nicht die erforderliche Rechtsgrundlage der Verarbeitung. Bei Gesundheitsdaten sind insbesondere Art. 6 und die einschlägige Ausnahme des Art. 9 DSGVO gemeinsam zu prüfen; Behandlung, Qualitätssicherung, Forschung und Modelltraining sind nicht automatisch derselbe Zweck. Gemeinsame Festlegung von Zwecken und Mitteln kann statt Auftragsverarbeitung eine Vereinbarung nach Art. 26 DSGVO verlangen. [22,42]

Die Regelung muss bei Auftragsverarbeitung insbesondere folgende gesetzliche Inhalte enthalten: dokumentierte Weisungen, Vertraulichkeit, Sicherheitsmaßnahmen, Unterauftragsverarbeiter, Unterstützung bei Betroffenenrechten und Datenschutzpflichten, Rückgabe beziehungsweise Löschung sowie Nachweis- und Kontrollmöglichkeiten. Eine abstrakte Zusage „DSGVO-konform“ ersetzt diese Inhalte nicht. Für internationale Datenflüsse sind zusätzlich Art. 44 ff. zu prüfen. Dabei sollten Hosting, Fernwartung, Supportzugriff und Unterauftragnehmer tatsächlich kartiert werden; ein europäischer Serverstandort beantwortet nicht alle Übermittlungsfragen. [22]

Darüber hinaus sind als vertragliche Governance-Empfehlungen sinnvoll: genaue Zweck- und Versionsbeschreibung, bekannte Leistungsgrenzen, Meldewege bei sicherheitsrelevanten Fehlern, Vorankündigung erheblicher Änderungen, Testmöglichkeiten, ein gesicherter Datenexport und eine praktisch nutzbare Beendigung des Einsatzes. Nicht jede dieser Klauseln ist für jedes System unmittelbar gesetzlich vorgeschrieben. Sie sollen der Einrichtung ermöglichen, ihre eigenen Pflichten tatsächlich zu erfüllen.

Besondere Aufmerksamkeit verdienen Klauseln zur „Produktverbesserung“. Sie sollten klären, ob Inhalte gespeichert, menschlich geprüft, weitergegeben oder für Training verwendet werden, welche Daten betroffen sind und wer darüber entscheidet. Eine unspezifische Zustimmung zum Anbietertraining ist keine belastbare Ersatzlösung für diese Prüfung. Bei Berufsgeheimnissen ist außerdem § 203 StGB einschließlich der Voraussetzungen der Einbindung weiterer mitwirkender Personen zu beachten. Eine datenschutzrechtlich zulässige Verarbeitung beantwortet die strafrechtliche Geheimnisfrage nicht automatisch. [22,37]

6.8 Information Einwilligung und Beteiligung

Patientenaufklärung, datenschutzrechtliche Information, eine gegebenenfalls erforderliche datenschutzrechtliche Einwilligung und die Befugnis zur Tonaufnahme sind getrennte Prüfgegenstände. Nicht jede interne KI-Unterstützung verlangt eine eigene Behandlungseinwilligung. Wird der Einsatz für Art, Risiken oder Alternativen einer medizinischen Maßnahme wesentlich, muss dies im Rahmen der §§ 630d und 630e BGB berücksichtigt werden. Die gesetzlich erforderliche mündliche Aufklärung wird nicht durch einen Chatbot ersetzt. [22,27,37,44]

Für Beschäftigte gilt: Einwilligung ist kein bequemes Standardinstrument zur Einführung von Überwachung. Bei ihrer Freiwilligkeit ist das Abhängigkeitsverhältnis ausdrücklich zu berücksichtigen. Die einschlägige Rechtsgrundlage ist einzelfallbezogen einschließlich DSGVO und anwendbarem Beschäftigtendatenschutz zu prüfen. Beteiligung der Interessenvertretung ersetzt weder individuelle Datenschutzrechte noch die Prüfung unzulässiger Diskriminierung. Umgekehrt ersetzt eine individuelle Einwilligung nicht die erforderliche Mitbestimmung. Für öffentliche und kirchliche Arbeitgeber sind die jeweiligen Sonderregime gesondert zu bestimmen. [22,24,25,43]

Art. 22 DSGVO verlangt besondere Aufmerksamkeit, wenn ausschließlich automatisierte Entscheidungen rechtliche oder ähnlich erhebliche Auswirkungen entfalten. Eine fachkundige Person muss für echte menschliche Entscheidung Spielraum, Informationen und Zeit haben; ein routinemäßiger Bestätigungsklick beseitigt das Problem nicht zuverlässig. Auch unterhalb dieser Schwelle bleiben die übrigen Datenschutz- und Gleichbehandlungspflichten bestehen. [22,25]

6.9 Konstruierte rechtliche Grenzfälle

Fall A Vom Gesprächsprotokoll zur Therapieempfehlung

Eine Klinik lässt Behandlungsgespräche durch einen Cloud-Dienst transkribieren. Zu klären sind insbesondere die Befugnis zur Aufnahme des nichtöffentlich gesprochenen Wortes, Geheimnisschutz, Datenverarbeitung und die Freigabe des erzeugten Textes. Ein unterschriebener Auftragsverarbeitungsvertrag löst die Aufnahmefrage nach § 201 StGB nicht. Eine vertretbare Einführungsentscheidung sollte einen nutzbaren Ablauf ohne Aufnahme vorsehen; dies ist hier eine Governance-Empfehlung, keine behauptete universelle gesetzliche Formvorgabe. [22,37]

Ergänzt der Dienst anschließend patientenspezifische Therapie- oder Dosierungsempfehlungen, muss die Medizinproduktequalifikation neu geprüft werden. Eine allgemeine Dokumentationsfunktion und eine medizinische Entscheidungsfunktion sind rechtlich nicht gleichzusetzen. Auch fachärztliche Endkontrolle hebt eine medizinische Zweckbestimmung nicht automatisch auf. Die Freigabe sollte bis zur Klärung auf den bereits bewerteten Funktionsumfang begrenzt bleiben. [23,27,36]

Fall B Administrative Bewerbungsorganisation oder entscheidende Rangliste

Ein Konzern nutzt KI zunächst zur Bereinigung doppelter Bewerbungsdateien. Später erstellt dasselbe Produkt Eignungswerte, nach denen nur die besten zehn Prozent menschlich gelesen werden. Für die Einordnung ist dieser tatsächliche Filter maßgeblich, nicht die Bezeichnung „Assistenz“. Bewerbungsanalyse, Filterung und Bewertung gehören zum Regelungsbereich des Anhangs III Nr. 4; die Voraussetzungen und Ausnahmen des Art. 6 bleiben im konkreten Fall zu prüfen. [20,21]

Unabhängig vom späteren Beginn der Hochrisikopflichten sind Datenschutz, Art. 22 DSGVO, AGG und Beteiligungsrechte schon jetzt relevant. Die Organisation muss erklären können, was mit aussortierten Personen geschieht, wie Fehler korrigiert werden und ob ein Mensch tatsächlich entscheidet. Biometrische Emotionserkennung im Arbeitsbereich ist zusätzlich an Art. 5 Abs. 1 Buchst. f KI-Verordnung zu messen; eine vermeintliche Verbesserung der Personalauswahl begründet keine medizinische Ausnahme. [20–25]

Fall C Klinikeigener Wissensassistent wird konzernweite Triage

Ein intern entwickelter Assistent beantwortet zunächst Fragen zu Dienstanweisungen. Anschließend priorisiert er anhand individueller Symptome die Behandlung; schließlich soll eine andere Konzerngesellschaft ihn übernehmen. Dies verändert Zweck, Wirkung und gegebenenfalls Anbieter- beziehungsweise Herstellerrolle. Eine allgemeine Wissenssuche ist nicht ohne Weiteres mit patientenbezogener Dringlichkeitsbewertung gleichzusetzen. Medizinproduktequalifikation, KI-Risikoeinstufung und Datenrollen sind neu zu prüfen. [20–23,36]

Die Bezeichnung „In-house“ beseitigt diese Prüfungen nicht. Gerade die andere juristische Person kann gegen eine Voraussetzung des Art. 5 Abs. 5 MDR sprechen. Ebenso wenig legitimiert ein lokaler Server den Einsatz: Er löst weder Leistungsnachweis noch Aufklärung, Zugriffsschutz oder Zweckbindung. Empfohlen ist ein getrenntes Freigabeverfahren für den neuen klinischen Zweck mit pflegerischer und ärztlicher Beteiligung sowie dokumentiertem Rückfallverfahren. [22,23,27,44,45]

6.10 Nachweise für die Rechtsprüfung

Tabelle 6 Prüfnachweise und rechtliche Auslöser

Nachweis	Rechtscharakter und Auslöser	Zuständige Vorbereitung
Einsatz-, Rollen- und Fristenblatt	Empfohlene Zusammenführung verschiedener Prüfungen; keine universell vorgeschriebene KI-Form.	Fachbereich und Rechtsabteilung, mit Datenschutz und Technik.
Rechtsgrundlagen- und Datenflussprüfung	Konkretisierung insbesondere von Art. 5, 6, 9 und 24 DSGVO; Verarbeitung muss rechtmäßig und nachweisbar sein.	Verantwortlicher, beraten durch Datenschutzbeauftragte.
Vertragliche Unterlagen	Art. 28 bei Auftragsverarbeitung; Art. 26 bei gemeinsamer Verantwortlichkeit; Kapitel V bei einschlägiger Drittlandübermittlung.	Einkauf und Recht mit Datenschutz.

Nachweis	Rechtscharakter und Auslöser	Zuständige Vorbereitung
DSFA oder begründete Schwellenprüfung	DSFA bei voraussichtlich hohem Risiko nach Art. 35; nicht allein wegen des Etiketts KI. Bei verbleibendem hohem Risiko Art. 36 prüfen.	Verantwortlicher; Datenschutzbeauftragte beraten.
Grundrechte-Folgen abschätzung	Art. 27 nur im gesetzlichen persönlichen/sachlichen Anwendungsbereich und ab einschlägigem Anwendungsbeginn; kein Ersatz der DSFA.	Künftiger verpflichteter Betreiber mit Fachfunktionen.
Produkt- und Zwecknachweise	Bei einschlägigem Produktrecht beispielsweise Konformitätsunterlagen oder Nachweise der In-house-Voraussetzungen; nicht jedes System verlangt CE.	Medizinprodukt-/Produktverantwortliche.
Informations-, Beteiligungs- und Kontrollkonzept	Je nach Fall DSGVO, Recht der Interessenvertretung, Behandlungsrecht; Art. 26 KI-Verordnung zusätzlich nach Anwendungsbeginn.	Fachleitung, Personal, Recht und Interessenvertretung.
Änderungs-, Vorfall- und Exit-Dossier	Rechtsabhängige Pflichten plus empfohlene organisatorische Zusammenführung; Fristen und Meldeschwellen je Rechtsgebiet gesondert festlegen.	Fachverantwortung, IT-Sicherheit, Qualität und Recht.

Ein ungeklärter Zulässigkeitsmangel ist kein „ethisch akzeptiertes Restrisiko“. Ein entscheidungsfähiges Dossier unterscheidet fehlende Tatsachen, juristische Auslegungsfragen und nach rechtlicher Klärung verbleibende ethische Zielkonflikte. Dafür sollte es jeweils einen benannten Entscheider, eine Frist und eine dokumentierte Eskalation geben. Das Ethikgremium kann Konflikte sichtbar machen und Bedingungen empfehlen; gesetzliche Verantwortung oder behördliche Zuständigkeit geht dadurch nicht auf das Gremium über.

7 Normative Diskussion und begründete Empfehlungen

7.1 Von wissenschaftlichen Befunden zu verantwortbaren Entscheidungen

Die entscheidende Schwierigkeit institutioneller KI-Ethik liegt nicht darin, eine Liste anerkannter Werte zusammenzustellen. Sie liegt darin, zwischen mehreren vertretbaren, mitunter unvereinbaren Handlungszielen zu entscheiden. Eine Einrichtung kann zugleich die Sicherheit erhöhen, Beschäftigte entlasten, Selbstbestimmung achten und Ressourcen schonen wollen. Eine konkrete Anwendung wird diese Ziele jedoch selten gleichzeitig und im gleichen Umfang erreichen. Wissenschaftliche Untersuchungen können die erwartbaren Folgen genauer beschreiben. Sie können nicht allein entscheiden, welche Nachteile einer Gruppe zugunsten welcher Vorteile einer anderen zugemutet werden dürfen.

Für die Interpretation der vorliegenden Analyse ist deshalb eine Unterscheidung zentral: Ein empirischer Befund beschreibt einen beobachteten Zusammenhang oder Effekt; eine normative Prämisse begründet, weshalb eine Folge als erstrebenswert, hinnehmbar oder unzulässig bewertet wird. Erst zusammen mit den Bedingungen des konkreten Einsatzes tragen beide eine Handlungsempfehlung. Aus einer Zeitersparnis folgt beispielsweise weder die moralische Berechtigung zum Personalabbau noch die Pflicht, sämtliche frei werdende Zeit einer bestimmten Aufgabe zuzuteilen. Dazu bedarf es zusätzlicher Annahmen über Versorgungsbedarf, Arbeitsbelastung und gerechte Verteilung.

Die folgenden Überlegungen machen solche Annahmen sichtbar. Sie sind eine begründete redaktionelle Synthese, kein Ergebnis einer Konsensuskonferenz, eines Delphi-Verfahrens oder einer eigenen Befragung. Auch die Empfehlungsdarstellung stellt keine formale Evidenzgraduierung dar. Ihre Funktion ist, die Verbindung zwischen Wissensgrundlage, Werturteil, institutioneller Zuständigkeit und überprüfbarem Handeln nachvollziehbar zu machen.

7.2 Autonomie Nicht jede Wahlmöglichkeit schafft Selbstbestimmung

Die WHO behandelt menschliche Autonomie als eigenständigen Bezugspunkt verantwortlicher KI-Governance.[1] Für die institutionelle Praxis reicht es nach unserer Einschätzung nicht, daraus lediglich eine Informationspflicht oder einen zusätzlichen Bestätigungsdialog abzuleiten. Selbstbestimmung setzt voraus, dass Menschen verstehen können, welche Bedeutung ein System für ihre Situation hat, und dass ihre begründeten Einwände den weiteren Prozess beeinflussen können. Wer zwar ablehnen darf, dann aber keine praktisch erreichbare Versorgung oder Dienstleistung erhält, besitzt möglicherweise nur eine formale Wahl.

Gleichwohl wäre die Forderung nach einer uneingeschränkten Alternative zu jeder automatisierten Funktion zu weitgehend. Sie könnte erhebliche Kosten erzeugen, gemeinsame Sicherheitsprozesse unterlaufen oder anderen Personen Ressourcen entziehen. Die angemessene Frage lautet daher, welche Entscheidungen persönliche Zustimmung erfordern, wo verständliche Information und wirksamer Einspruch genügen und welche Alternativen unter den konkreten Bedingungen zumutbar verfügbar sind. Diese ethische Abwägung ist von der gesonderten Prüfung gesetzlicher Informations-, Einwilligungs- und Widerspruchsrechte zu unterscheiden.

Besonders sorgfältig sollte die Prüfung ausfallen, wenn Abhängigkeit das Verhältnis prägt: zwischen Patientin und Krankenhaus, pflegebedürftigem Menschen und Einrichtung oder Beschäftigten und Arbeitgeber. Zustimmung darf hier nicht vorschnell als Beweis freiwilliger Akzeptanz verstanden werden. Unsere Empfehlung ist deshalb eine Prüfung der tatsächlichen Handlungsmöglichkeiten: Kann jemand eine fehlerhafte Einordnung ansprechen, ohne Nachteile befürchten zu müssen? Wird ein Einwand von einer befugten Person bearbeitet? Ist eine Korrektur noch rechtzeitig möglich? Ein gut gestaltetes Verfahren stärkt Autonomie nicht dadurch, dass es jede Unsicherheit beseitigt, sondern dadurch, dass es Menschen trotz Unsicherheit handlungsfähig hält.

