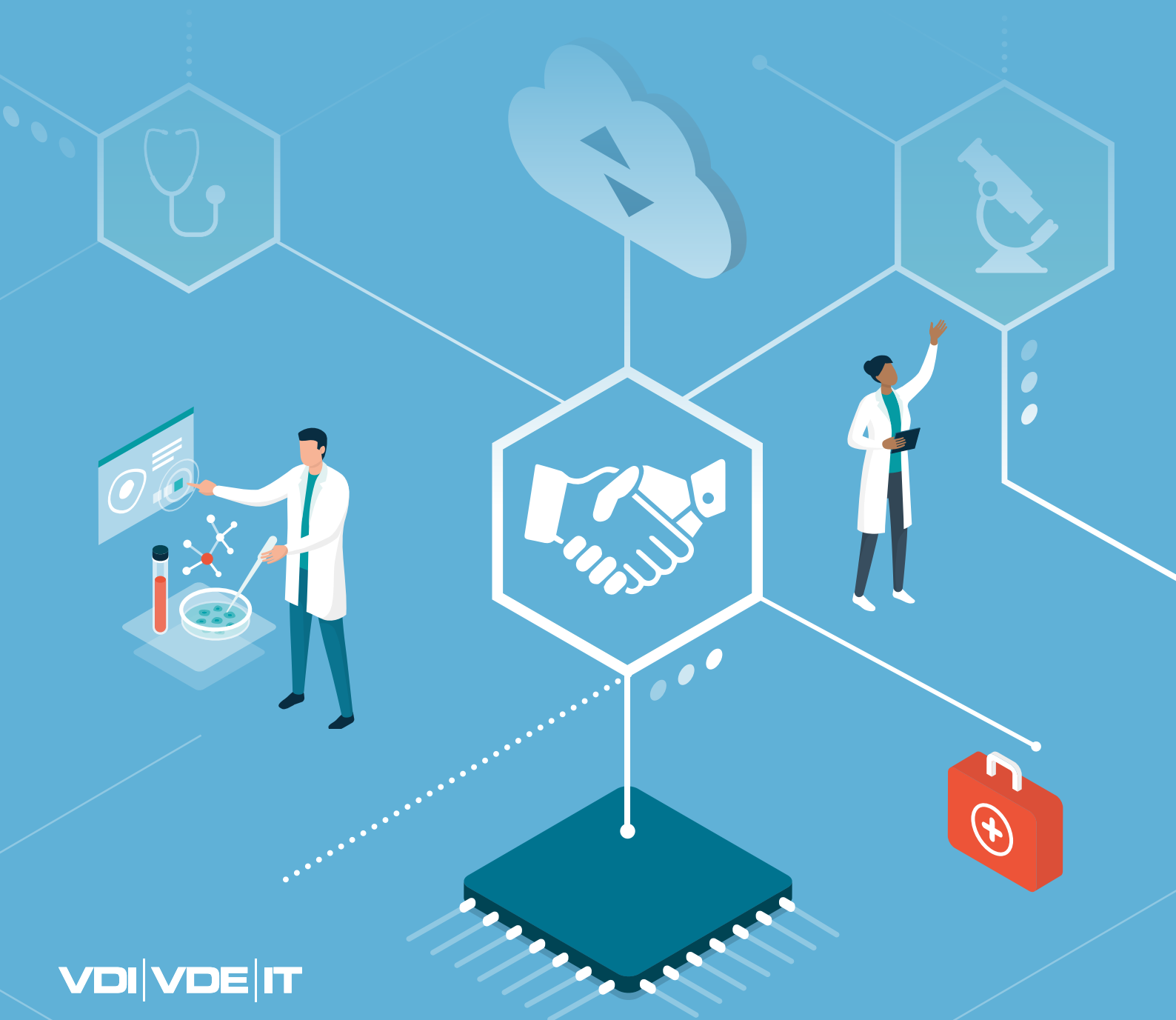


Mensch und KI als Team in der Medizintechnik: Vom KI-Hype zur Teamarbeit

Maxie Lutze, Bettina Schmietow, Benedikt P. Krieger



Mensch und KI als Team in der Medizintechnik: Vom KI-Hype zur Teamarbeit

KI-Systeme verändern die Medizintechnik: Von Bildanalytik über Entscheidungsunterstützung bis hin zu robotischen Assistenzsystemen entstehen neue Formen der Mensch-Maschine-Zusammenarbeit. Oft heißt es, KI werde ärztliche und pflegerische Expertise nicht ersetzen, sondern lediglich ergänzen. Als Team, so die Annahme, könnten deutlich bessere Ergebnisse erzielt werden. Aber wie sieht die Realität tatsächlich aus? Führt die Zusammenarbeit von Mensch und KI im Gesundheitssektor zwangsläufig zu einem besseren (Behandlungs-)Ergebnis?

