Robert Langner, Michael B. Steinborn

II. Psychologische Aspekte der Nutzung Künstlicher Intelligenz (KI) in der radiologischen Diagnostik

In der klinischen Radiologie werden heute zahlreiche KI-basierte Anwendungen entwickelt, erprobt oder bereits genutzt, um Arbeitsschritte und -abläufe im diagnostischen Prozess zu automatisieren. Solche radiologischen KI-Anwendungen zielen beispielsweise darauf ab, (i) Bildgebungsdaten zu strukturieren und relevante Bildbereiche visuell zu akzentuieren, (ii) Bildmuster diagnostisch einzuordnen und daraus (vorläufige) Befunde abzuleiten, (iii) Untersuchungen nach Dringlichkeit zu ordnen und damit den klinischen Arbeitsfluss zu steuern oder (iv) visuelle und quantitative Befunde sprachlich darzustellen. Der vielfältige Einsatz von KI verändert dabei den diagnostischen Arbeitsprozess nicht nur auf technischer Ebene, sondern greift in kognitive Prozesse, das ärztliche Selbstbild, die Verteilung von Verantwortung, berufliche Beziehungen sowie die Organisation der Arbeit ein. Leistungsfähigkeit und Sicherheit KI-gestützter radiologischer Diagnostik ergeben sich also nicht allein aus der Genauigkeit und Korrektheit der eingesetzten KI-Systeme, sondern auch aus der Art, wie KI-Anwendungen in die diagnostische Arbeit eingebunden sind. Maßgeblich sind dabei psychologische Mechanismen, die durch den Automatisierungsgrad, die Gestaltung der Automatisierung, die funktionale Platzierung des KI-Einsatzes im Entscheidungsablauf sowie durch organisationale und rechtliche Einsatzbedingungen beeinflusst werden (Parasuraman et al., 2000; Parasuraman & Riley, 1997; Sittig & Singh, 2010). Eine an menschlichen Eigenschaften orientierte Analyse dient dazu, diese Mechanismen systematisch zu bestimmen und aufzuzeigen, wie die Nutzung von KI diagnostisches Denken und Handeln in der Radiologie verändern kann.

Radiologische Diagnostik findet unter Bedingungen statt, die Fehlentscheidungen besonders folgenreich machen. Relevante Befunde sind in der Gesamtheit des zu beurteilenden Bildmaterials vergleichsweise selten und visuell oft schwer zu detektieren, während Zeitdruck und hoher Durchsatz die verfügbare Verarbeitungszeit begrenzen. Hinzu kommt eine asymmetrische Kostenstruktur von Fehlern: Das Übersehen klinisch relevanter Veränderungen wiegt meist schwerer als ein falsch-positiver Befund (»Fehlalarm«). Diese Konstellation beeinflusst Such- und Entscheidungsprozesse bereits implizit. Unter hoher Belastung konzentriert sich die Wahrnehmung auf erwartete Bildmerkmale, zugleich verschieben sich die Maßstäbe der Entscheidungsbildung, abhängig davon, ob zügiges Vorgehen oder Absicherung im Vordergrund steht. Frühe Festlegungen begünstigen vorschnelle Urteile, stärker angehobene Urteilsschwellen erhöhen dagegen die Wahrscheinlichkeit des Übersehens. Beide Fehlermodi sind funktional miteinander gekoppelt und entstehen, weil nur ein Teil der verfügbaren diagnostischen Informationen in die Entscheidung einfließt. Vor diesem Hintergrund liegt der Einsatz von KI nahe, doch ihr Beitrag zur diagnostischen Qualität hängt davon ab, wie zusätzliche Informationen in Wahrnehmung und Bewertung eingebunden sind. Förderlich wirken sie nur dann, wenn sie diagnostische Suchprozesse strukturieren, ohne diese zu überlagern oder zu stören (Drew et al., 2013). Entsprechend beeinflusst KI diagnostische Leistung auch über ihre Wirkungen auf Aufmerksamkeit, emotionale Sicherheit und Prüfverhalten, da diese Faktoren den Verlauf und die Qualität diagnostischer Entscheidungen mitbestimmen.

Die kognitiven Folgen des KI-Einsatzes lassen sich nur erfassen, wenn seine Funktionen entlang des diagnostischen Arbeitsgangs getrennt betrachtet werden. Eingriffe in klinische Priorisierung, visuelle Orientierung, diagnostische Klassifikation oder Befundformulierung greifen jeweils in unterschiedliche Arbeitsphasen ein und verändern entsprechend verschiedene kognitive Prozesse. Eine undifferenzierte Gesamtbewertung des KI-Einsatzes in der Radiologie griffe daher zu kurz. Da Priorisierung, visuelle Orientierung, Klassifikation und Befundformulierung jeweils andere kognitive Prozesse betreffen, lassen sich die Folgen auch nicht in einer einzelnen Leistungskennzahl abbilden, sondern nur als Kombinationen aus Veränderungen diagnostischer Genauigkeit, charakteristischen Fehlermustern sowie Verschiebungen von Tempo, Belastung und emo-

tionaler Sicherheit im Arbeitsvollzug. Darüber hinaus verändert der KI-Einsatz organisationale Abläufe. Er beeinflusst, wem eine Befundentscheidung zugerechnet wird, welche zusätzlichen Prüf- und Dokumentationsschritte erforderlich sind und wie technische Hinweise in die ärztliche Entscheidung eingehen. Diese Anpassungen betreffen nicht das System isoliert, sondern den gesamten diagnostischen Prozess. Die folgende Analyse verknüpft sie mit konkreten Eingriffspunkten und macht sichtbar, wo KI entlastet und wo neue Belastungen oder Unsicherheiten auftreten.