7.3 Nutzen und Nichtschaden Auch Nichtstun hat Folgen

Das Wohlergehen der Betroffenen bildet einen zweiten Ausgangspunkt. Es rechtfertigt, Anwendungen mit nachvollziehbarem Nutzen ernsthaft zu prüfen, anstatt Technik allein wegen ihrer Neuartigkeit abzulehnen. Zugleich wäre es verfehlt, jeden statistischen Leistungsgewinn als hinreichende Begründung einer Einführung anzusehen. Maßgeblich ist, ob der Gewinn im tatsächlichen Arbeitszusammenhang eine relevante Verbesserung bewirkt und welche neuen Belastungen dafür entstehen. Die WHO empfiehlt für große multimodale Modelle klar definierte Aufgaben und eine Bewertung ihrer Folgen; DECIDE-AI betont die transparente Darstellung früher klinischer Evaluation.[2,28]

Unsere Schlussfolgerung daraus ist eine vergleichende Bewertung, die auch Unterlassungs- und Opportunitätskosten einbezieht. Geld, Personal und Aufmerksamkeit, die in eine KI-Anwendung fließen, stehen zunächst nicht für andere Verbesserungen zur Verfügung. Ein geringfügig genaueres Modell kann gegenüber einer zusätzlichen Pflegefachperson, einer besseren Alarmorganisation oder einer zuverlässigeren konventionellen Software die schlechtere Verwendung knapper Mittel sein. Umgekehrt kann der Verzicht auf eine tatsächlich hilfreiche Assistenz vermeidbare Belastungen fortschreiben. Welche Alternative vorzuziehen ist, lässt sich weder aus Technikbegeisterung noch aus einem abstrakten Vorsorgeprinzip ableiten.

Bei dieser Abwägung sollten Schaden und Nutzen nicht unbesehen zu einem einzigen Punktwert verrechnet werden. Ein Vorteil für viele kann einen schweren, konzentrierten Nachteil für wenige nicht automatisch rechtfertigen. Zwingende rechtliche Grenzen bleiben ohnehin außerhalb einer freien Nutzenrechnung. Innerhalb des zulässigen Handlungsspielraums empfehlen wir, zunächst die schwersten plausiblen Schäden und die verfügbaren Schutzmaßnahmen zu prüfen und erst danach verbleibende Vor- und Nachteile gegeneinander abzuwägen. Je weniger reversibel ein Schaden ist, desto stärker sollte nach unserer normativen Position die Begründungslast für einen Einsatz unter erheblicher Unsicherheit ausfallen.

7.4 Gerechtigkeit Verteilung statt bloßer Durchschnittsleistung

Eine Anwendung ist nicht bereits deshalb gerecht, weil ihre durchschnittliche Genauigkeit hoch ist. Die relevante Verteilung kann zwischen Patientengruppen, Standorten, Berufsgruppen oder Schichten verlaufen. Ebenso wichtig ist, wer zusätzliche Kontrollarbeit übernimmt und wer wirtschaftliche Vorteile erhält. STANDING Together liefert konsenterte Empfehlungen für transparentere und besser beurteilbare Gesundheitsdatensätze. Diese helfen, Fragen der Repräsentation zu untersuchen; sie entscheiden jedoch nicht abschließend, welche Verteilung von Chancen und Belastungen eine Einrichtung für gerecht halten sollte.[30]

Hier schlagen wir eine doppelte Prüfung vor. Erstens ist zu untersuchen, ob Unterschiede auf vermeidbare Benachteiligungen hindeuten. Zweitens muss begründet werden, welche Unterschiede trotz Schutzmaßnahmen verbleiben und warum sie vertretbar sein sollen. Gleichbehandlung kann dafür unzureichend sein: Menschen mit höherem Unterstützungsbedarf benötigen möglicherweise zusätzliche Hilfe, um eine nominell gleiche Leistung überhaupt nutzen zu können. Umgekehrt bedeutet jede gruppenspezifische Anpassung eine begründungspflichtige

Entscheidung; sie darf nicht auf pauschalen Zuschreibungen beruhen.

Eine schwierige Grenze entsteht, wenn belastbare Daten zu kleinen Gruppen fehlen. Es wäre wissenschaftlich falsch, daraus Fairness abzuleiten. Ebenso problematisch wäre ein vollständiger Ausschluss dieser Gruppen von potenziell nützlichen Anwendungen, ohne die Folgen des Ausschlusses zu betrachten. Vertretbar können eng begrenzte Verwendungen, zusätzliche fachliche Kontrolle oder gezielte weitere Evaluation sein. Welche Lösung passt, hängt von Schadensschwere, Alternativen und Datenlage ab. Die erforderliche Datenerhebung muss rechtlich tragfähig sein; ein gutes Fairnessziel rechtfertigt keine beliebige Sammlung sensibler Merkmale.[18] Wesentlich ist, dass Unsicherheit sichtbar bleibt und nicht in einem positiven Gesamturteil verschwindet.

7.5 Beziehung und Fürsorge Der Kontakt ist nicht nur eine Prozessvariable

Für Krankenhäuser und Pflegeeinrichtungen besitzt Beziehung einen Stellenwert, der sich nicht vollständig durch Zufriedenheitswerte oder Gesprächsdauer abbilden lässt. Menschen können auf persönliche Wahrnehmung, Vertrauen und die Möglichkeit angewiesen sein, uneindeutige Bedürfnisse auszudrücken. Der ICN-Kodex fordert deshalb für KI und entsprechende Geräte personenzentrierte Pflege sowie Unterstützung statt Ersetzung menschlicher Beziehungen. Auch die ZEKO behandelt die Bedeutung der zwischenmenschlichen Ebene ärztlicher Kommunikation.[12,13]

Unsere normative Interpretation lautet nicht, dass eine technische Interaktion grundsätzlich minderwertig ist. Ein System kann bei alltäglichen Aufgaben Selbstständigkeit ermöglichen oder eine persönliche Begegnung vorbereiten. Entscheidend ist, ob die institutionelle Gestaltung Menschen zusätzliche Möglichkeiten eröffnet oder ihnen die erreichbare menschliche Unterstützung entzieht. Eine freiwillig gewählte technische Hilfe ist anders zu beurteilen als eine faktische Ersatzversorgung, die ausschließlich aus fehlendem Personal entsteht. Dennoch wäre es ebenso unangemessen, einen unter schwierigen Bedingungen tatsächlich hilfreichen Einsatz wegen eines unerreichbaren Idealzustands pauschal abzulehnen.

Daraus folgt eine besondere Aufgabe für die pflegerische Leitung: Sie sollte die Veränderung des gesamten Versorgungsgeschehens mitbeurteilen, nicht nur die Bedienbarkeit eines Geräts. Zu fragen ist, welche Begegnungen entfallen, welche neuen Kontakte entstehen und wer bemerkt, wenn jemand mit dem Angebot nicht zurechtkommt. Neben strukturierten Rückmeldungen können begründete Fallbesprechungen hierfür wichtig sein. Solche Beobachtungen ersetzen keine Wirksamkeitsstudie; sie erschließen aber Wertkonflikte und unbeabsichtigte Folgen, die ein enger technischer Endpunkt übersehen kann. Ein Nutzenversprechen sollte sich auch daran messen lassen, ob die Organisation die Voraussetzungen für zwischenmenschliche Zuwendung tatsächlich erhält.

7.6 Verantwortung Befugnisse müssen zur übernommenen Aufgabe passen

Eine Organisation kann Verantwortung auf mehrere Rollen verteilen, ohne dass sie dadurch verschwindet. Problematisch wird die Verteilung, wenn die offiziell verantwortliche Person weder die notwendigen Informationen noch Zeit oder Eingriffsmöglichkeiten besitzt. Die ZEKO weist darauf hin, dass ärztliche Verantwortung nicht zur vollständigen Verlagerung der Lasten auf einzelne Ärzte führen darf. Die AEM verlangt geklärte Mandate und angemessene Voraussetzungen für Ethikberatung.[13,14] Diese Vorgaben tragen unsere Empfehlung, Verantwortlichkeit konsequent mit Befugnissen und Ressourcen zu verbinden.

Eine Ethikbeauftragte sollte demnach nachvollziehbare Einwände formulieren und eine Entscheidung zur erneuten Prüfung vorlegen können. Daraus folgt kein allgemeines persönliches Vetorecht. Eine institutionelle Leitung kann begründet anders entscheiden als das beratende Gremium, sollte dann aber den Dissens und die tragenden Gründe dokumentieren. Umgekehrt darf ein Ethikvotum nicht als Unterschrift verwendet werden, hinter der Geschäftsführung, Produktverantwortliche oder Fachbereiche ihre eigenen Pflichten verbergen. Die AEM-Empfehlungen zur Leitlinienentwicklung unterscheiden entsprechend zwischen Erarbeitung und autorisierter Inkraftsetzung.[15]

Verantwortung betrifft außerdem den Zeitpunkt nach der Freigabe. Wer eine Anwendung einführt, sollte vorher festlegen, wer neue Erkenntnisse bewertet und Änderungen anordnen kann. NIST bietet dafür einen organisationsweiten Risikomanagementrahmen; ISO/IEC 42001 beschreibt Anforderungen an ein KI-Managementsystem.[16,31] Beide können konsistente Prozesse unterstützen. Sie ersetzen nicht die

Entscheidung, welche Risiken im konkreten Fall moralisch vertretbar sind. Auch eine vollständig dokumentierte Entscheidung kann falsch sein. Dokumentation ist deshalb ein Mittel der Nachprüfbarkeit, nicht das Endziel verantwortlichen Handelns.

7.7 Zielkonflikte, die ausdrücklich entschieden werden sollten

Drei Konflikte verdienen besondere Aufmerksamkeit. Erstens kann höhere Empfindlichkeit eines Warnsystems zusätzliche Fehlalarme und damit neue Belastungen erzeugen. Die ethische Frage betrifft dann nicht nur verpasste Ereignisse, sondern auch die Versorgung der Menschen, denen Aufmerksamkeit entzogen wird. Ein Schwellenwert sollte deshalb nicht allein von einem technischen Team oder allein nach betriebswirtschaftlichen Vorgaben gewählt werden. Fachliche Einschätzung, verfügbare Kapazitäten und die Folgen unterschiedlicher Fehlertypen gehören in dieselbe Entscheidung.

Zweitens können Transparenz und Sicherheit miteinander in Spannung geraten. Betroffene benötigen verständliche Gründe und Grenzen einer Anwendung; daraus folgt nicht, dass jede technische Einzelheit öffentlich zugänglich sein muss. Informationen über Schwachstellen oder vertrauliche Daten können ihrerseits Schutzbedarf besitzen. Empfehlenswert ist eine adressatengerechte Transparenz: verständliche Informationen für Betroffene, ausreichend detaillierte Unterlagen für befugte Prüfer und kontrollierter Zugang zu besonders schutzbedürftigen Inhalten. Die Orientierungshilfen zu KI und RAG sowie die Sicherheitshinweise des BSI machen deutlich, dass Datenzugang und technische Angriffsflächen eigenständig zu bearbeiten sind.[18,19,34]

Drittens kostet Kontrolle selbst Ressourcen. Eine ausufernde Prüfung kann sinnvolle Verbesserungen verzögern und kleine Projekte unverhältnismäßig belasten. Zu wenig Prüfung kann dagegen Schäden verursachen, deren Behebung wesentlich aufwendiger ist. Unsere Empfehlung ist deshalb keine maximale Bürokratie, sondern eine begründete Prüftiefe. Risiken, Reichweite, Handlungsmacht und Reversibilität sollten das Verfahren bestimmen. Dieselben Fragen dürfen bei einer risikoarmen Formulierungshilfe knapp und bei einer klinischen Priorisierung ausführlich beantwortet werden. Proportionalität bedeutet dabei nicht, unangenehme Fragen auszulassen, sondern ihre Bearbeitung an der tatsächlichen Bedeutung auszurichten.

7.8 Empfehlungen Wissensgrundlage, Werturteil und Nachweis verbinden

Die folgende Evidence-to-Decision-Darstellung enthält neun eigene Empfehlungen. R1 bis R9 sind zur Rückverfolgung stabil gekennzeichnet. „Evidenzbasis“ bezeichnet die jeweils herangezogene Wissens- oder Orientierungsgrundlage. Sie unterscheidet ausdrücklich zwischen Forschungssynthese, fachlichem Konsens, berufsethischer Orientierung und Managementrahmen. Keiner der Verweise belegt die Wirksamkeit des hier vorgeschlagenen Gesamtverfahrens. Zuständigkeiten und Prüfnachweise sind Umsetzungsvorschläge, die eine Einrichtung an ihr Organisationsrecht und ihre tatsächlichen Möglichkeiten anpassen muss.

R1 Zweck und realistische Alternativen vor Beschaffung vergleichen.

Wissensgrundlage. WHO-Orientierung [1,2]; NIST-Rahmen [16]. **Ethische Begründung.** Wohlergehen und verantwortliche Verwendung knapper Mittel.

Verantwortung. Fachbereich und Leitung; Einkauf unterstützt. **Prüfnachweis.** Vergleich mit bisherigem Verfahren und nicht KI-basierten Alternativen.

Verbleibende Unsicherheit. Lokale Nutzen- und Aufwandsschätzungen können unzuverlässig sein.

R2 Tatsächliche Einflussmöglichkeiten Betroffener vorsehen.

Wissensgrundlage. WHO-Ethik [1]; AEM-Leitlinienempfehlung [15]. **Ethische Begründung.** Autonomie und faire Verfahren.

Verantwortung. Prozessverantwortliche; Ethikberatung und Vertretungen. **Prüfnachweis.** Verständliche Information, Rückmeldeweg und dokumentierte Bearbeitung.

Verbleibende Unsicherheit. Beteiligte Personen bilden nicht automatisch alle Interessen ab.

R3 Nutzen im vollständigen Arbeitsablauf bewerten.

Wissensgrundlage. Arbeitsablauf- und Mensch-KI-Befunde [4,7,41]; Berichtsempfehlungen DECIDE-AI und TRIPOD-LLM [28,29]. **Ethische Begründung.** Nutzen und Nichtschaden statt bloßer Modelloptimierung.

Verantwortung. Fachliche Leitung; methodische und Qualitätskompetenz. **Prüfnachweis.** Geeigneter Vergleich, relevante Endpunkte und erfasster Kontrollaufwand.

Verbleibende Unsicherheit. Kleine oder kurze Evaluationen erfassen seltene Folgen unzureichend.

R4 Verteilungswirkungen gesondert begründen.

Wissensgrundlage. Audit zur ungeeigneten Zielvariable [9]; Konsensempfehlungen STANDING Together [30].

Ethische Begründung. Gerechtigkeit und Schutz vor vermeidbarer Benachteiligung.

Verantwortung. Fachbereich; Datenkompetenz, Datenschutz und Ethikberatung. **Prüfnachweis.** Gruppenbezogene Befunde, Zugangsprüfung und ausgewiesene Datenlücken.

Verbleibende Unsicherheit. Gerechte Behandlung besitzt keinen universell ausreichenden Einzelindikator.

R5 Menschliche Unterstützung praktisch erreichbar halten.

Wissensgrundlage. ICN-Kodex und ZEKO-Stellungnahme [12,13]. **Ethische Begründung.** Beziehungsschutz, Fürsorge und Selbstbestimmung.

Verantwortung. Pflegerische und ärztliche Leitung; im Konzern zuständiger Dienstleistungsbereich. **Prüfnachweis.** Beschriebene Kontaktwege und Prüfung unerfüllten Unterstützungsbedarfs.

Verbleibende Unsicherheit. Beziehungserleben ist kontextabhängig und nur teilweise messbar.

R6 Verantwortung mit Informationen, Zeit und Befugnissen ausstatten.

Wissensgrundlage. AEM-Standards [14]; NIST-Rahmen [16]. **Ethische Begründung.** Rechenschaft ohne bloße Verantwortungsverschiebung.

Verantwortung. Leitung; Personal- und Prozessverantwortliche. **Prüfnachweis.** Rollenmandat, Ressourcen und erprobte Eskalationswege.

Verbleibende Unsicherheit. Formale Regeln garantieren keine unterstützende Organisationskultur.

R7 Datenzugriff und Handlungsvollmacht risikoadäquat begrenzen.

Wissensgrundlage. DSK-Orientierung [18,19]; NIST und BSI [17,34]. **Ethische Begründung.** Vertraulichkeit, Schadensvermeidung und Kontrolle.

Verantwortung. IT-Sicherheit und Prozessverantwortliche; Datenschutz berät. **Prüfnachweis.** Berechtigungskonzept, nachvollziehbare Datenflüsse und geeignete Tests.

Verbleibende Unsicherheit. Neue Angriffsmuster können vorhandene Schutzmaßnahmen umgehen.

R8 Freigaben bei relevanten Änderungen erneut prüfen.

Wissensgrundlage. NIST-Risikomanagement [16,17]; ISO-Managementansatz [31]. **Ethische Begründung.** Dauerhafte Verantwortlichkeit bei veränderlichen Bedingungen.