Der EU AI Act schreibt in diesem Kontext vor, dass KI-basierte Medizintechnik stets unter der Aufsicht von Menschen stehen muss. Das heißt gleichzeitig aber auch: Das Gesundheitspersonal muss KI-Entscheidungen verstehen, hinterfragen – und im Zweifel korrigieren. Was bisherige Studien über die Performance von Mensch-KI-Teams zeigen, welche paradoxen Effekte auftreten können, und welche Faktoren eine vertrauenswürdige Teamkultur fördern können, soll im folgenden Beitrag genauer unter die Lupe genommen werden.

Mensch versus KI oder hybride Intelligenz? Ein Leistungsscheck

Existierende Studien vergleichen die Leistungsfähigkeit und Genauigkeit von KI-Modellen mit der von medizinischen Expert:innen, vor allem im Bereich der Diagnose. Zumeist wird dabei eine menschliche Leistung mit der Leistung einer KI gegenübergestellt, nicht aber die Leistung von unterschiedlichen Konstellationen von Mensch-KI-Teams untersucht ([Lutze et al. 2025](#)). Die Befunde zeigen ein gemischtes Bild:

Synergie-Potenzial: In einigen Feldern übertrifft die Kombination aus Mensch und KI beide Einzelleistungen. So konnten in einer Dermatologie-Studie Kliniker:innen mit Unterstützung von KI Hautkrebs genauer diagnostizieren als der menschliche Experte oder die KI allein. Besonders weniger erfahrene Mediziner:innen profitieren von der KI-Unterstützung ([Tschanl et al. 2020](#)). Eine Metaanalyse von 2024 bestätigt, dass KI-Unterstützung die Sensitivität und Spezifität der Hautkrebsdiagnose signifikant erhöhte ([Krakowski et al. 2024](#)), also wie zuverlässig echte Fälle erkannt und falsche ausgeschlossen wurden.

Kein Automatismus: Andere Untersuchungen zeigen, dass die Unterstützung durch KI nicht automatisch zu besseren Ergebnissen führt. Die KI-Assistenz verbesserte unter anderem die Diagnosequalität gegenüber einem modellierten Expertise-Standard bei Fällen, in denen die teilnehmenden Ärzt:innen unsicher waren. Bei eindeutig erscheinenden Fällen reduzierte sich die Ergebnisqualität sogar. Die Bereitstellung weiterer Informationen der Krankengeschichte verbesserte die Diagnosequalität im Durchschnitt, nicht jedoch die Nutzung der KI. Der potenzielle Nutzen hängt somit stark vom Kontext und von der Fähigkeit der KI ab, relevante Faktoren einzuordnen und systematische Verzerrungen (Bias) zu vermeiden ([Agarwal et al. 2023](#)).

Wo KI heute schon hilft: Im deutschen Mammographie-Screening-Programm erzeugte eine KI-Mensch-Kooperation eindrucksvolle Resultate. In der landesweiten PRAIM-Studie mit über 460.000 Frauen erhöhte KI die Brustkrebs-Entdeckungsrate um 18 Prozent gegenüber rein menschlicher Doppelbefundung – ohne mehr Fehlalarme. Gleichzeitig konnte die KI viele unauffällige Fälle aussortieren, sodass ohne Qualitätsverlust weniger befundet werden musste ([Eisemann et al. 2025](#)). Dieses Beispiel unterstreicht, dass richtig eingesetzte KI einen deutlichen Nutzen bringen kann.