Verantwortung. Benannte Betriebsverantwortliche; Leitung entscheidet wesentliche Änderungen. **Prüfnachweis.** Versionsbezug, Prüfkalender, Vorfalls- und Änderungsprotokoll.

Verbleibende Unsicherheit. Schleichende Veränderungen können unentdeckt bleiben.

R9 Ressourcen, Abhängigkeit und Ausstieg gemeinsam bewerten.

Wissensgrundlage. IEA-Analyse [33]; WHO-Nachhaltigkeitsprinzip [1]. **Ethische Begründung.** Langfristiges Wohlergehen und verantwortliche Ressourcenverteilung.

Verantwortung. Leitung, Einkauf, Betrieb und Nachhaltigkeitsverantwortliche. **Prüfnachweis.** Aufgabenbezogene Ressourcenabschätzung und erprobbares Ersatzverfahren.

Verbleibende Unsicherheit. Verbrauchszurechnung und künftige Anbieterbedingungen sind unsicher.

Die Empfehlungen sollten nicht als gleich gewichtete Punkte eines Summenscores verwendet werden. Ein schwerwiegendes Problem wird nicht dadurch vertretbar, dass viele andere Felder positiv ausfallen. Ebenso wenig verlangt jede unvollständige Information den endgültigen Verzicht auf eine Anwendung. Die entscheidende Aufgabe besteht darin, verbleibende Unsicherheit, mögliche Folgen und vorhandene Alternativen offen zu benennen. Wo eine begrenzte Erprobung vertretbar erscheint, sollte deren Erkenntnisziel ausdrücklich feststehen; eine bloß verlängerte Unsicherheit ist noch kein lernender Prozess.

7.9 Umsetzung Gemeinsamkeiten und Grenzen zwischen Krankenhaus und Konzern

Krankenhäuser und große Unternehmen benötigen beide ein verlässliches Zusammenspiel aus fachlicher Zuständigkeit, Recht, Datenschutz, Sicherheit und Leitung. Daraus folgt jedoch keine austauschbare Organisationslösung. Im Krankenhaus können akute Erkrankung, Abhängigkeit und zeitkritische Entscheidungen die Möglichkeiten zu Information und Widerspruch einschränken. Die pflegerische Perspektive bleibt unverzichtbar, weil Kontinuität, Beobachtung und tatsächliche Bearbeitungskapazität Teil der Versorgung sind. Ein neues zentrales KI-Gremium sollte vorhandene Ethik- und Qualitätsstrukturen ergänzen, nicht verdrängen.

Im Konzern unterscheiden sich Risiken stärker zwischen Geschäftsbereichen. Eine interne Textassistentin, ein Auswahlverfahren im Personalbereich und eine sicherheitsrelevante Produktionssteuerung benötigen unterschiedliche fachliche Bewertung. Ein einheitlicher Mindestprozess kann Orientierung schaffen; die konkrete Entscheidung sollte dennoch bei identifizierbaren Verantwortlichen mit Kenntnis des Anwendungskontexts bleiben. Übergreifende Ethikkompetenz ist besonders dort hilfreich, wo Kosten, Nutzen und Verantwortung zwischen Bereichen auseinanderfallen. Ein erfolgreicher Pilot an einem Standort begründet noch keine ungeprüfte Übernahme in andere Länder oder Arbeitsabläufe.

Die Reichweite des vorgeschlagenen Verfahrens bleibt begrenzt. Es kann Abwägungen transparenter machen, aber weder alle Fehler verhindern noch gesellschaftliche Macht- und Verteilungsprobleme allein lösen. Schwerwiegende Risiken allgemeiner KI, wie sie der International AI Safety Report behandelt, erfordern auch Forschung, staatliche Aufsicht und internationale Zusammenarbeit.[32] Einrichtungen tragen Verantwortung für ihren Einflussbereich; daraus sollte weder institutionelle Selbstüberschätzung noch die Behauptung folgen, externe Risiken gingen sie nichts an.

Eine freiwillige unabhängige Evaluation des Verfahrens wäre sinnvoll: Mehrere unterschiedlich ausgestattete Einrichtungen könnten dessen Verständlichkeit, Prüfaufwand, Entscheidungsqualität und unbeabsichtigte Folgen prospektiv untersuchen. Diese externe Prüfung hat für den vorliegenden Beitrag nicht stattgefunden.

Die tragende Schlussfolgerung ist damit anspruchsvoller als die Forderung nach vertrauenswürdiger Technik. Verantwortlicher KI-Einsatz benötigt begründbare Entscheidungen über Zwecke, Belastungen und Grenzen. Wissenschaft kann dafür Unsicherheit reduzieren und falsche Nutzenversprechen korrigieren. Welche Zukunft eine Organisation ihren Patientinnen, Beschäftigten oder Kunden zumuten will, bleibt jedoch eine menschliche und institutionelle Verantwortung, die auch ein leistungsfähiges System nicht übernehmen kann.

8 Zuständigkeiten und institutionelle Verantwortung

Die AEM-Standards sehen Ethikberatung auch als organisationsethische Unterstützung an. Sie betonen ein geklärtes Mandat, multiprofessionelle Kompetenz, inhaltliche Unabhängigkeit und Ressourcen. Die AEM-Empfehlungen zur Leitlinienentwicklung unterscheiden wiederum die Erarbeitung einer Empfehlung von ihrer autorisierten Inkraftsetzung.[14,15]

Daran anschließend schlagen wir folgende Aufgabenverteilung vor. Sie ist ein Organisationsmodell, keine allgemeingültige gesetzliche Zuständigkeitsordnung:

Tabelle 7 Vorgeschlagene Zuständigkeiten

Rolle	Aufgabe im vorgeschlagenen KI-Verfahren	Grenze der Rolle
Träger, Geschäftsführung oder Vorstand	Zweck, Budget, Verantwortliche und Freigabebedingungen festlegen; verbleibende Organisationsrisiken begründet entscheiden.	Verantwortung nicht vollständig an Anwender oder Ethikstelle delegieren.
Pflegerische Leitung	Auswirkungen auf Pflegeprozess, Schichtbetrieb, Alarmbearbeitung, Personalressourcen und Beziehungsgestaltung prüfen; Umsetzung mitgestalten.	Nicht auf nachträgliche „Akzeptanzbeschaffung“ reduzieren.
Pflegefachpersonen und andere unmittelbar Anwendende	Alltagstauglichkeit, Fehler und unvorhergesehene Folgen in Tests und Rückmeldungen sichtbar machen.	Keine persönliche Sicherheitsgarantie für unbekannte Produktmängel übernehmen.
Ärztliche beziehungsweise sonstige fachliche Leitung	Fachliche Eignung, Vergleichsstandard und angemessene Reaktion auf Systemausgaben prüfen.	Nicht allein über organisatorische oder grundrechtliche Risiken entscheiden.
Ethikbeauftragte oder Ethikgremium	Werte- und Interessenkonflikte strukturieren, Betroffene beteiligen, Empfehlungen und Dissens dokumentieren.	Keine automatische Zulassungs-, Datenschutz- oder Haftungsinstanz.
Datenschutz, Informationssicherheit, Recht und gegebenenfalls Medizinproduktekompetenz	Jeweiligen Fachbereich prüfen und Anforderungen dokumentieren.	Getrennte Prüfungen ersetzen sich nicht; Datenschutzbeauftragte beraten und überwachen in ihrer gesetzlichen Rolle.
Beschäftigten- sowie Patienten- oder Nutzervertretung	Belastungen, verständliche Informationen, Beschwerdewege und Verteilungsfolgen einbringen.	Beteiligung nicht mit Zustimmung aller Betroffenen gleichsetzen.

Im Konzern treten an die Stelle der klinischen Rollen die fachlich verantwortlichen Geschäftsbereiche, etwa Personal, Kundenservice oder Produktion. Bei länderübergreifendem Einsatz sind lokale Perspektiven und Rechtslagen einzubeziehen. Ein zentrales Gremium sollte nicht verdecken, wer einen konkreten Prozess tatsächlich betreibt.

Die pflegerische Leitung ist besonders wichtig, weil die Wirkung einer Technik häufig erst zwischen Berufsgruppen und Schichten entsteht. Ein Alarm kann ärztlich sinnvoll und dennoch pflegerisch nicht bearbeitbar sein. Ein Roboter kann Wege sparen, aber zusätzliche Übergabe- und Aufsichtsarbeit erzeugen. Diese Konflikte sollten vor Beschaffung und Prozessfestlegung untersucht werden, nicht erst nach einer Produktdemonstration.

Die ZEKO betont bereits in ihrer Stellungnahme von 2021, dass die Verantwortung nicht allein auf einzelne Ärzte verlagert werden darf; auch Auswahl, Vorbereitung der Mitarbeitenden, Wartung und Schutz vor Missbrauch betreffen die Einrichtung.[13] Für die heutige Umsetzung empfehlen wir zusätzlich klar dokumentierte Eskalationsrechte, Zugang zu relevanten Unterlagen und geschützte Zeit für Prüfung und Schulung. Ein gesetzliches Universalveto der Ethikbeauftragten wird damit nicht behauptet.

Klinische Ethikberatung und Forschungsethik sind schließlich zu unterscheiden. Ob ein Vorhaben eine gesonderte Bewertung durch eine zuständige Forschungsethikkommission und weitere Genehmigungen benötigt, ist eigenständig zu klären. Die Bezeichnung „Pilot“ hebt entsprechende Anforderungen nicht auf.

9 Risikoadäquates Entscheidungsverfahren

Abbildung 3 – Sechs Prüfstufen vor und nach einer Freigabe

Stufe 1 · Auftrag, Vergleich und Grenzen festlegen Was soll besser werden, verglichen womit?
Stufe 2 · Belege statt Vorführungen verlangen Endpunkte, Population, Vergleichsgruppe
Stufe 3 · Rechte, Betroffene und Zielkonflikte prüfen Datenschutz, Beteiligung, Verteilungsfolgen
Stufe 4 · Begrenzten Pilot mit vorab festgelegten Kriterien planen Abbruchgründe vor dem Start
Stufe 5 · Begrenzte, begründete Freigabe erteilen Umfang, Dauer und Auflagen benennen
Stufe 6 · Betrieb, Änderungen und Ausstieg organisieren Rückkehr zum Alternativverfahren möglich halten

Gliederung nach Kapitel 9 dieser Studie. Redaktioneller Verfahrensvorschlag, kein wissenschaftlich validiertes Gesamtverfahren; Abschnitt 11.1 nennt die Grenzen.

Die folgenden Stufen sind unsere redaktionelle Umsetzungsempfehlung. Sie greifen Gedanken aus Risikomanagement und früher klinischer Evaluation auf, sind aber selbst nicht als Gesamtverfahren validiert.[16,28] Prüftiefe und Aufwand sollen mit Schadenspotenzial, Reichweite, Autonomie und Unumkehrbarkeit wachsen. Ein interner Formulierungsvorschlag braucht nicht denselben Prozess wie eine Therapiepriorisierung.

9.1 Auftrag, Vergleich und Grenzen festlegen

Die verantwortliche Stelle formuliert den Bedarf ohne Produktnamen: Welches Problem soll für wen verbessert werden? Wie wird es heute gelöst? Welche nicht KI-basierten Alternativen bestehen? Anschließend werden zulässige Aufgabe, Nutzer, Daten und ausgeschlossene Verwendungen festgehalten. Ohne diese Beschreibung ist auch die Nutzenprüfung unscharf.

Die Organisation erstellt dafür ein Anwendungsregister. Es soll auch KI-Funktionen innerhalb bereits genutzter Standardsoftware erfassen. Ein risikoadäquater Meldeweg für informelle Nutzung ist hilfreicher als ein Register, das nur offiziell beschaffte Großprojekte kennt.

9.2 Belege statt Vorführungen verlangen

Hersteller sollten Nachweise für die angebotene Version und den beabsichtigten Einsatz vorlegen. Zu prüfen sind Vergleichsgruppe, Datenherkunft, externe Validierung, relevante Endpunkte, Unsicherheiten, Finanzierung und Interessenkonflikte. Eine Veröffentlichung mit Autorenbezug zum Hersteller ist nicht automatisch unbrauchbar; die Abhängigkeit gehört aber in die Bewertung.

Je nach Frage kommen unterschiedliche Methoden infrage. Eine externe Validierung prüft Übertragbarkeit von Vorhersagen, eine geeignete vergleichende Wirkungsstudie den Nutzen des Einsatzes. Randomisierung ist für manche kausalen Fragen besonders hilfreich, aber nicht für jede Prüfaufgabe sinnvoll oder möglich. Methodische Beratung sollte deshalb vor Auswahl von Design und Fallzahl erfolgen.

9.3 Rechte, Betroffene und Zielkonflikte prüfen

Hier laufen Rechts-, Datenschutz-, Sicherheits- und Ethikprüfung parallel. Erfasst werden auch indirekt Betroffene: Menschen, die weniger Aufmerksamkeit erhalten, Beschäftigte, deren Arbeit verdichtet wird, oder Gruppen, die vom digitalen Zugang ausgeschlossen bleiben.

Die Ethikberatung sollte nicht lediglich fragen, ob die Beteiligten „KI akzeptieren“. Sie sollte prüfen, welche Alternativen sie haben, ob sie widersprechen können und wer mögliche Nachteile trägt. Gesetzlich erforderliche Folgenabschätzungen werden identifiziert; darüber hinaus kann die Organisation eine freiwillige ethische Folgenprüfung vorsehen.

9.4 Einen begrenzten Pilot mit vorab festgelegten Kriterien planen

Vor dem Einsatz werden Nutzenziel, Sicherheitsendpunkte, Messdauer, Datenzugriff und Abbruchkriterien festgelegt. Ein zunächst stiller Parallelbetrieb kann bei geeigneten Systemen prüfen, welche Ausgaben entstanden wären, ohne unmittelbar Entscheidungen zu steuern. Er beweist allerdings noch nicht, wie Menschen auf die Ausgaben reagieren oder ob sich Ergebnisse verbessern.

In einem anschließenden kontrollierten Einsatz sind auch Korrekturarbeit, Verzögerungen, übersehene Warnungen und Auswirkungen auf andere Aufgaben zu erfassen. Es sollten nicht nur frühe Enthusiasten teilnehmen. Nachtdienst, Vertretung, Stress, Sprachbarrieren und technische Ausfälle gehören je nach Anwendung zum Testkontext.

9.5 Eine begrenzte, begründete Freigabe erteilen

Eine Freigabe benennt Zweck, Version, Datenquellen, Nutzerkreis, Aufsicht, Ausschlüsse und nächstes Prüfdatum. Sie dokumentiert erwarteten Nutzen, verbleibende Risiken, offene Evidenzlücken und abweichende Voten. Eine Organisation kann eine Anwendung freigeben, weiter begrenzt erproben, zurückstellen oder ablehnen. Die Entscheidung und ihre Bedingungen sollten voneinander unterscheidbar sein.

Unsere Empfehlung ist, wesentliche Interessenkonflikte offenzulegen und bei besonders folgenreichen Einsätzen unabhängige Expertise einzubeziehen. Diese kann ein interner Bereich ohne Projektverantwortung oder eine externe Stelle sein. Eine einzige zusätzliche Unterschrift ist noch keine unabhängige Prüfung.

9.6 Betrieb, Änderungen und Ausstieg organisieren

Nach der Einführung werden sowohl Leistung als auch Folgen überwacht. Auslöser einer Neubewertung sind beispielsweise Modell- oder Datenänderungen, ein anderer Nutzerkreis, neue Aufgaben, auffällige Fehler oder veränderte Arbeitsbedingungen. Nicht jedes Sicherheitsupdate verlangt einen vollständigen Neustart; die Tiefe der erneuten Prüfung ist nach Art und Tragweite der Änderung zu begründen.

Vorfälle brauchen eine erreichbare Meldestelle, klare Zuständigkeit und einen Weg zurück zur sicheren Versorgung oder Arbeitsweise. Für Beschwerden empfehlen wir einen benannten Empfänger, eine nach Dringlichkeit festgelegte Bearbeitungsfrist und eine fachlich befugte Überprüfung mit Korrekturmöglichkeit. Dringlicher Schadensverdacht braucht einen sofort nutzbaren Eskalationsweg. Eine Pause sollte für Betroffene nicht unbemerkt zum Versorgungsabbruch werden. Ausfall- und Ausstiegsverfahren sind deshalb praktisch zu testen. Hinweise von Beschäftigten dürfen nicht durch eine Kultur sanktioniert werden, die ausschließlich auf Einführungsziele und Nutzungsquoten schaut.

10 Messung und lokale Evaluation

10.1 Warum 95 Prozent Genauigkeit wenig sagen kann

Ein bewusst vereinfachtes Rechenbeispiel, keine Studienzahl: Unter 10.000 untersuchten Personen haben 100 den gesuchten Zustand. Bei 90 Prozent Sensitivität werden 90 davon erkannt. Bei 95 Prozent Spezifität werden von den 9.900 übrigen Personen 495 fälschlich positiv markiert. Damit sind nur 90 von 585 positiven Meldungen richtig: rund 15,4 Prozent.

Das System kann trotzdem nützlich sein, wenn die anschließende Prüfung angemessen und der Nutzen ausreichend groß ist. Aber seine Folgen hängen entscheidend von Häufigkeit des Zustands, Fehlerkosten und verfügbarer Bearbeitungskapazität ab. Eine Kennzahl allein beantwortet diese Fragen nicht.

Tabelle 8 Dimensionen der lokalen Wirkungsmessung

Dimension	Beispiele für geeignete Messgrößen	Typische Fehlinterpretation
Relevanter Nutzen	Patientenergebnis, korrekte Fallbearbeitung, nachvollziehbar bessere Dienstleistung	Modellscore wird mit Versorgungserfolg gleichgesetzt.
Sicherheit	Schweregrad von Fehlern, übersehene kritische Ereignisse, unnötige Folgeaktionen	Viele harmlose richtige Antworten überdecken seltene schwere Fehler.

Dimension	Beispiele für geeignete Messgrößen	Typische Fehlinterpretation
Arbeitsablauf	Nettozeit einschließlich Kontrolle, Unterbrechungen, Reaktionszeit, Zusatzaufgaben anderer Teams	Eingesparte Schreibzeit wird als gesamte Arbeitersparnis ausgegeben.
Verteilung	Gruppenbezogene Ergebnisse, Zugangsbarrieren, ungleiche Belastung	Ein guter Durchschnitt gilt als Beleg für Fairness.
Selbstbestimmung und Beziehung	Verständnis der Rolle des Systems, erreichbare Menschen, Beschwerden, berichtete Wahlmöglichkeiten	Hohe Nutzung wird mit freiwilliger Zustimmung verwechselt.
Robustheit und Betrieb	Ausfälle, relevante Veränderungen, Wiederherstellungszeit, Abhängigkeiten	Eine einmalige Abnahme gilt als dauerhafte Sicherheit.

Diese Tabelle enthält eigene Messvorschläge, keine universellen Mindestwerte. Schwellen müssen aus Schadensschwere, Baseline und Ziel des jeweiligen Vorhabens begründet werden. Beobachtungsdauer und Stichprobengröße brauchen methodische Planung. „Kein beobachteter Schaden“ in einem kleinen Pilot ist kein Beweis, dass seltene schwere Schäden ausgeschlossen sind.

Für sprachgenerierende Systeme sollten fachlich definierte Fehlertypen statt bloßer sprachlicher Gefälligkeit bewertet werden. Bei kritischen Inhalten sind nachvollziehbare Bewertungsregeln, geeignete unabhängige Beurteilungen und der Umgang mit Uneinigkeit zu planen. TRIPOD-LLM unterstützt die transparente Beschreibung solcher Untersuchungen, ersetzt aber keine Qualitätsprüfung.[29]

10.2 Lokales Evaluationsprotokoll

Zweck und Status des Vorschlags

Das folgende Verfahren ist ein eigener methodischer Vorschlag für Einrichtungen, die eine konkrete KI-Anwendung bewerten möchten. Es wurde nicht im Rahmen dieser Literaturstudie praktisch erprobt. Die Vorlage beschreibt deshalb, welche Entscheidungen ein lokales Untersuchungsteam begründen sollte; sie garantiert weder Sicherheit noch regulatorische Konformität. Fachliche Leitung, pflegerische Leitung beziehungsweise betroffener Unternehmensbereich, Methodik, Datenschutz, Informationssicherheit und Betroffenenvertretung müssen sie an den Einsatz anpassen. Bei patientenbezogenen Anwendungen ist zusätzlich zu klären, ob Forschung, Qualitätssicherung oder eine andere Tätigkeit vorliegt und welche Prüfungen erforderlich sind. Die Bezeichnung „Pilot“ beantwortet diese Frage nicht.

Untersucht wird nicht allein ein Modell, sondern eine festgelegte Anwendungsversion innerhalb eines Arbeitsablaufs. Zur Beschreibung gehören Zweck, Nutzende, betroffene Personen, Eingangsdaten, Schnittstellen, Zugriffsrechte, Ausgabeform und Entscheidungsspielraum. Änderungen an Modell, Wissensbestand, Benutzeroberfläche oder nachgelagerten Handlungen können den Prüfgegenstand verändern. DECIDE-AI und TRIPOD-LLM helfen, Untersuchungen nachvollziehbar zu beschreiben; die Orientierung an Berichtsempfehlungen ersetzt jedoch keine geeignete Studienplanung.[28,29]

Fragestellung und geeigneter Vergleich

Vor Beginn sollte eine überprüfbare Frage formuliert werden. Statt „Verbessert KI unsere Pflege?“ wäre beispielsweise zu untersuchen, ob eine festgelegte Dokumentationshilfe bei definierten Aufnahmegesprächen den gesamten personellen Dokumentationsaufwand reduziert, ohne die Vollständigkeit sicherheitsrelevanter Angaben zu verschlechtern. Im Konzern könnte der Vergleich betreffen, ob ein Assistenzsystem die fachlich korrekt abgeschlossenen Vorgänge beschleunigt, ohne Beschwerden oder ungleiche Zugangschancen zu erhöhen. Verbesserung, Schaden und betroffene Gruppen werden damit ausdrücklich benannt.

Der Vergleich muss den tatsächlich verfügbaren Alternativen entsprechen. Dazu können der bisherige Ablauf, eine verbesserte Schulung, zusätzliche personelle Unterstützung oder eine einfache nicht KI-basierte Automatisierung gehören. Ein absichtlich schwacher Vergleich würde den Nutzen überschätzen. Ebenso sollte die Kontrollgruppe nicht durch den Entzug bisher üblicher Unterstützung schlechter gestellt werden. Bei begrenzten Mitteln ist zusätzlich zu fragen, ob dieselbe Investition an anderer Stelle mehr Nutzen schaffen könnte. Die Evaluation prüft damit eine konkrete Entscheidung, nicht die abstrakte Überlegenheit einer Technologiekatgorie.

Ausgangslage und Endpunkte

Eine Baseline erfasst den bisherigen Prozess unter vergleichbaren Bedingungen. Der Beobachtungszeitraum sollte typische Schichten, Fallarten und relevante Schwankungen abdecken; eine universelle Zahl von Tagen ist dafür ungeeignet. Urlaubszeiten, Personalwechsel, ungewöhnliche Auslastung und parallel eingeführte Maßnahmen werden dokumentiert. Messung selbst kann Verhalten verändern. Deshalb sollten Instrumente und Beobachtungsintensität zwischen Vergleichsbedingungen möglichst ähnlich sein. Selbstauskünfte zur Entlastung sind wertvoll, ersetzen aber nicht die Erfassung tatsächlich aufgewendeter Arbeitszeit.

Das Protokoll legt vor der Auswertung einen primären Nutzenendpunkt und die für die Freigabe unverzichtbaren Sicherheitskriterien fest. Weitere klinische, soziale und wirtschaftliche Größen werden als sekundäre oder explorative Endpunkte bezeichnet. Eine günstige Nebenkennzahl darf einen verfehlten primären Endpunkt nicht nachträglich ersetzen. Bei mehreren gleichrangigen Zielen ist das Auswertungsverfahren entsprechend zu planen. Nichtunterlegenheit verlangt eine vorab sachlich begründete Grenze; ein nicht signifikanter Unterschied belegt sie nicht. Patientennutzen, Prozessqualität und wirtschaftlicher Effekt bleiben getrennte Aussagen.

Messwörterbuch für die Umsetzung

Die Tabelle ist ein vorgeschlagenes Mindestgerüst, kein universeller Pflichtkatalog. Jede Einrichtung wählt die für ihren Zweck relevanten Größen und ergänzt Instrument, Messzeitpunkt, Zuständigkeit und Umgang mit fehlenden Angaben.

Tabelle 9 Messwörterbuch für die lokale Evaluation

Messgröße	Vorgeschlagene operationale Definition	Wesentliche Interpretationsgrenze
Klinischer Nutzen	Vorab bestimmtes patientenrelevantes Ereignis oder geeignete Zustandsveränderung im festgelegten Zeitraum	Verbesserte Dokumentation ist nicht automatisch bessere Gesundheit.
Sicherheitsrelevanter Fehler	Fachlich definierter Fehlertyp, nach Schweregrad beurteilt; Anzahl je prüfbarem Vorgang	Gemeldete Ereignisse allein unterschätzen möglicherweise unentdeckte Fehler.
Alarmbelastung	Alarmer und erforderliche Reaktionen je Schicht oder Fallzahl, einschließlich Bearbeitungszeit	Hohe Sensitivität kann mit erheblicher Belastung durch Fehlalarme einhergehen.
Nettoarbeitszeit	Gesamte menschliche Bearbeitungszeit einschließlich Prüfung, Korrektur, Rückfragen und Betriebsaufwand	Eine kürzere Einzeltätigkeit kann Arbeit an andere Berufsgruppen verlagern.
Beziehungsqualität	Vorab gewähltes geeignetes Befragungsinstrument, ergänzt um offene Erfahrungen	Zufriedenheit, Kontaktzeit und erlebte Zuwendung sind unterschiedliche Größen.
Verteilungswirkung	Nutzen, Fehler oder Zugangsquote in begründeten Gruppen, jeweils mit Nenner und Unsicherheit	Kleine Gruppen erlauben oft keine präzise Entwarnung.
Menschliche Eingriffsfähigkeit	Erkennen relevanter Fehler und erfolgreicher Eingriff in definierten Testsituationen	Eine hohe Übersteuerungsquote kann gutes Eingreifen oder schlechte Systemqualität bedeuten.
Ökonomischer Nettoeffekt	Kosten und nutzbare Kapazitätsgewinne im definierten Zeitraum aus festgelegter Perspektive	Freie Minuten sind nicht ohne Weiteres auszahlbare Einsparungen.
Kompetenzentwicklung	Wiederholte, vergleichbare Aufgaben mit und ohne Assistenz; begründete Leistungsmaße	Lerneffekte und Personalwechsel können Veränderungen erklären.
Akzeptanz und Zugang	Nutzung, begründete Ablehnung, Beschwerden und Verfügbarkeit einer geeigneten Alternative	Hohe Nutzung kann organisatorischen Druck statt freiwillige Zustimmung widerspiegeln.

Datenschutz und Qualität der Messung

Der Evaluationsdatensatz sollte nur Informationen enthalten, die für die festgelegten Fragen benötigt werden. Zugriff, Aufbewahrung, Löschung und zulässige Weiterverwendung werden vorab geklärt. Wo möglich werden Auswertungsdaten von unmittelbaren Identifikatoren getrennt. Pseudonymisierung ist nicht mit Anonymität gleichzusetzen. Die Übertragung vertraulicher Inhalte in ein zusätzliches Auswertungssystem schafft eigene Risiken und darf nicht beiläufig erfolgen. Besonders problematisch wäre ein Versuch, Ethik durch unverhältnismäßige Überwachung von Beschäftigten zu messen.

Gruppenbezogene Prüfungen erfordern eine eigenständige Begründung. Datenschutz und Fairness dürfen weder gegeneinander ausgespielt noch durch das wahllose Sammeln sensibler Merkmale scheinbar gelöst werden. Datenlücken und nicht ausreichend vertretene Gruppen gehören ausdrücklich in den Ergebnisbericht. STANDING Together bietet hierfür eine Orientierung zur Beurteilung von Datengrundlagen.[30] Für fachliche Fehler sollten nachvollziehbare Bewertungsregeln, möglichst unabhängige Beurteilungen und ein Verfahren bei Uneinigkeit festgelegt werden. Soweit praktikabel erfolgt die Beurteilung ohne Kenntnis der Gruppenzuordnung. Automatisch erzeugte Qualitätsbewertungen dürfen nicht ungeprüft als objektiver Referenzmaßstab dienen.

Designoptionen und ihre Grenzen

Ein vorgeschalteter Schattenbetrieb kann untersuchen, welche Ausgaben ein System unter realistischen Eingangsdaten erzeugt, ohne Entscheidungen direkt zu beeinflussen. Damit lassen sich technische und fachliche Auffälligkeiten erkennen. Nicht beurteilt werden jedoch die tatsächlichen Reaktionen der Nutzenden, zusätzliche Rückfragen oder der Nutzen im vollständigen Ablauf. Ein erfolgreicher Schattenbetrieb ersetzt deshalb keine Untersuchung der Mensch-System-Interaktion.

Für manche Anwendungen bietet sich ein randomisierter Vergleich an. Individuelle Randomisierung kann ungeeignet sein, wenn Nutzende zwischen Bedingungen wechseln und Erfahrungen übertragen. Eine Randomisierung von Teams, Stationen oder Standorten reduziert bestimmte Vermischungen, verlangt aber die Berücksichtigung der Ähnlichkeit von Beobachtungen innerhalb dieser Einheiten. Ein gestaffelter randomisierter Rollout kann organisatorisch hilfreich sein, ist jedoch gegenüber zeitlichen Veränderungen empfindlich; er setzt außerdem voraus, dass die geplante Einführung ethisch vertretbar bleibt. Ein Wechselvergleich kann an fortbestehenden Lern- oder Gewöhnungseffekten scheitern.

Wenn Randomisierung nicht möglich ist, kann ein gleichzeitig beobachteter, hinreichend vergleichbarer Kontrollbereich die Aussagekraft verbessern. Verbleibende Unterschiede werden dadurch nicht automatisch beseitigt. Eine unterbrochene Zeitreihe benötigt ausreichend wiederholte Messpunkte und eine begründete Behandlung zeitlicher Trends. Ein einfacher Vorher-nachher-Vergleich ist leichter umzusetzen, trennt Systemwirkung jedoch häufig schlecht von Personal-, Saison- oder Organisationsveränderungen. Qualitative Interviews und Beobachtungen können in allen Designs erklären, warum ein Ablauf funktioniert oder scheitert. Sie sind kein Ersatz für eine Effektabschätzung, aber auch nicht deren bloßes Beiwerk.

Stichprobe und Auswertung

Die Zahl der Fälle wird aus Fragestellung, erwarteter Ausgangshäufigkeit, relevanter Veränderung, gewünschter Präzision und gewählttem Design abgeleitet. Hinzu kommen Ausfälle, wiederholte Messungen und gegebenenfalls Clusterstrukturen. Eine pauschale Empfehlung wie „hundert Fälle reichen“ wäre wissenschaftlich nicht begründbar. Seltene schwere Schäden können auch in einer sorgfältigen lokalen Evaluation unbeobachtet bleiben. Geplante Teilgruppenanalysen benötigen eigene Überlegungen zur Aussagekraft; fehlende statistische Signifikanz bei wenigen Personen rechtfertigt keine Gleichheitsaussage.

Vor Beginn sollten Auswertungsplan, Umgang mit fehlenden Daten und Abweichungen vom vorgesehenen Einsatz festgehalten werden. Abbruchkriterien orientieren sich an Schwere und Plausibilität möglicher Schäden, nicht allein an einer später berechneten Erfolgsquote. Ergebnisse werden mit geeigneten absoluten Angaben, Bezugsgrößen und Unsicherheitsintervallen berichtet. Nachträgliche Untersuchungen sind als explorativ zu kennzeichnen. Für Entscheidungen kann entscheidender sein, welche nachteiligen Effekte noch mit den Daten vereinbar sind, als ob ein einzelner p-Wert eine Schwelle unterschreitet.

Hypothetisches Rechenbeispiel zur Nettoentlastung

Das folgende Beispiel ist vollständig erfunden und dient ausschließlich der Erklärung. Es enthält weder Messdaten einer Einrichtung noch Leistungsangaben eines Produkts. Angenommen werden 400 gleichartige Dokumentationsvorgänge pro Woche. Bisher benötigt ein vollständig bearbeiteter und geprüfter Vorgang zehn Minuten. Der Ausgangsaufwand beträgt damit 4.000 Minuten, also 66 Stunden und 40 Minuten.

Mit Assistenz entfallen je Vorgang vier Minuten auf Erstellung beziehungsweise Übernahme und drei Minuten auf die menschliche Prüfung. Bei angenommenen 20 Prozent der Vorgänge kommen fünf Minuten Nachbearbeitung hinzu.