Aufwind durch LLM: Neuere Studien fokussieren Teamleistungen mithilfe großer Sprachmodelle (LLM). In einem Experiment ([Zöller et al. 2025](#)) wurden unterschiedliche Teamkonstellationen verglichen – von Einzelpersonen über menschliche Teams, KI-Modelle und Kollektive von KI-Modellen hin zu gemischten Mensch-KI-Teams. Dabei konnten die besten Entscheidungen getroffen werden, wenn verschiedene Personen und mehrere KI-Modelle zusammenarbeiteten. Schon das Hinzufügen eines einzelnen KI-Modells zu einer Gruppe von Diagnostiker:innen verbesserte das Ergebnis erheblich. Der Erfolg resultiert aus komplementären Fehlerprofilen. Wenn die KI versagt, erkennen Menschen den Fehler und umgekehrt. So gleichen sich individuelle Schwächen aus.

Paradoxien und Herausforderungen der Zusammenarbeit von Mensch-KI-Teams

Trotz der Chancen gibt es in der Human-Factors-Forschung gut dokumentierte Herausforderungen in Mensch-Maschine-Interaktionen, die bei KI-Systemen und KI-basierter Medizintechnik neu oder verstärkt hervortreten:

Technisch-bedingte, soziale und kommunikative Effekte sowie rechtlich-normative Spannungen



Abbildung: Paradoxien und Herausforderungen der Zusammenarbeit von Mensch-KI-Teams (Quelle: Institut für Innovation und Technik (iit), eigene Darstellung)

Automationsparadox: Je verlässlicher KI-Systeme erscheinen, desto größer ist die Gefahr, dass Fachkräfte in eine passive Überwachungsrolle geraten und ihr Situationsbewusstsein verlieren. So übernahmen Patholog:innen unter Zeitdruck fehlerhafte KI-Empfehlungen, obwohl ihre ursprünglichen Einschätzungen korrekt gewesen wären (Rosbach et al. 2024). Übersichtsarbeiten belegen, dass vorausgefüllte oder standardisierte Systeme kritische Rückmeldeschleifen reduzieren (Levy et al. 2021). Dadurch kann eine trügerische Sicherheit entstehen (Out-of-the-loop-Problem) mit der Folge, dass klinische Intuition, Fehlererkennung und Urteilsfähigkeit verkümmern. Es kommt zum sogenannten „De-Skilling“ (Bratan et al. 2024).

Kontrollparadox in Mensch-KI-Teams: Ein zentrales Problem des Human-in-the-loop-Prinzips ist das Kontrollparadox in Mensch-KI-Teams. Gemeint ist hier, dass die Maschine liefert – und der Mensch prüft. Doch je leistungs-

fähiger die KI, desto schwieriger wird genau das. Studien zeigen, dass die menschliche Genauigkeit deutlich sinkt, wenn fehlerhafte KI-Vorschläge vorangehen (Agudo et al. 2024). Besonders kritisch wird es, wenn Systeme keine Hinweise auf Unsicherheiten oder Alternativen bieten.

Generalist-Spezialist-Paradox: KI-Systeme zeigen ihre Stärke vor allem in eng umrissenen Spezialaufgaben, etwa bei der Hautbild- oder EEG-Analyse oder der Klassifikation seltener Pathologien. In solchen Fällen erreichen sie Expert:innen-niveau oder übertreffen dieses. Ihre Leistung sinkt jedoch deutlich, sobald Aufgaben kontextabhängig, widersprüchlich oder interdisziplinär werden (Li et al. 2025). Im realen Klinikalltag bleibt der Output genereller Modelle oft zu unspezifisch. Besonders auf Intensivstationen zeigt sich, dass Ärzt:innen KI-Empfehlungen nicht blind folgen, sondern diese im Team „verhandeln“, so dass klinische Entscheidungen algorithmisch nicht direkt abbildbar sind (Sivaraman et al. 2023).

KI als kultureller Akteur im Gesundheitswesen: KI-Systeme beeinflussen zunehmend Sprache, Werte und Denkstile in klinischen Kontexten. Große Sprachmodelle wie ChatGPT imitieren nicht nur menschliche Sprache, sondern prägen sie aktiv durch bestimmte Ausdrucks- und Argumentationsmuster. Dieses Phänomen trägt zur Herausbildung einer „Machine Culture“ bei, in der technische Systeme kulturelle Routinen erzeugen und in soziale Interaktionen übertragen (Brinkmann et al. 2023; Yakura et al. 2025). Im medizinischen Umfeld zeigt sich das beispielsweise bei Medizinstudierenden, die KI-generierte Antworten zur Prüfungsvorbereitung nutzen. Ihre sprachlichen Formulierungen und klinischen Priorisierungen gleichen sich dabei stark den KI-Vorgaben an (Vicente et al. 2025). Überträgt sich dieser Effekt in den Klinikalltag, kann die unkritische Übernahme ökonomisierter Begriffe wie „low value patient“ oder standardisierter Textbausteine das Rollenverständnis und die Einschätzungen in Teams verändern. Pflegerische und psychosoziale Perspektiven drohen dabei zunehmend marginalisiert zu werden.