Im Durchschnitt sind das $4 + 3 + 0,20 \times 5 = 8$ Minuten. Für 400 Vorgänge ergeben sich 3.200 Minuten. Zusätzlich werden in diesem Beispiel wöchentlich insgesamt 300 Minuten für Betrieb, zusätzliche Qualitätssicherung und anteilig berücksichtigte Schulung angesetzt. Der neue Gesamtaufwand beträgt 3.500 Minuten. Die hypothetische Nettoentlastung liegt folglich bei 500 Minuten oder acht Stunden und zwanzig Minuten, entsprechend 12,5 Prozent des Ausgangsaufwands.

Die Rechnung reagiert empfindlich auf ihre Annahmen. Bei 45 Prozent nachzubearbeitenden Vorgängen entsteht unter unveränderten übrigen Bedingungen kein Zeitgewinn mehr. Bei 60 Prozent beträgt der neue Aufwand 4.300 Minuten; die Anwendung verursacht dann fünf zusätzliche Arbeitsstunden. Keine dieser Schwellen ist allgemein gültig. Selbst acht freie Stunden erlauben noch keine Aussage über Stellenabbau oder Geldersparnis. Entscheidend ist, bei wem, wann und in welcher nutzbaren Form Zeit entsteht und ob die fachliche Qualität mindestens ausreichend bleibt.

Entscheidung und laufende Nachprüfung

Die Freigabe sollte auf einer begrenzten, nachvollziehbar beschriebenen Verwendung beruhen. Ergebnisbericht und Entscheidung dokumentieren auch verbleibende Unsicherheit, ausgeschlossene Einsatzbereiche, offene Datenlücken und Zuständigkeiten. Nutzen und Schaden werden nicht in einem beliebig gewichteten Gesamtscore versteckt. Ein wirtschaftlicher Vorteil kann beispielsweise eine unvertretbare Gefährdung nicht durch bloße Verrechnung aufheben.

Nach Einführung werden relevante Qualitäts- und Belastungsgrößen weiter beobachtet. Der Aufwand dafür gehört in die Wirtschaftlichkeitsbetrachtung. Neue Versionen, veränderte Fallgruppen oder zusätzliche Handlungsrechte lösen eine begründete Neubewertung aus. Bei Auffälligkeiten muss eine praktikable Einschränkung oder Rückkehr zum alternativen Verfahren möglich sein. Das Verfahren endet deshalb nicht mit einer einmaligen Unterschrift.

11 Limitationen und Forschungsagenda

11.1 Grenzen der vorliegenden Untersuchung

Diese Arbeit ist eine selektive, evidenzinformierte Literatur- und Dokumentenanalyse mit eigenständig vorgeschlagenen Handlungsempfehlungen. Sie beruht nicht auf einer vorregistrierten systematischen Suche und beansprucht keine vollständige Erfassung aller relevanten Studien. Dokumentierte Ergänzungsabfragen erhöhen die Nachprüfbarkeit, beseitigen aber keine Auswahlverzerrung. Es erfolgte kein unabhängiges Doppelscreening. Suchmaschinenranking, Sprachverfügbarkeit und der unterschiedliche Zugang zu Volltexten können die Auswahl beeinflusst haben.

Die betrachteten Anwendungen, Populationen und Endpunkte sind zu unterschiedlich, um aus ihnen einen gemeinsamen „KI-Effekt“ zu berechnen. Eine nachträglich zusammengesetzte Studienanzahl wäre ebenfalls kein Qualitätsmaß. Nicht alle Interessenkonflikte, Protokollabweichungen oder veröffentlichten Korrekturen konnten über sämtliche Quellen hinweg in gleicher Tiefe beurteilt werden. Es wurde keine vollständige formale Verzerrungsbewertung für jedes Studiendesign durchgeführt. Die Orientierung an SANRA, PRISMA und SWiM beschreibt Transparenzanforderungen; sie ist weder Zertifizierung noch Beleg einer vollständig entsprechend durchgeführten systematischen Übersicht.[38–40]

Viele empirische Befunde betreffen begrenzte Zeiträume oder spezifische Arbeitsumgebungen. Übertragbarkeit auf deutsche Kliniken, Pflegeeinrichtungen oder andere Unternehmen muss jeweils begründet werden. Langfristige Kompetenz-, Beschäftigungs- und Beziehungswirkungen bleiben in dieser Auswahl unzureichend erfasst. Aus dieser Lücke folgt keine Aussage darüber, ob solche Effekte außerhalb des gesichteten Materials belegt sind. Ebenso sind normative Wertentscheidungen nicht allein aus Studienresultaten ableitbar. Die vorliegenden Empfehlungen benötigen fachliche Diskussion und lokale Legitimation. Das Manuskript ist für externe fachliche Begutachtung vorgesehen, nicht bereits begutachtete Originalforschung.

11.2 Prioritäten weiterer Forschung

Priorität sollten Untersuchungen erhalten, die Systemleistung mit Folgen für Menschen verbinden. In Kliniken betrifft das patientenrelevante Endpunkte ebenso wie die Verteilung zusätzlicher Kontrollarbeit zwischen Berufsgruppen.

Pflegewissenschaftliche Kompetenz und die Perspektive pflegerischer Leitung sollten nicht erst bei der Interpretation fertiger Ergebnisse hinzukommen. Bei sozialen Robotern sind freiwillige Nutzung, Beziehungserleben, menschliche Kontaktmöglichkeiten und unerwünschte Bindungs- oder Ausschlusseffekte gemeinsam zu untersuchen.

Für Unternehmen fehlen im hier ausgewerteten Material hinreichend langfristige Antworten darauf, wie Assistenz Kompetenzentwicklung, Arbeitsintensität und berufliche Chancen verändert. Künftige Untersuchungen sollten deshalb Nutzungsdauer, Aufgabenverlagerung und Veränderungen ohne Assistenz mitberücksichtigen. Besonders wertvoll wären unabhängige Replikationen, vergleichbare Endpunkte und Studien mit unterschiedlichen Standorten. Auch negative oder uneindeutige Ergebnisse sowie abgebrochene Einführungen sollten zugänglich werden. Lokale Projekte können hierzu beitragen, wenn Fragestellung, Vorgehen und Grenzen transparent dokumentiert werden, statt lediglich eine gelungene Einführung zu präsentieren.

12 Schlussfolgerung

Verantwortbarer KI-Einsatz verlangt keine pauschale Entscheidung für oder gegen Technik. Er verlangt eine begründete Entscheidung über eine bestimmte Anwendung unter bestimmten Bedingungen. Dafür müssen Nutzen, Risiken und Verteilungswirkungen beobachtbar sein; Verantwortung und Eingriffsmöglichkeiten müssen im Alltag tatsächlich bestehen. Ein professionelles Evaluationsverfahren macht Unsicherheit nicht unsichtbar, sondern bearbeitbar. Genau darin liegt der praktische Wert dieses Vorschlags für Ethikverantwortliche, pflegerische und ärztliche Leitung sowie die Unternehmensführung.

Antwort auf F1. Die untersuchten Anwendungen liefern unterschiedliche Arten von Nutzen- und Risikohinweisen. Das rechtfertigt weder ein universelles Verbot noch eine pauschale Freigabe. Aus diesem ausgewählten Bestand lässt sich keine Erfolgsquote aller KI-Systeme ableiten. Eine Entscheidung muss genau den Nutzen benennen, für den ein geeigneter Nachweis vorliegt; benachbarte, nicht untersuchte Vorteile bleiben Hypothesen.

Antwort auf F2. Übertragbarkeit ist eine zu begründende Annahme. Zu vergleichen sind mindestens Zielgruppe, Aufgabe, Systemversion, Einbindung, vorhandene Fachkompetenz und die tatsächliche Alternative. Besonders wichtig ist die Reaktionsfähigkeit der Einrichtung. Ein nützlicher Hinweis kann ohne rechtzeitige Bearbeitung wirkungslos bleiben; eine zusätzliche Kontrolle kann andere notwendige Arbeit verdrängen. Lokale Prüfung soll diese Bedingungen klären, nicht lediglich einen allgemeinen Herstellerwert wiederholen.

Antwort auf F3. Rechte, Selbstbestimmung, Gerechtigkeit und Fürsorge begrenzen eine reine Leistungsrechnung. Das gilt auch dann, wenn sich eine Anwendung wirtschaftlich lohnt. Umgekehrt sind Nichtstun und der Verzicht auf eine hilfreiche Technik nicht automatisch die ethisch bessere Wahl. Maßgeblich ist ein offener Vergleich der verfügbaren Handlungsoptionen einschließlich ihrer Belastungen, Ausschlüsse und Opportunitätskosten.

Antwort auf F4. Empfohlen wird eine institutionelle Aufgabenverteilung mit benannten Entscheidern, ausreichenden Ressourcen und dokumentiertem Dissens. Ethikberatung strukturiert Konflikte; sie ersetzt weder die Fachprüfung noch die Verantwortung der Leitung. Die pflegerische Leitung gestaltet Arbeitsablauf, Reaktionskapazität und Beziehungsschutz mit. Im Konzern müssen die jeweils betroffenen Geschäftsbereiche und Interessenvertretungen entsprechend wirksam einbezogen werden.

Antwort auf F5. Eine begründete Freigabe bleibt überprüfbar und veränderbar. Sie benötigt geeignete Ergebnisdaten, verständliche Beschwerdewege, eine zuständige Stelle für Änderungen und eine tatsächlich nutzbare Alternative bei Ausfall oder unvertretbarer Leistung. Die Empfehlungen dieses Berichts sind dafür eine transparente Arbeitsgrundlage. Ihre Wirksamkeit als Gesamtverfahren ist noch zu untersuchen. Der nächste wissenschaftliche Schritt ist daher keine zusätzliche Behauptung von Sicherheit, sondern eine unabhängige, prospektive Prüfung ihrer Umsetzung und Folgen.

13 Informationswege und fachliche Unterstützung

Die folgende Wegweisung soll Informationsarbeit verkürzen. Kein einzelnes Dokument deckt alle Fragen ab. Bei Internetseiten sind Fassungsdatum und Rechtsstand zu prüfen; ein offizieller Herausgeber garantiert nicht, dass jede Unterseite bereits an die neueste Gesetzesänderung angepasst wurde.

Tabelle 10 Informationswege und fachliche Anlaufstellen

Konkrete Frage	Geeigneter Einstieg	Was damit zu tun ist – und was er nicht leistet
Wie sollte mein Ethikmandat ausgestaltet sein?	AEM: Standards und Empfehlungen [14,15]	Geschäftsordnung, Kompetenz, Unabhängigkeit und Ressourcen abgleichen; keine KI-Zulassung.
Was verlangt personenorientierte Pflege?	ICN-Ethikkodex, Volltext, besonders 1.11 [12]	Pflegeethische Ziele in lokale Anforderungen übersetzen; der Kodex ersetzt kein nationales Recht.
Wie verteilen sich klinische Verantwortungen?	ZEKO: KI-Entscheidungsunterstützung [13]	Rollen und fachliche Verantwortung diskutieren; Rechtsangaben aus 2021 gesondert aktualisieren.
Welche ethischen Aspekte werden leicht übersehen?	WHO: KI-Governance und generative Modelle [1,2]	Lokale Prüfung auf Autonomie, Gerechtigkeit und weitere Folgen kontrollieren; keine Einzelproduktbewertung.
Welche KI-Pflichten gelten wann?	AI Act bei EUR-Lex und Änderungsverordnung 2026/1744 [20,21]	Rechtsakt, Änderungen, Geltungsbereich und Anwendungsdatum zusammen lesen.
Welche deutsche Stelle ist zuständig?	KI-MIG § 2 und zentrale Anlaufstelle § 6 [26]	Zuständigkeit anhand System und Betreiber klären; nicht jede Anfrage gehört allein an BNetzA oder BfArM.
Wie prüfe ich personenbezogene Datenverarbeitung?	DSK: Orientierungshilfen [18,19]	Datenflüsse und Prüffragen mit Datenschutzbeauftragten durcharbeiten; konkrete Aufsicht je nach Träger klären.
Handelt es sich um ein Medizinprodukt?	BfArM: rechtlicher Status und Klassifizierung, MDR und MDCG [23,35,36]	Zweckbestimmung und Einstufung fachkundig prüfen; nicht nur auf die Vertriebsbezeichnung vertrauen.
Wie werden Risiken von Medizinprodukten gemeldet?	BfArM: Risiken melden [35]	Den einschlägigen Meldeweg in die internen Vorfallsprozesse integrieren.
Wie strukturiere ich KI-Risikomanagement?	NIST AI RMF und GenAI-Profil [16,17]	Aufgaben zuordnen und Nachweise planen; freiwillige Rahmenwerke sind kein EU-Konformitätsnachweis.
Welche technischen Angriffsmöglichkeiten sind zu prüfen?	BSI: Künstliche Intelligenz [34]	IT-Sicherheit mit konkreter Architektur und Rechten befassen; kein allgemeiner Prompt ersetzt Sicherheitstests.
Wie prüfe ich die Qualität einer Studie?	DECIDE-AI, TRIPOD-LLM, STANDING Together [28–30]	Bericht und Datengrundlage gezielt prüfen; Checklisten sind keine Gütesiegel. Biometrische und fachliche Beratung hinzuziehen.
Was belegt ein KI-Zertifikat?	ISO/IEC 42001: Erläuterung [31]	Geltungsbereich und Ausschlüsse des Zertifikats prüfen; Produktnutzen zusätzlich belegen lassen.

Für eigene Literaturrecherchen ist PubMed ein sinnvoller Einstieg in medizinische Studien. Eine Suchanfrage sollte nicht nur „AI ethics“ enthalten, sondern konkrete Aufgabe, Produkt oder Modell, Population und Evaluationstyp. Beispielhaft: ("artificial intelligence" OR "machine learning") AND sepsis AND ("external validation" OR randomized). Produktnamen und aktuelle Versionen sind ergänzend zu suchen.

Anschließend sind Originalarbeit, Studienregistrierung soweit vorhanden, Finanzierung, Korrekturen und spätere Untersuchungen zu prüfen. Abstracts helfen beim Auffinden, reichen aber für anspruchsvolle Beschaffungsentscheidungen häufig nicht aus. Wenn die eigene Einrichtung diese Beurteilung nicht leisten kann, ist methodische Unterstützung durch Biometrie, klinische Epidemiologie oder fachlich geeignete externe Einrichtungen sinnvoll. Das ist kein Kompetenzmangel der Ethikstelle, sondern Teil einer guten Aufgabenverteilung.

A Arbeitsvorlagen

Die folgenden Vorlagen sind eigene Organisationsvorschläge. Sie werden erst durch die jeweils zuständige Einrichtung und im Rahmen des geltenden Rechts zu internen Vorgaben.

A.1 Arbeitsstruktur für die ersten 90 Tage

Der Zeitraum ist ein Planungsbeispiel, keine gesetzliche Frist und kein Versprechen, innerhalb von 90 Tagen jede Anwendung freigeben zu können.

Tabelle 11 Arbeitsstruktur der ersten 90 Tage

Zeitraum	Priorität	Greifbares Ergebnis
Tage 1–15	Mandat, Ansprechpartner und bereits verwendete Systeme erfassen.	Vorläufiges Register; verantwortliche Leitung; erreichbarer Meldeweg.
Tage 16–30	Anwendungen nach Folgen, Daten, Autonomie und Reichweite priorisieren.	Begründete Prüfreihefolge; sofortiger Umgang mit dringlichen Sicherheits- oder Rechtsproblemen.
Tage 31–60	Für priorisierte Anwendungen Evidenz, Rechte, Arbeitsabläufe und Betroffenenperspektiven prüfen.	Entscheidungsunterlagen; gegebenenfalls methodisch geplanter, zulässiger Pilot.
Tage 61–90	Erste Entscheidungen dokumentieren; Vorfalls- und Änderungsprozesse erproben.	Begrenzte Freigaben oder begründete Zurückstellungen; Prüfkalendar; Ausfallübung.

Ein laufender Einsatz mit ernstem Schadensverdacht wartet nicht bis Tag 61. Er benötigt unverzügliche Bewertung und gegebenenfalls eine sichere Aussetzung mit Ersatzverfahren. Umgekehrt darf der Zeitplan nicht dazu führen, dass unzureichende Daten aus einem kurzen Pilot als ausreichender Sicherheitsnachweis erklärt werden.

A.2 Fragen vor einer Beschaffung

1. Welchen genau abgegrenzten Zweck decken Produkt und Nachweise ab, und welche Verwendung ist ausdrücklich ausgeschlossen?
1. Welche Modell-, Software- und Datenversion wurde untersucht, und entspricht sie der angebotenen Konfiguration?
1. Welche unabhängige oder externe Evaluation liegt vor; wie wurden Vergleichsgruppe und relevante Endpunkte gewählt?
1. Welche bekannten Fehler, gefährlichen Auslassungen und Grenzen bestehen, auch bei seltenen Situationen?
1. Wie variiert die Leistung zwischen relevanten Gruppen und Standorten; wo fehlen belastbare Daten?
1. Welche personenbezogenen und vertraulichen Informationen werden verarbeitet, weitergegeben, gespeichert oder für zusätzliche Zwecke verwendet?
1. Welche Handlungen kann das System selbst ausführen, und wie werden Zugriffe, fremde Inhalte und Manipulationsversuche begrenzt?
1. Welche konkrete Arbeit verbleibt bei Nutzern, wer bearbeitet Warnungen, und welcher Schulungs- und Kontrollaufwand entsteht?
1. Welche Protokolle, Auswertungen und Informationen über Fehler erhalten Betreiber tatsächlich?
1. Wie werden Modell- und Produktänderungen angekündigt, bewertet und gegebenenfalls zurückgenommen?
1. Welche vertraglichen Regelungen bestehen zu Unterstützung, Vorfällen, Prüfzugang, Verfügbarkeit, Datenexport und Vertragsende?
1. Welche sichere Alternative steht bei Ausfall, unvertretbarer Leistung oder Anbieterwechsel zur Verfügung?

Fehlende Antworten sind nicht automatisch ein endgültiges Verbot. Sie bestimmen aber, welche Unsicherheit noch besteht und ob sie für den konkreten Zweck vertretbar ist. Verbotene Verwendungen oder fehlende zwingende

Voraussetzungen lassen sich nicht durch einen hohen erwarteten wirtschaftlichen Nutzen ausgleichen.

A.3 Entscheidungsvorlage

Eine angemessen kurze Vorlage sollte folgende Felder enthalten:

- **Anwendung:** Bezeichnung, konkrete Aufgabe, Standort, Version, Nutzerkreis, Verantwortliche; rechtliche Rolle der Einrichtung sowie relevante Eigenentwicklungen oder Veränderungen.
- **Vergleich und Ziel:** Ausgangslage, nichttechnische Alternativen, erwarteter menschlicher Nutzen.
- **Betroffene:** direkt und indirekt betroffene Gruppen, Beteiligung, wesentliche Einwände.
- **Evidenz:** wichtigste Nachweise, lokale Übertragbarkeit, ungeklärte Fragen, Interessenkonflikte.
- **Recht und Daten:** einschlägige Einordnung, geprüfte Grundlagen, erforderliche Beteiligungen und Folgenabschätzungen.
- **Sicherheit und Betrieb:** Fehlerfolgen, Rechte, Aufsicht, Kapazitäten, Ausfall- und Ersatzverfahren.
- **Entscheidung:** freigegeben, begrenzt erproben, zurückgestellt oder abgelehnt; jeweils mit Begründung.
- **Bedingungen:** erlaubte Verwendung, Ausschlüsse, Messgrößen, Schwellen, Zuständigkeit für Eingriffe.
- **Fortschreibung:** nächster Prüftermin und anlassbezogene Auslöser.
- **Dokumentation:** Entscheidungsträger, abweichende Voten, Datum und zugängliche Fassung.

Ein Freigabedokument sollte beispielsweise nicht nur „Ethik geprüft“ festhalten, sondern den tatsächlichen Konflikt und seine Bearbeitung benennen: etwa, warum eine zusätzliche Warnbelastung angesichts eines erwarteten Nutzens vertretbar ist und wie sie begrenzt werden soll.

A.4 Auslöser einer Neubewertung

Tabelle 12 Auslöser einer Neubewertung

Beobachtung	Empfohlenes Vorgehen
Plausibler Verdacht auf schweren Schaden oder eine verbotene Funktion	Unverzügliche qualifizierte Bewertung, gegebenenfalls sichere Aussetzung; einschlägige Meldepflichten prüfen und erfüllen.
Datenzugriff oder Weitergabe außerhalb des vorgesehenen Rahmens	Sicherheits- und Datenschutzprozess aktivieren; Zugriff begrenzen, Sachverhalt sichern und bewerten.
Relevante Leistungsver schlechterung oder neue Benachteiligung	Ursachenanalyse und begründete Begrenzung, Anpassung oder Pause nach Schadenspotenzial.
Warnungen bleiben mangels Personal unbearbeitet	Arbeitsprozess und Kapazität ändern; Nutzenannahme und Freigabe neu bewerten.
Modell, Zweck oder Nutzerkreis ändert sich wesentlich	Änderung nicht stillschweigend von der alten Freigabe abdecken lassen.
Keine sichere Versorgung oder Arbeit bei Ausfall möglich	Abhängigkeit reduzieren und tragfähigen Ersatz vorbereiten, bevor weitere Ausweitung erfolgt.

Die konkrete Reaktion sollte verhältnismäßig sein und Risiken eines abrupten Abschaltens einbeziehen. Gesetzliche Melde- und Handlungspflichten gehen vor; die Tabelle behauptet keine einheitliche Rechtsfolge für alle Fälle.

A.5 Glossar

Sensitivität: Anteil tatsächlich positiver Fälle, die das System erkennt. **Spezifität:** Anteil tatsächlich negativer Fälle, die es korrekt nicht als positiv markiert. **Positiver Vorhersagewert:** Anteil tatsächlich positiver Fälle unter allen positiven Meldungen; stark von der Häufigkeit des Zustands abhängig. **AUROC:** Maß für die Rangtrennung positiver und negativer Fälle über verschiedene Schwellen; kein direkter Nachweis klinischen Nutzens. **Kalibrierung:** Übereinstimmung vorhergesagter Wahrscheinlichkeiten mit beobachteten Häufigkeiten. **Externe Validierung:** Prüfung an Daten außerhalb der Entwicklungsgrundlage. **Automation Bias:** unangemessene Orientierung an

automatisierten Hinweisen. **RAG:** Abruf von Dokumenten oder Informationen zur Ergänzung des Modellkontexts; keine Wahrheitsgarantie. **DSFA:** Datenschutz-Folgenabschätzung. **FRIA:** Grundrechte-Folgenabschätzung nach dem einschlägigen KI-Rechtsrahmen. **Betreiber:** eine rechtliche Rolle, deren konkrete Zuordnung zu prüfen ist; nicht automatisch die Person, die eine ethische Stellungnahme verfasst.

B Rechercheprotokoll der empirischen Vertiefung

B.1 Zugang zu den empirischen Originalarbeiten

Die folgende Liste dokumentiert erneut geprüfte öffentliche Originalzugänge. „Volltext“ bedeutet zugänglicher Hauptartikel und Prüfung relevanter Methoden-, Ergebnis- und Offenlegungsabschnitte, nicht vollständige Begutachtung aller Zusatzdateien. Technische Nichterreichbarkeit ist keine Feststellung, dass der Text generell nicht frei verfügbar ist.

Tabelle 13 Zugang zu den empirischen Originalarbeiten

Quelle	Tatsächlich geöffnete URL	Ergebnis
[3]	Originalzugang	HTTP 403; kein Volltext
[3]	Originalzugang	Autorenabstract und bibliographische Angaben
[4]	Originalzugang	Volltext
[5]	Originalzugang	Volltext einschließlich Korrekturhinweis
[6]	Originalzugang	Volltext
[7]	Originalzugang	HTTP 403
[7]	Originalzugang	Direkter Abruf: Prüfseite; Originalabschnitte über Suchindex
[7]	Originalzugang	Technisch nicht erreichbar
[7]	Originalzugang	HTTP 402; kein Volltext
[8]	Originalzugang	Öffentliche Original-PDF im Universitätsrepositorium
[9]	Originalzugang	Öffentliche Original-PDF bei US-Behörde
[10]	Originalzugang	Finale Originalpublikation auf Autorinnenwebsite
[11]	Originalzugang	Verlagsvolltext
[41]	Originalzugang	Offener Verlagsvolltext; DOI 10.1038/s41591-026-04503-6

B.2 Ergänzende Volltextsuche

Es wurde der Websuchdienst mit seinen beiden Suchsystemen genutzt, keine vollständige MEDLINE-, Embase- oder Studienregistersuche. Nachstehend stehen die exakt abgesendeten Abfragen dieser ergänzenden Zugriffsrunde. Bei drei Abfragen meldete der Dienst ausdrücklich eine interne Umformulierung; deren tatsächlich intern ausgeführte Form wurde nicht offengelegt. Treffer- oder Ausschlusszahlen werden nicht behauptet.

Suchabfrage 1. System: 1; Abgesendete Abfrage: "AIoa2501000" "Funding" "conflicts"; Zweck/Ergebnis: Originaloffenlegungen gesucht; keine ausreichende direkte Volltextausbeute

Suchabfrage 2. System: 1; Abgesendete Abfrage: "S0140-6736(25)02464-X" pdf; Zweck/Ergebnis: MASAI-Volltext gesucht; kein zusätzlich gelesener Volltext

Suchabfrage 3. System: 2; Abgesendete Abfrage: site:pmc.ncbi.nlm.nih.gov/articles/PMC12768499 "Funding" "Disclosures"; Zweck/Ergebnis: Unzureichend passende Treffer; nicht als Datengrundlage genutzt

Suchabfrage 4. System: 2; Abgesendete Abfrage: site:pmc.ncbi.nlm.nih.gov/articles/PMC12768499 "9.5" "238"; Zweck/Ergebnis: Originaleintrag auffindbar, daneben sachfremde Treffer

Suchabfrage 5. System: 2; Abgesendete Abfrage: site:pmc.ncbi.nlm.nih.gov/articles/PMC12768499 "registration" "January"; Zweck/Ergebnis: Keine belastbare Registerhistorie erhalten

Suchabfrage 6. System: 1; Abgesendete Abfrage: "Ambient AI Scribes in Clinical Practice" "Funding"; Zweck/Ergebnis: Vom Dienst umformuliert; Original- und Sekundärtreffer

Suchabfrage 7. System: 1; Abgesendete Abfrage: "Ambient AI Scribes in Clinical Practice" "9.5" "238"; Zweck/Ergebnis: Vom Dienst umformuliert; Originalauszüge, Ovid-Nachweis

Suchabfrage 8. System: 1; Abgesendete Abfrage: "Ambient AI Scribes in Clinical Practice" "randomization" "blinded"; Zweck/Ergebnis: Vom Dienst umformuliert; keine eigenständige Verblindungsprüfung

Suchabfrage 9. System: 2; Abgesendete Abfrage: "Ambient AI Scribes in Clinical Practice: A Randomized Trial" "Supported"; Zweck/Ergebnis: Indexierte PMC-Originalabschnitte einschließlich Förderangaben

Suchabfrage 10. System: 2; Abgesendete Abfrage: "Ambient AI Scribes in Clinical Practice: A Randomized Trial" "9.5%"; Zweck/Ergebnis: Originalergebnis in PubMed/PMC gegengeprüft

Suchabfrage 11. System: 2; Abgesendete Abfrage: "Ambient AI Scribes in Clinical Practice: A Randomized Trial" "METHODS"; Zweck/Ergebnis: Originalabschnitte zu Zuteilung, Produkten und Durchführung

Innerhalb der geöffneten Originaltexte wurden zusätzlich die Suchbegriffe Funding/Support, Funding, acknowledge, ACKNOWLEDGMENTS, Disclosures, Limitations, TABLE II, 43,539, 19.8 und blind verwendet. Für [41] erfolgten ein direkter URL-Aufruf, die interne Suche nach OpenAI, Competing interests und Statistical analysis sowie direkte Abschnittsaufrufe; keine zusätzliche freie Webabfrage. Ergänzend wurden passende Textabschnitte direkt geöffnet. Die Auswertung berichtet keine unabhängige Doppelbegutachtung durch menschliche Reviewer und keine vollständige Reanalyse der Originaldaten.

Recherchetag: 24.09.2026. Zweck: gezielte Prüfung von Lücken des bereits vorhandenen, selektiv zusammengestellten Neun-Studien-Korpus. Dies ist eine dokumentierte Ergänzungssuche, **keine systematische Übersichtsarbeit**. Weder vollständige Datenbankabdeckung noch unabhängiges Doppelscreening, Präregistrierung oder formale Risk-of-Bias-Bewertung wurden durchgeführt. Die unten genannten Prüfentscheidungen erfolgten durch eine KI-gestützte redaktionelle Sichtung; eine fachwissenschaftliche Zweitprüfung steht aus.

B.3 Ergänzende Gegenrecherche

Es wurden vier Web-Suchabfragen ausgeführt. Das Datum gilt für alle vier. Die Suchmaschine lieferte zusammengeführte, rangierte Ergebnisse; deshalb werden weder eine vermeintliche Gesamttrefferzahl noch PRISMA-Ausschlusszahlen für alle Treffer behauptet.

Suchabfrage 1. ID: GS-01; Suchsystem: Websuche, System 2; Exakte Abfrage: 2025 2026 randomized trial human AI decision making assistance long term overreliance learning randomized study; Zweck / Filter: Mensch-KI-Interaktion, Arbeits- und Lerneffekte; keine Domainfilter

Suchabfrage 2. ID: GS-02; Suchsystem: Websuche, System 2; Exakte Abfrage: 2025 2026 care social robot randomized controlled trial nursing older adults loneliness long term; Zweck / Filter: Pflege, soziale Robotik, Beziehungsfolgen; keine Domainfilter

Suchabfrage 3. ID: GS-03; Suchsystem: Websuche, System 2; Exakte Abfrage: 2025 randomized controlled study AI chatbot emotional dependence loneliness psychosocial four week; Zweck / Filter: Psychosoziale Folgen und Abhängigkeit; keine Domainfilter

Suchabfrage 4. ID: GS-04; Suchsystem: Websuche, System 1; Exakte Abfrage: "Generative AI-enabled clinical decision support system in primary care" "14-day" trial 2026; Zweck / Filter: Gezielte Gegenprüfung eines

einschlägigen Treffers; angeforderte Domainfilter: nature.com, pubmed.ncbi.nlm.nih.gov

Die Domainfilter der vierten Abfrage verhinderten nicht sämtliche fachfremden Suchergebnisse; Aussagen wurden ausschließlich aus der Originalpublikation übernommen. Anschließend wurden Originalseiten und einzelne Textabschnitte geöffnet. Direktaufrufe sind keine zusätzlichen Suchabfragen. Die Aussagen über Prüfentscheidungen beziehen sich nur auf die fünf ausdrücklich aufgeführten Kandidaten, nicht auf jeden von der Suchmaschine ausgegebenen Treffer.

B.4 Auswahlentscheidungen zu fünf Kandidaten

G1 Aufgenommen

Agweyu et al. [41] wurde als zehnte Studie in den empirischen Kern aufgenommen. Begründung: ein vergleichender Versuch mit realem patientenbezogenem Endpunkt ergänzt den bisherigen Schwerpunkt auf Modellgüte, Arbeitsablauf und experimenteller Bearbeitung. Datenauszug und Bewertung stehen in Kapitel 4. Zugang: Originalvolltext.

G2 Soziale Robotik mit offenem Volltextzugang

Evaluating the Effectiveness of Digital Social Robots in Reducing Loneliness Among Community-Dwelling Older Adults in Japan: Randomized Controlled Trial and Qualitative Analysis. *JMIR Aging*. 2025;8:e74422. DOI: [10.2196/74422](https://doi.org/10.2196/74422). PubMed.

Zugang: Originalpublikations- und PubMed-Suchindex; wiederholter Volltextzugang scheiterte an Browserprüfung, die nicht umgangen wurde. Sichtbare Angaben: randomisierte Untersuchung mit qualitativer Ergänzung, 68 ältere, zuhause lebende Menschen, mittleres Alter 82,3 Jahre, 94 % Frauen. Intervention, Dauer, Instrument, Ausfälle und Auswertungsplan konnten in dieser Ergänzung nicht ausreichend verifiziert werden. **Deshalb diesen Zahlenbefund nicht ungeprüft in eine bewertete Evidenztabelle übernehmen.** Erkenntnisinteresse: relationale Wirkungen und mögliche positive Effekte sozialer Robotik. Eine Übertragung auf Demenz, stationäre Pflege oder Personalersatz wäre selbst bei bestätigtem Effekt nicht begründet. Der Konferenzabstract DOI 10.1093/geroni/igaf122.3049 erscheint eng verwandt; nicht als unabhängige zweite Studie zählen, solange Stichprobenidentität nicht geprüft ist.