Zu viel oder zu wenig Vertrauen: Studien zeigen, dass Vertrauen in KI sowohl über- als auch unterdosiert sein kann, mit negativen Folgen für die Patientensicherheit (Patil et al. 2025; Kostick-Quenet et al. 2024). Es zeigt sich noch ein weiterer Effekt: die Überkorrektur durch den Menschen aus Misstrauen gegenüber der KI. Dieser sogenannte „Verschlimmbesserungseffekt“ tritt auf, wenn korrekte KI-Vorschläge vorschnell verworfen oder unnötig verändert werden – was paradoxerweise zu einem schlechteren Gesamtergebnis führen kann (Rieger et al. 2023). Beide Phänomene untergraben die funktionierende Zusammenarbeit und können zu einer Diffusion von Verantwortung führen.

Haftung & Verantwortung: Vertrauen in KI hängt nicht nur von technischer Performance ab, sondern auch von klaren rechtlichen Rahmenbedingungen. Aktuell haftet meist der Mensch, selbst wenn Entscheidungen durch algorithmische Empfehlungen vorgeprägt sind. Zentrale juristische Analysen sprechen von einem „responsibility gap“, weil unklar ist, wo menschliche Verantwortung endet und maschinelle beginnt ([Maroudas 2024](#)). Ärzt:innen haften tendenziell für Nutzung oder Missachtung von KI-Warnungen, Hersteller dagegen meist nur bei technischen Mängeln ([Price et al. 2024](#)).

Besonders kritisch wird das Problem bei „Black-Box“-Systemen: Je weniger nachvollziehbar die Entscheidung, desto schwerer ist der Beweis im Schadensfall ([Maroudas 2024](#)). Der AI Act fordert zwar menschliche Aufsicht, bleibt aber vage in der Haftungsverteilung ([Van Staalduinen 2024](#)). In der Folge nutzen viele Ärzt:innen KI nur, wenn sie Entscheidungen überstimmen oder vollständig dokumentieren können, was Effizienzgewinne konterkariert und Verantwortung unklar verteilt.

Moralische Divergenzen: KI-Systeme entscheiden schneller und neigen in einem Dilemma zu utilitaristischen Mustern: Effizienz und Outcome-Maximierung wiegen schwerer als Einzelfallgerechtigkeit oder Fürsorge. In Simulationsstudien priorisierten KI-Modelle rechnerisch nachvollziehbar das Überleben der Mehrheit, blendeten aber Würde, Chancengleichheit oder soziale Verantwortung aus ([Kirch et al. 2024](#)). In ressourcenarmen Kontexten zeigt sich, dass etwa ältere oder pflegebedürftige Menschen systematisch benachteiligt werden, da die Algorithmen auf Effizienzkennzahlen optimiert sind ([Hendricks-Sturup et al. 2023](#)). Verstärkt wird das Problem durch sogenanntes „Moral Outsourcing“, wobei Menschen die Verantwortung scheinbar objektiven Systemen überlassen ([Lang et al. 2023](#)).

Erfolgsfaktoren: Die Kooperation in Mensch-KI-Teams verbessern

Damit die Zusammenarbeit zwischen Gesundheitspersonal und KI-basierter Medizintechnik im Gesundheitswesen (hybriden Teams) klappt, braucht es mehr als funktionierende Algorithmen. Entscheidend ist, wie Fachpersonal und KI-Systeme gemeinsam handeln, wie Zuständigkeiten geregelt, Kommunikation ermöglicht und Fehler vermieden oder aufgefangen werden. Die folgenden Erfolgsfaktoren basieren auf Konzepten der Human-Factors-Forschung und sind speziell adaptiert für den medizinischen Kontext.

Technik, Workflow und Integration: Bevor KI-Systeme im klinischen Alltag genutzt werden, braucht es eine verlässliche technische Basis: validierte Algorithmen, transparente

Leistungsschwellen und eine sichere Einbindung in bestehende IT-Strukturen wie KIS (Krankenhausinformationssystem) oder PACS (Picture Archiving and Communication System). Entscheidend ist weiterhin, wie die KI-basierte Medizintechnik in den realen Arbeitsalltag eingebettet wird. Systeme, die sich nahtlos in bestehende Workflows integrieren, zum Beispiel direkt in Bildarchivierungs- oder Befundungssysteme, stoßen auf deutlich höhere Akzeptanz als isolierte Tools. Gleichzeitig kann es notwendig sein, bestehende Abläufe grundlegend zu hinterfragen, denn viele Workflows wurden ohne KI entwickelt.

Gute Mensch-KI-Teams zeichnen sich dadurch aus, dass die Interaktion intuitiv erfolgt: Die KI liefert Unterstützung dort, wo sie gebraucht wird, und hält sich zurück, wo sie nicht nötig ist. Interoperabilität und passgenaue Einbindung sind dabei eine Voraussetzung für Vertrauen und Kontrolle. Methoden wie die Cognitive Task Analysis helfen, Schnittstellen, Entscheidungspunkte und Übergaben zwischen Mensch und Maschine systematisch zu analysieren und zu gestalten.

Adaptive Rollen und Interaktion: Feste Aufgabenzuweisungen nach dem Prinzip „Men Are Best At, Machines Are Best At“ – zum Beispiel KI für die Vorverarbeitung von Bildern und großen Datenmengen, während ein interprofessionelles Team Empfehlungen oder Entscheidungen im klinischen und ethischen Kontext einordnet ([Moehring et al. 2025](#)), geben Orientierung. Gleichzeitig erfordert die Dynamik medizinischer Situationen gegebenenfalls Rollen Anpassungen. Nach diesem Ansatz wird die KI zu einem kognitiven Teampartner (Joint Cognitive System). Entscheidend ist, dass Zuständigkeiten je nach Falltyp angepasst, aber nie unklar sind. Es gilt die realen Aufgaben transparent zu machen, inklusive kognitiver Anforderungen, Abhängigkeiten und Ausnahmen sowie Entscheidungspunkten der Akteure. Task Analysis liefert hier die methodische Grundlage: Wer macht was – wann – warum – mit welchen Mitteln – und mit welchen Risiken?

Erklärbarkeit und Transparenz: Klinische Entscheidungen beruhen selten auf simplen Ja-Nein-Logiken, sondern Erfahrungswissen, Unsicherheiten und situativem Abwägen. Deshalb müssen die Entscheidungswege der Systeme nachvollziehbar sein (Explainable AI). Studien deuten darauf hin, dass Verständlichkeit das Vertrauen erhöht und sogar die diagnostische Genauigkeit verbessern kann ([Chanda et al. 2025](#)). Wichtig ist jedoch, dass die Erklärungen korrekt und nützlich sind – eine scheinbare Erklärung, die falsch ist, schafft trügerische Sicherheit. Erklärbarkeit sollte daher vor allem eingesetzt werden, um die Unsicherheiten, Zusatzinformationen und Alternativen aufzuzeigen und so die Aufmerksamkeit der Nutzenden zu lenken.

Kompetenzen und Vertrauen: Viele Fehler in der Mensch-KI-Interaktion entstehen durch missverstandene Rollen, falsche Erwartungen oder fehlenden Kontext statt durch Technikversagen. Deshalb bedarf es gezielter Schulungen. Das umfasst sowohl das technisch-funktionale Verständnis (zum Beispiel was bedeuten KI-Wahrscheinlichkeiten?) als auch das kognitive Verständnis (zum Beispiel Was kann delegiert werden?), um Vertrauen richtig zu dosieren. Simulationen und Mensch-KI-Übungen können helfen, typische Fehlerquellen zu erkennen und als Team auf KI-Antworten zu reagieren. Einige Kliniken in Deutschland bieten entsprechende Weiterbildungen bereits an, damit das Fachpersonal lernt, KI-Ergebnisse kritisch zu beurteilen (vergleiche [KI-Campus](#)).