G3 Patientennahe Laieninteraktion

Bean AM et al. Reliability of LLMs as medical assistants for the general public: a randomized preregistered study. *Nature Medicine*. 2026;32:609–615. DOI: [10.1038/s41591-025-04074-y](https://doi.org/10.1038/s41591-025-04074-y).

Zugang: Originalvolltext, Abstract und Studiendesign gelesen; Korrekturhinweis vom 17.04.2026 vorhanden, Korrektur nicht gesondert geprüft. Randomisierte Szenarienstudie mit 1.298 britischen Laien und zehn medizinischen Situationen. Modellleistung allein übertraf deutlich die Leistung der Personen im Dialog; KI-unterstützte Personen waren bei Zustandsidentifikation und Handlungsempfehlung nicht besser als die Kontrollgruppe. Kein reales Behandlungs- oder Schadensereignis. Ergänzt Goh um Laieninteraktion, aber der methodische Kern — Modellbenchmark ist nicht Teamleistung — ist bereits vertreten. Innerhalb der begrenzten Ergänzungsrecherche wurde die reale klinische Untersuchung G1 priorisiert. Bei künftigen Schwerpunkt Patienten-Chatbots ist G3 erneut zu prüfen, einschließlich Korrektur.

G4 Psychosoziale Folgen von Chatbotnutzung

Fang CM et al. How AI and Human Behaviors Shape Psychosocial Effects of Extended Chatbot Use: A Longitudinal Randomized Controlled Study. arXiv:2503.17473, **Version 2 vom 02.10.2025**, DOI [10.48550/arXiv.2503.17473](https://doi.org/10.48550/arXiv.2503.17473).

Zugang: aktuelle Original-Abstractseite und Versionsgeschichte. Vierwöchiges Experiment, n=981; randomisiert wurden Modalität und Gesprächsart. Das Abstract der zweiten Version meldet keine signifikanten Effekte dieser experimentellen Bedingungen. Freiwillig stärkere Nutzung war mit ungünstigeren psychosozialen Ergebnissen assoziiert. **Die Nutzungsintensität wurde damit nicht randomisiert; die Studie beweist nicht, dass häufige Nutzung Einsamkeit verursacht.** Kein Langzeitbeleg über Jahre, kein Nachweis für stationäre Pflege. Als peer-reviewte Langzeitstudie oder Nachweis menschlicher Beziehungsersetzung nicht verwendbar. Die gegenüber der Erstfassung aktualisierte Ergebnisdarstellung zeigt, warum Versionierung Bestandteil der Recherche sein muss.

G5 Qualitative Beziehungserfahrungen

Can social robots reduce the loneliness of older adults? — Interaction and communication between social robots and older adults. *Geriatric Nursing*. 2025. DOI: [10.1016/j.gerinurse.2025.04.032](https://doi.org/10.1016/j.gerinurse.2025.04.032). PubMed.

Zugang: PubMed-Originalabstract und Interessenkonflikterklärung. Qualitative Interviews und Beobachtungen, 15 alleinlebende Personen ab 65 Jahren mit mindestens dreimonatiger Robotererfahrung; konstruktivistische Grounded Theory. Zeigt unterschiedliche Beziehungsgestaltung und die Rolle vermittelnder Angehöriger oder Betreuungspersonen. Keine randomisierte Vergleichsgruppe, kein belastbarer kausaler Nachweis einer Einsamkeitsreduktion. Für partizipative Gestaltung und Implementierungsfragen durchaus relevant; nicht für die Behauptung, Roboter ersetzen soziale Kontakte erfolgreich. Wegen des begrenzten Ergänzungsumfangs hier nicht zusätzlich aufgenommen, nicht wegen vermeintlicher genereller Minderwertigkeit qualitativer Forschung.

B.5 Reichweite der Ergänzung

1. **Eine sofortige, klar begründete Ergänzung:** G1 erweitert das empirische Korpus um patientenbezogene reale Endpunkte. Eine zweite Quelle sollte nicht allein zur Erhöhung der Studienzahl aufgenommen werden.
1. **Robotik bleibt eine offengelegte Lücke:** G2 ist ein geeigneter Kandidat für die nächste Volltextprüfung. Bis dahin keine quantitative Robotik-Wirksamkeitssynthese behaupten.
1. **Langfristige Arbeit und Beziehungen bleiben offen:** Die gezielte Suche hat in den fünf näher geprüften Kandidaten keine über Jahre laufende Untersuchung zu Beschäftigung, Dequalifizierung oder Verdrängung menschlicher Beziehungen erschlossen. Das ist eine Grenze dieser Recherche, **keine Behauptung, dass weltweit keine entsprechenden Studien existieren.**
1. **Keine negative Gewissheit aus Unsicherheit:** Fehlender Nachweis eines Schadens ist keine Sicherheitsgarantie; fehlender signifikanter Nutzen kein Beweis von Wirkungslosigkeit. Ein kurzfristiger Produktivitätseffekt beantwortet keine langfristige Beschäftigungsfrage.
1. **Keine erfundene Systematik:** Die transparenten vier Suchabfragen verbessern Nachprüfbarkeit, rechtfertigen aber weder einen PRISMA-Vollständigkeitsanspruch noch eine formale GRADE-Einstufung. Für eine echte systematische Review wären vorab abgegrenzte Fragestellungen, Datenbanken, vollständige Suchstrategien, dokumentierte Auswahlverfahren und designgerechte Qualitätsbewertung erforderlich.

C Rechercheprotokoll der Rechtsvertiefung

Es wurde eine gezielte, nicht erschöpfende Primärquellenrecherche am 24. September 2026 durchgeführt. Ausgangspunkt waren die vorhandene Leitfaden und die Quellen [20]–[27], [36]–[37]. Die folgenden Zeichenfolgen wurden in diesem Arbeitsdurchlauf tatsächlich an die Websuche übergeben. Die Suchmaschine kann Anfragen intern normalisieren; insbesondere bei mehreren Anfragen des dritten Blocks meldete sie ähnliche statt exakt identischer Suchanfragen. Keine Trefferzahlen, systematische Vollständigkeit oder unabhängige juristische Begutachtung werden behauptet.

C.1 Ausgeführte Suchabfragen

1. site:eur-lex.europa.eu "2026/1744" "Date of effect" "27/07/2026"
1. site:recht.bund.de "KI-Marktüberwachungs" "2026" "22. Juli"
1. site:bmds.bund.de "Neues KI-Gesetz tritt in Kraft" "29.07.2026"
1. site:eur-lex.europa.eu "2026/1744" "Article 113" "2 December 2027"
1. site:gesetze-im-internet.de/ki-mig "Ausfertigungsdatum" "BGBl"
1. site:recht.bund.de "KI-MIG"
1. site:edpb.europa.eu guidelines 07/2020 concepts controller processor final version 2.0 purposes means own purposes processor
1. site:gesetze-im-internet.de/bdsg_2018/___26.html Beschäftigten Daten Einwilligung Abhängigkeit

1. site:gesetze-im-internet.de/bgb/_630d.html Einwilligung Aufklärung 630e
1. site:eur-lex.europa.eu "2017/745" "Article 5" "within health institutions" "quality management system"
1. site:eur-lex.europa.eu "2026/1744" "Article 4" "support the development"
1. site:eur-lex.europa.eu "2026/1744" "2 December 2026" "Article 50(2)"
1. site:eur-lex.europa.eu "2024/1689" "Article 25" "substantial modification"
1. site:eur-lex.europa.eu "2017/745" "Article 5" "not transferred to another legal entity"
1. site:eur-lex.europa.eu "32026R1744" "4. Providers of AI systems" "2 December 2026"
1. site:eur-lex.europa.eu "2026/1744" "placed on the market before 2 August 2026"
1. site:eur-lex.europa.eu "2016/679" "Article 28" "audits" "Article 44"
1. site:curia.europa.eu "C-34/21" "30 March 2023" employment Article 88

C.2 Zugriff und Grenzen

Tabelle 14 Zugang zu rechtlichen Originalquellen

Quelle/Prüfziel	Tatsächlich erreichter Zugriff
KI-MIG, Gesetze-im-Internet	Amtlich indexierter Text mit Ausfertigungsdatum, Vollzitat, BGBl.-Nummer und Inkrafttretensfußnote; direkter Volltextabruf in diesem Durchlauf Timeout. Der Treffer vermerkt „Nachgewiesener Text dokumentarisch noch nicht bearbeitet“.
Bundesgesetzblatt 2026 I Nr. 223	Direkter Versuch https://www.recht.bund.de/bgb/1/2026/223/VO.html : HTTP 403; nicht als gelesener Originalvolltext gewertet. Die beiden recht.bund-Suchen lieferten keinen zusätzlich auswertbaren Originalnachweis.
AI-Omnibus	Amtlich indexierte EUR-Lex-Metadaten und operative Auszüge aus Art. 1 Nr. 5, 39 und 40; direkte ELI-Seite nicht lesbar, EN/ALL-Seite nur dreizeilige technische Antwort. Keine Behauptung einer vollständigen authentifizierten PDF-Lektüre in diesem Durchlauf.
EDSA-Leitlinien 07/2020	PDF als 51-seitiger Volltext abrufbar; Rollen- und Vertragsabschnitte gezielt ausgewertet. Die Suchanfrage nannte Version 2.0; der abgerufene Text weist Version 2.1 aus.
§ 26 BDSG und § 630e BGB	Amtliche Einzelnormen direkt lesbar; relevante Absätze geprüft.
§ 630d BGB	Amtlicher indexierter Gesetzesauszug; direkter Abruf mit internem Fehler.
MDCG 2023-1	PDF als 20-seitiger Volltext abrufbar; gezielte Suche nach „legal entity“ und „does not apply“, einschlägige In-house-Abschnitte ausgewertet.
BMDS-Mitteilung	Amtlicher indexierter Text zur Gegenprüfung des Inkrafttretens.
EuGH C-34/21	Amtliche Fundstelle/Metadaten und Pressemitteilung gefunden; Entscheidung nicht vollständig geprüft und deshalb nicht als zusätzliche tragende Rechtsquelle dieser Ergänzung verwendet.

Die Aussagen zu bereits im Leitfaden belegten Normen bauen ergänzend auf dessen Quellenprüfung auf; sie werden hier nicht als sämtlich neu vollständig gelesene Gesetzestexte ausgegeben. Nicht vertieft geprüft wurden insbesondere Landeskrankenhaus-, Personalvertretungs- und kirchliches Datenschutzrecht, einzelne Berufsordnungen, besondere Forschungsregime, Vergaberecht sowie sämtliche sektorspezifischen Meldepflichten. Vor einer konkreten Freigabe sind diese nach Trägerschaft und Einsatz zu ergänzen. Die drei Fälle sind konstruierte Prüfzenarien, keine dokumentierten Schadensfälle oder rechtsverbindlichen Einzelfallentscheidungen.

C.3 Abschließende Gegenprüfung

Bei der abschließenden Quellenkontrolle am 24.09.2026 lieferten die System-1-Abfragen site:eur-lex.europa.eu "2026/1744" "2 December 2027" und site:gesetze-im-internet.de/ki-mig/"22. Juli 2026" "29. Juli 2026" keine geeigneten Treffer. Die anschließenden System-2-Abfragen Verordnung EU 2026 1744 AI Omnibus 2 December 2027 und KI-MIG 22 Juli 2026 223 Gesetz ergaben passende amtlich indexierte Texte und Bestätigungen. Direkte

Normabrufe blieben technisch eingeschränkt. Die Abfolge dokumentiert, weshalb ein leerer oder unpassender Suchlauf nicht mit dem Nichtbestehen eines Rechtsakts gleichgesetzt wurde.

D Transparenz und Publikationsstatus

Berichtstyp. Strukturierte kritische Literaturstudie mit normativer und rechtsbezogener Governanceanalyse. Keine eigene empirische Untersuchung und kein vollständiger systematischer Review.

Erstellung. KI-gestützte Recherche und Ausarbeitung für das vorliegende FluxLane-Projekt. Es wurden keine menschlichen Autorinnen oder Gutachter benannt, die an dieser Ausarbeitung nicht beteiligt waren. Eine wissenschaftlich verantwortliche menschliche Herausbergerschaft ist vor einer formalen Veröffentlichung festzulegen.

Begutachtung. Recherche-, Konsistenz- und Satzprüfungen innerhalb des Erstellungsprozesses; keine externe fachwissenschaftliche, klinische oder juristische Peer Review. Eine institutionelle Anwendung verlangt eine kontextbezogene Prüfung. Der Bericht selbst ist kein Konformitäts-, Sicherheits- oder Ethikzertifikat.

Finanzierung und Interessenkonflikte. Dem Erstellungsprozess lagen keine gesonderten Angaben zu einer Studienfinanzierung oder zu Interessenbindungen einer künftigen menschlichen Autorenschaft vor. Eine verifizierte Erklärung „keine Interessenkonflikte“ wird daher nicht abgegeben. Finanzierungs- und Interessenkonflikthinweise zu Originalstudien sind davon zu unterscheiden und stehen in Kapitel 4.

Patienten und Öffentlichkeit. Keine Patientendaten erhoben und keine Betroffenen als Studiengruppe einbezogen. Eine Beteiligung an lokalen Einführungen und an einer künftigen Evaluation des Verfahrens wird empfohlen, hat hier jedoch nicht stattgefunden. Ein Votum einer Forschungsethikkommission wurde nicht eingeholt und wird nicht behauptet.

Daten und Materialien. Alle berichteten empirischen Zahlen stammen aus bezeichneten Originalpublikationen. Es liegen keine eigenen Rohdaten vor. Anhänge B und C dokumentieren die ergänzenden Recherchehandlungen und ihre Grenzen; sie bilden kein vollständiges Sucharchiv der ersten Quellenzusammenstellung. Auswertungsbeispiele sind ausdrücklich hypothetisch gekennzeichnet.

Version und Aktualisierung. Version 1.1, Recherche- und Rechtsstand 24. September 2026. Neue Studien, Produktversionen, Sicherheitsinformationen und Rechtsänderungen können die Bewertung verändern. Vor einer verbindlichen Nutzung sind die für den konkreten Einsatz tragenden Originalnachweise erneut zu prüfen und zu archivieren.

Prüfbedarf vor Veröffentlichung. Fachwissenschaftliche Gegenprüfung besonders der nur teilweise zugänglichen Arbeiten [3,7]; Klärung der Analysepopulation bei [41]; juristische Gegenprüfung der zeitkritischen Rechtsangaben und Archivierung der authentischen Rechtsakte; Beteiligung von Pflegeleitung und Betroffenen an der Umsetzbarkeitsprüfung. Diese offenen Punkte werden nicht durch das professionelle Layout erledigt.

Literatur und Primärquellen

Die Referenznummern bleiben zur Rückverfolgung der Ausgangsrecherche stabil; die Auflistung ist numerisch geordnet. Zuletzt geprüft am 24. September 2026. Die vier nicht in den empirischen Kern aufgenommenen Recherchekandidaten werden mit Fundstellen in Anhang B dokumentiert.

[1] World Health Organization. *Ethics and governance of artificial intelligence for health*. 2021. [Originalpublikation](#); [offizielle Erläuterung der sechs Prinzipien](#).

[2] World Health Organization. *Ethics and governance of artificial intelligence for health: Guidance on large multi-modal models*. 2024. [Publikation](#); [Einführung vom 18.01.2024](#). Das abweichende Datum der aktuellen Landingpage ist kein Nachweis einer neuen Ausgabe.

[3] Gommers J et al. *Interval cancer, sensitivity, and specificity comparing AI-supported mammography screening with standard double reading without AI in the MASAI study*. The Lancet. 2026;407:505–514. DOI: [10.1016/S0140-6736\(25\)02464-X](#). [Zugänglicher Originalabstract](#), Universität Lund. Für diesen Beitrag

insbesondere der Originalabstract ausgewertet.

[4] Goh E et al. *Large Language Model Influence on Diagnostic Reasoning: A Randomized Clinical Trial*. JAMA Network Open. 2024;7:e2440969. DOI: [10.1001/jamanetworkopen.2024.40969](https://doi.org/10.1001/jamanetworkopen.2024.40969). Originalvolltext.

[5] Wong A et al. *External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients*. JAMA Internal Medicine. 2021;181:1065–1070. DOI: [10.1001/jamainternmed.2021.2626](https://doi.org/10.1001/jamainternmed.2021.2626). Originalvolltext. Historische Produktgeneration.