Vom Prinzip zur Praxis: Hybride Intelligenz strukturiert verankern

Es gibt zahlreiche Entwicklungen, die auf das Potenzial erfolgreicher Mensch-KI-Teams in der Medizintechnik hinweisen. Dieser Beitrag hat zentrale Paradoxien aufgezeigt und Erfolgsfaktoren benannt – doch damit diese auch Wirkung entfalten, braucht es systematische Forschung, strukturierte Lernprozesse und eine gezielte Umsetzung.

Aktuell fehlt es an systematischen Übersichten darüber, in welchen medizinischen Anwendungsbereichen, etwa in Onkologie, Notaufnahme oder Pflege, welche Formen der Mensch-KI-Zusammenarbeit tatsächlich funktionieren ([Lutze et al. 2025](#)). Einzelstudien liefern wertvolle Einblicke, doch für Praxis, Regulierung und Weiterentwicklung braucht es vergleichende Übersichten, qualitative Tiefenanalysen und konkrete Leitlinien für verschiedene Settings im Gesundheitswesen. Eine öffentlich zugängliche Datenbank zur hybriden Intelligenz könnte Konzepte und Evidenzlagen sichtbar machen sowie Studiendesigns standardisieren ([Vaccaro et al. 2024](#)).

Darüber hinaus ist eine vertiefte Interaktionsforschung nötig, die Mensch-KI-Zusammenarbeit als kooperatives Handeln untersucht und Einflüsse auf die Kultur beleuchtet: Wie entstehen geteilte Rollen, wie wird Verantwortung wahrgenommen, und wie werden Entscheidungen koordiniert beziehungsweise mit Unsicherheiten umgegangen? Besonders die zunehmende Flexibilität generativer KI etwa in Bezug auf Anpassungsfähigkeit oder kreative Problemlösung macht es erforderlich, die KI als aktiven Teampartner zu begreifen, und nicht mehr nur als Werkzeug ([Schmutz et al. 2025](#)).

Implementierungsforschung kann ergänzend zeigen, in welchen Settings welche Formen hybrider Zusammenarbeit besonders effektiv sind etwa bei komplexen Fällen, vulnerablen Gruppen oder dynamischen Entscheidungsprozessen. Dabei sollten auch unterschiedliche Praxislogiken von Pflege, ärztlicher Versorgung und Management stärker berücksichtigt werden.

Partizipation und Teamkultur sind entscheidend für die erfolgreiche Einführung. KI-Systeme sollten nicht von oben eingeführt werden, sondern gemeinsam mit dem klinischen Personal in [Reallaboren, Simulationen oder Pilotprojekten](#) entstehen. In solchen Umgebungen lassen sich ethische Fragen, organisatorische Prozesse und technische Grenzen praxisnah reflektieren und anpassen. Die Perspektive des Teams muss von Anfang an Teil des Designs sein, nur so entsteht Vertrauen und klinische Anschlussfähigkeit ([Wysocki et al. 2022](#)).

Embedded Ethics spielt dabei eine zentrale Rolle, gerade für kleine und mittlere Unternehmen (KMU), die viele KI-Innovationen vorantreiben ([Lutze et al. 2025](#)). Ansätze wie „Ethics by Design“ oder „[Integrierte Forschung](#)“ helfen, Prinzipien wie Fairness, Erklärbarkeit und menschliche Kontrolle direkt in die Produktentwicklung einzubetten ([Willem et al. 2025](#)). Das steigert nicht nur die Marktakzeptanz, sondern hilft auch regulatorische Anforderungen frühzeitig zu erfüllen. Institutionelle Unterstützung, zum Beispiel durch Gütesiegel, Standards oder Förderprogramme, kann hier gezielt Orientierung bieten.

Abschließend braucht es strukturierte Lernprozesse über Einzeleinrichtungen hinaus. In sogenannten „Communities of Practice“ können Gesundheitseinrichtungen, Entwickler:innen und Forschungseinrichtungen Erfahrungen austauschen, Good Practices etablieren und Fehlentwicklungen frühzeitig erkennen. Solche Netzwerke sind essenziell, um Wissen zu bündeln und kontinuierlich an den klinischen Alltag rückzukoppeln.

Herausgeber:

VDI/VDE Innovation + Technik GmbH
Steinplatz 1 | 10623 Berlin
www.vdivde-it.de

Bildnachweis:

elenabs/istockphoto

© VDI/VDE-IT 2025