[6] Wong A, Currey D, Schwinne M et al. *Multicenter Prospective Validation of an Updated Proprietary Sepsis Prediction Model*. JAMA Network Open. 2026;9(2):e260181. DOI: [10.1001/jamanetworkopen.2026.0181](https://doi.org/10.1001/jamanetworkopen.2026.0181). Originalvolltext.

[7] Lukac PJ et al. *Ambient AI Scribes in Clinical Practice: A Randomized Trial*. NEJM AI. 2025;2(12). DOI: [10.1056/AIoa2501000](https://doi.org/10.1056/AIoa2501000). PubMed; PMC. Originalabstract und öffentlich indexierte Originalpassagen ausgewertet, direkter Volltextabruf teilweise eingeschränkt.

[8] Dratsch T et al. *Automation Bias in Mammography: The Impact of Artificial Intelligence BI-RADS Suggestions on Reader Performance*. Radiology. 2023;307:e222176. DOI: [10.1148/radiol.222176](https://doi.org/10.1148/radiol.222176). Originalvolltext im Hochschulrepositorium.

[9] Obermeyer Z, Powers B, Vogeli C, Mullainathan S. *Dissecting racial bias in an algorithm used to manage the health of populations*. Science. 2019;366:447–453. DOI: [10.1126/science.aax2342](https://doi.org/10.1126/science.aax2342). Originalartikel als PDF bei der FTC.

[10] Brynjolfsson E, Li D, Raymond LR. *Generative AI at Work*. Quarterly Journal of Economics. 2025;140:889–942. DOI: [10.1093/qje/qjae044](https://doi.org/10.1093/qje/qjae044). Finale Originalfassung. Nicht mit den abweichenden Zahlen früher Arbeitspapiere vermischen.

[11] Dell’Acqua F et al. *Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of Artificial Intelligence on Knowledge Worker Productivity and Quality*. Organization Science. 2026;37:403–423. DOI: [10.1287/orsc.2025.21838](https://doi.org/10.1287/orsc.2025.21838). Finale Veröffentlichung vom 11.03.2026, nicht das frühere Arbeitspapier.

[12] International Council of Nurses. *The ICN Code of Ethics for Nurses*. 2021, insbesondere Abschnitt 1.11. Vollständiger englischer Kodex.

[13] Zentrale Ethikkommission bei der Bundesärztekammer. *Entscheidungsunterstützung ärztlicher Tätigkeit durch Künstliche Intelligenz*. 2021. DOI: [10.3238/arztebl.zeko_sn_cdss_2021](https://doi.org/10.3238/arztebl.zeko_sn_cdss_2021). Stellungnahme und Original-PDF. Ethische Orientierung; nicht als aktueller AI-Act-Kommentar verwenden.

[14] Vorstand der Akademie für Ethik in der Medizin. *Standards für Ethikberatung im Gesundheitswesen*. Verabschiedet am 21.12.2022, publiziert 2023. Original-PDF.

[15] Vorstand der Akademie für Ethik in der Medizin. *Empfehlungen zur Erstellung von Ethik-Leitlinien in Einrichtungen des Gesundheitswesens*. 2024. DOI: [10.1007/s00481-024-00820-x](https://doi.org/10.1007/s00481-024-00820-x). Original-PDF.

[16] National Institute of Standards and Technology. *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. 2023. DOI: [10.6028/NIST.AI.100-1](https://doi.org/10.6028/NIST.AI.100-1). Original-PDF. Freiwilliges Rahmenwerk; laufende Überarbeitung nicht mit veröffentlichter Neufassung verwechseln.

[17] NIST. *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile*. 2024. DOI: [10.6028/NIST.AI.600-1](https://doi.org/10.6028/NIST.AI.600-1). Original-PDF.

[18] Datenschutzkonferenz. *Orientierungshilfe Künstliche Intelligenz und Datenschutz*. 06.05.2024. Original-PDF. Behördliche Orientierung, nicht Gesetz.

[19] Datenschutzkonferenz. *Orientierungshilfe zu datenschutzrechtlichen Aspekten von Retrieval-Augmented Generation*. Oktober 2025, Version 1.0. Original-PDF.

[20] Verordnung (EU) 2024/1689, AI Act, in der am Recherchetag maßgeblichen Fassung. Ursprünglicher Rechtsakt. Insbesondere Art. 3–6, 25–27, 50, 111 und 113 sowie Anhänge I und III; zusammen mit [21] lesen.

- [21] Verordnung (EU) 2026/1744 vom 08.07.2026, Amtsblatt vom 24.07.2026, Digital Omnibus on AI. **Amtlicher Rechtsakt**; **amtlicher englischer Text**; **Kommission: Inkrafttreten am 27.07.2026**. Die Änderung des Art. 113 verschiebt bestimmte Hochrisikovorschriften, nicht den gesamten AI Act.
- [22] Verordnung (EU) 2016/679, Datenschutz-Grundverordnung. **Amtlicher Rechtsakt**. Insbesondere Art. 5, 6, 9, 13–15, 22, 28, 32 und 35; weitere Vorschriften können je nach Datenfluss relevant sein.
- [23] Verordnung (EU) 2017/745 über Medizinprodukte, MDR. **Amtlicher Rechtsakt**. Insbesondere Art. 2, Erwägungsgrund 19 und Anhang VIII Regel 11; jeweils geltende Fassung prüfen.
- [24] Betriebsverfassungsgesetz. **§ 87, § 90, § 118**. Bei öffentlichen und kirchlichen Trägern einschlägiges Vertretungsrecht gesondert klären.
- [25] Allgemeines Gleichbehandlungsgesetz. **Amtlicher Gesetzestext**.
- [26] Gesetz zur Marktüberwachung und Innovationsförderung von künstlicher Intelligenz (KI-Marktüberwachungs-und-Innovationsförderungs-Gesetz – KI-MIG). Vom 22.07.2026, BGBl. 2026 I Nr. 223; in Kraft seit 29.07.2026. Zugangsbeschränkung des Gesetzblatts siehe Anhang C. **Amtlicher Gesetzestext**; **§ 2 Zuständigkeiten, § 6 Anlaufstelle, § 8 Beschwerdestelle, § 12 Information und Beratung**.
- [27] Bürgerliches Gesetzbuch. **§ 630a Abs. 2**, allgemein anerkannter fachlicher Standard bei medizinischer Behandlung.
- [28] Vasey B et al. *Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI*. Nature Medicine. 2022;28:924–933. DOI: [10.1038/s41591-022-01772-9](https://doi.org/10.1038/s41591-022-01772-9). Berichtsempfehlung; hier öffentlich zugängliche Verlagsinformationen ausgewertet.
- [29] Gallifant J et al. *The TRIPOD-LLM reporting guideline for studies using large language models*. Nature Medicine. 2025;31:60–69. DOI: [10.1038/s41591-024-03425-5](https://doi.org/10.1038/s41591-024-03425-5). Publikationsjahr 2025 trotz „2024“ im DOI.
- [30] Alderman JE et al. *Tackling algorithmic bias and promoting transparency in health datasets: the STANDING Together consensus recommendations*. The Lancet Digital Health. 2025;7(1):e64–e88. DOI: [10.1016/S2589-7500\(24\)00224-3](https://doi.org/10.1016/S2589-7500(24)00224-3). **Originalabstract**. Konsensarbeit, keine Wirksamkeitsstudie; online erstmals am 18.12.2024 veröffentlicht.
- [31] ISO/IEC 42001:2023. *Artificial intelligence – Management system*. **Offizielle Standardbeschreibung**; **offizielle Erläuterung**. Hier wurden die frei zugänglichen Beschreibungen, nicht der kostenpflichtige Normvolltext ausgewertet.
- [32] *International AI Safety Report 2026*. Veröffentlichung 03.02.2026. **Vollständiger Bericht**. Internationale wissenschaftliche Evidenzsynthese zu allgemeiner KI; keine verbindliche sektorale Richtlinie.
- [33] International Energy Agency. *Key Questions on Energy and AI*. 2026. **Executive Summary**. Beobachtungen und Prognosen auseinanderhalten.
- [34] Bundesamt für Sicherheit in der Informationstechnik. **Themenbereich Künstliche Intelligenz**; **Warnung zu indirekten Prompt-Injection-Angriffen**. Ergänzend fachbehördlicher Einstieg; einzelne Direktabrufe eingeschränkt.
- [35] Bundesinstitut für Arzneimittel und Medizinprodukte. **Rechtlicher Status und Klassifizierung**; **Risiken melden**. Fachbehördliche Einstiegs- und Meldeinformationen, keine allgemeine Freigabe sämtlicher Klinik-KI.
- [36] Medical Device Coordination Group. *MDCG 2019-11 rev.1: Qualification and classification of software*, Juni 2025. **Offizielle Veröffentlichung**. Ergänzend MDCG 2025-6, Zusammenspiel von MDR/IVDR und AI Act: **offizielle Veröffentlichung**. Unverbindliche fachbehördliche Orientierung; ältere Fristenangaben mit [21] abgleichen.
- [37] Strafgesetzbuch. **§ 201: Vertraulichkeit des Wortes**; **§ 203: Verletzung von Privatgeheimnissen**. Aufnahme, Verarbeitung und Einbindung externer Dienstleister anhand der konkreten Architektur und Berechtigung prüfen.
- [38] Baethge C, Goldbeck-Wood S, Mertens S. *SANRA—a scale for the quality assessment of narrative review articles*. Research Integrity and Peer Review. 2019;4:5. DOI [10.1186/s41073-019-0064-8](https://doi.org/10.1186/s41073-019-0064-8). Methodische Orientierung, kein durchgeführtes Bewertungssiegel.

- [39] Page MJ et al. *The PRISMA 2020 statement: an updated guideline for reporting systematic reviews*. BMJ. 2021;372:n71. DOI 10.1136/bmj.n71. Zur Abgrenzung des Berichtsanspruchs; diese Studie ist kein PRISMA-konformer systematischer Review.
- [40] Campbell M et al. *Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline*. BMJ. 2020;368:l6890. DOI 10.1136/bmj.l6890. Originalvolltext bei UCL.
- [41] Agweyu A, Mwaniki P, Menon V et al. *Generative AI-enabled clinical decision support system in primary care: a pragmatic, cluster-randomized trial*. Nature Medicine. 2026;32:3032–3039. DOI 10.1038/s41591-026-04503-6. Originalvolltext ausgewertet.
- [42] Europäischer Datenschutzausschuss. *Guidelines 07/2020 on the concepts of controller and processor in the GDPR*. Version 2.1; angenommen 07.07.2021, kleinere Korrekturen 20.09.2022. Originaldokument. Behördliche Auslegungshilfe.
- [43] Bundesdatenschutzgesetz. § 26, insbesondere Absatz 2 zur Freiwilligkeit im Beschäftigungsverhältnis. Einschlägigkeit und unionsrechtliche Anforderungen im konkreten Fall prüfen.
- [44] Bürgerliches Gesetzbuch. § 630d und § 630e, Einwilligung und Aufklärungspflichten.
- [45] Medical Device Coordination Group. *MDCG 2023-1 Guidance on the health institution exemption under Article 5(5)*. Januar 2023. Originaldokument. Unverbindliche Auslegungshilfe zu MDR und IVDR.
- [46] Bundesministerium für Digitales und Staatsmodernisierung. *Neues KI-Gesetz tritt in Kraft*. 29.07.2026. Amtliche Mitteilung. Zeitliche Gegenprüfung, kein Ersatz für den Normtext.

13 Rechtliche Hinweise

13.1 Keine Beratung

Diese Veröffentlichung ist ein redaktionelles Erzeugnis. Sie stellt keine Rechtsberatung, keine medizinische, pflegerische, arbeitsrechtliche oder datenschutzrechtliche Beratung und keine Beschaffungs- oder Vergabeberatung dar. Sie ersetzt insbesondere nicht die Prüfung durch Rechtsanwaltschaft, Datenschutzaufsicht, Medizinprodukteberatung, Vergabestelle, Ethikkommission nach Medizinprodukterecht oder Arbeitnehmervertretung. Wer auf Grundlage dieses Dokuments entscheidet, entscheidet auf eigene Verantwortung und hat den Sachverhalt vorher eigenständig fachlich prüfen zu lassen.

13.2 Keine Zusicherung, keine Konformitätsaussage

Die beschriebenen Verfahren, Prüfstufen und Checklisten sind redaktionelle Vorschläge. Sie sind keine Zusicherung einer Eigenschaft, kein garantiertes Ergebnis und kein Nachweis der Einhaltung einer Rechtsvorschrift. Wer ein Verfahren aus diesem Dokument anwendet, erfüllt damit keine gesetzliche Anforderung; ob eine Anforderung erfüllt ist, entscheidet die zuständige Stelle am Einzelfall.

13.3 Rechtsstand

Rechtsstände, Fristen und Fundstellen beziehen sich auf den im Dokumentkopf genannten Tag. Der KI-Rechtsrahmen wird laufend geändert; ein älterer Fristenplan darf nicht unverändert weiterverwendet werden. Genannte Gesetze und Verordnungen sind in ihrer jeweils geltenden Fassung heranzuziehen.

13.4 Haftung

Die Inhalte wurden mit Sorgfalt erstellt und, soweit im Methodenteil beschrieben, gegengeprüft. Für Richtigkeit, Vollständigkeit und Aktualität wird gleichwohl keine Gewähr übernommen. Eine Haftung des Herausgebers ist auf Vorsatz und grobe Fahrlässigkeit beschränkt; dies gilt nicht bei der Verletzung des Lebens, des Körpers oder der Gesundheit sowie bei zwingender gesetzlicher Haftung.

13.5 Verweise auf fremde Inhalte

Dieses Dokument verweist auf Inhalte Dritter, insbesondere auf Behörden-, Fachgesellschafts-, Normen- und Fachverlagsseiten. Auf deren Gestaltung und Inhalt hat der Herausgeber keinen Einfluss; für sie ist jeweils der dortige Anbieter verantwortlich. Die Adressen waren zum im Literaturverzeichnis genannten Prüfzeitpunkt erreichbar. Die Nennung eines Produkts, Herstellers oder Dienstes ist weder eine Empfehlung noch eine Bewertung seiner Eignung.

13.6 Marken und Produktbezeichnungen

Genannte Produkt-, Marken- und Unternehmensbezeichnungen sind Kennzeichen ihrer jeweiligen Inhaber und werden ausschließlich zur Beschreibung des Sachverhalts verwendet. Aus der Nennung folgt keine Verbindung zwischen dem Herausgeber und dem jeweiligen Unternehmen und keine Aussage über Zulassungs-, Konformitäts- oder Zertifizierungsstatus.

13.7 Urheberrecht und Nutzung

Dieses Dokument, sein Text, seine Tabellen und seine Abbildungen sind urheberrechtlich geschützt. Rechteinhaberin ist die WebTrades GmbH. Zulässig sind das Lesen, das Speichern und der Ausdruck für den eigenen Gebrauch sowie das Zitieren einzelner Stellen mit vollständiger Quellenangabe nach der im Impressum genannten Zitierweise. Die Vervielfältigung, Bearbeitung, Verbreitung, öffentliche Zugänglichmachung oder kommerzielle Verwertung des Dokuments oder wesentlicher Teile daraus bedarf der vorherigen schriftlichen Zustimmung der Rechteinhaberin. Das gilt ausdrücklich auch für die Übernahme von Tabellen und Abbildungen.

13.8 Keine personenbezogenen Daten

Für dieses Dokument wurden keine personenbezogenen Daten verarbeitet. Es beruht auf veröffentlichter Literatur, amtlichen Rechtsquellen und Veröffentlichungen von Fachgesellschaften; es wurden weder Beschäftigte befragt noch Patientendaten verwendet. Wer die hier beschriebenen Verfahren auf eine reale Einrichtung anwendet, verlässt diesen Rahmen: Die Auswertung von Tätigkeits-, Alarm- oder Quittierungsdaten unterliegt der Datenschutz-Grundverordnung und in der Regel der Mitbestimmung.

13.9 Unabhängigkeit

Dieses Dokument ist ohne Auftrag, ohne Finanzierung und ohne inhaltliche Beteiligung eines Herstellers, Krankenhausträgers, Kostenträgers, Verbands oder Beratungsunternehmens entstanden. Es besteht kein Vergütungs- oder Beteiligungsverhältnis zu einem der genannten Unternehmen oder Institutionen. Genannt werden Systeme, Anbieter und Dokumente, weil sie in den beschriebenen Gegenstand fallen, nicht weil sie beworben werden sollen.

13.10 Anwendbares Recht

Es gilt das Recht der Bundesrepublik Deutschland unter Ausschluss des UN-Kaufrechts. Gesetzliche Verbraucherschutzvorschriften bleiben unberührt.