Die Bewertung des KI-Einsatzes in der Radiologie durch verschiedene Akteursgruppen folgt deren unterschiedlichen Zielsetzungen und Anforderungen. Aus radiologischer Perspektive ist z. B. maßgeblich, wie geeignet eine gegebene KI-Anwendung für bestimmte Aufgaben ist, wie sie in bestehende Abläufe integriert ist, welche zusätzlichen Prüf- und Absicherungsschritte entstehen und wie Verantwortung für Befunde im diagnostischen Prozess zugeordnet bleibt. Ob ein vorgeschlagener Befund als unterstützende Rückversicherung, als Unterbrechung des Arbeitsflusses oder als faktische Entscheidungsvorgabe wirkt, hängt von seiner funktionalen Platzierung und Gewichtung im Ablauf ab. Aus Patientenperspektive stehen demgegenüber Fragen der wahrgenommenen Sicherheit, der Fairness diagnostischer Entscheidungen und der Qualität der ärztlichen Interaktion im Vordergrund. Da Akzeptanz neuer medizinischer Technologien aus unterschiedlichen Bedingungen hervorgeht, ist eine analytische Trennung dieser Perspektiven erforderlich. Ärztliche Akzeptanz einer (neuen) Technologie wird typischerweise über wahrgenommene Nützlichkeit und Nutzerfreundlichkeit beschrieben, patientenseitige Akzeptanz über Vertrauen und Kommunikation. Die Kapitelstruktur folgt dieser Differenzierung und entwickelt die Analyse von kognitiven Automatisierungseffekten über affektive und verantwortungsbezogene Aspekte bis hin zu Fragen der Akzeptanz und organisatorischen Einbettung.

1. Kognitive Konsequenzen KI-basierter Automatisierung

1.1 Theoretischer Rahmen

Eine kognitive Betrachtung des KI-Einsatzes in der Radiologie beginnt damit, festzulegen, welche Aufgaben automatisiert werden und

wie menschliches Handeln daran anschließt. Automatisierung kann die ärztliche Aufmerksamkeit ausrichten, Entscheidungen vorbereiten oder die Reihenfolge und Weiterleitung von zu beurteilenden Fällen bestimmen. Da diese Eingriffe an unterschiedlichen Punkten der Informationsverarbeitung ansetzen, verändern sie die Art, wie Fehler und Abhängigkeiten entstehen (Parasuraman et al., 2000). Abhängig von dieser Ausgestaltung bleibt diagnostisches Entscheiden eine primär menschliche Tätigkeit oder nimmt die Form einer nachgeordneten Prüfung an. Eine solche Verschiebung in Richtung Nachprüfung stellt erhöhte Anforderungen an die kognitive Kontrolle, da geringere aktive Beteiligung mit einer erhöhten Anfälligkeit für Nachlässigkeit einhergehen kann (Parasuraman & Riley, 1997). Dasselbe KI-System kann dadurch die diagnostische Treffgenauigkeit erhöhen und zugleich die Art verändern, wie Fehler zustande kommen. Vor diesem Hintergrund lässt sich zwischen kurzfristigen Veränderungen einzelner Bearbeitungssituationen und längerfristigen Anpassungen diagnostischer Routinen unterscheiden. Die folgenden Abschnitte behandeln die zugrunde liegenden Mechanismen sowie Ansatzpunkte zur Begrenzung vorhersehbarer Fehlentwicklungen.

1.2 Kurzfristige kognitive Auswirkungen

Radiologische Bildbefundung beruht auf visueller Suche unter Bedingungen niedriger Ereignishäufigkeit und hoher Ähnlichkeit zwischen relevanten Signalen und Hintergrundstrukturen. Auffälligkeiten treten selten auf und heben sich oft nur schwach vom übrigen Bildmaterial ab. Unter diesen Bedingungen verändert sich mit zunehmender Suchdauer die statistische Gewichtung möglicher Befunde: Die Wahrscheinlichkeit, weitere Auffälligkeiten zu identifizieren, nimmt ab, und die gezielte Aufrechterhaltung von Aufmerksamkeit über die Zeit wird schwieriger, selbst bei hoher Expertise (Wolfe et al., 2005). KI greift in diesen Suchkontext ein, indem sie Hinweise bereitstellt oder bestimmte Bildregionen vorab markiert. Dadurch verschiebt sich, welche Bereiche bevorzugt geprüft werden, für wie lange die Suche fortgeführt wird und in welchem Maß Exploration offen bleibt. Solche Vorgaben können den Zugang zu relevanten Bildbereichen beschleunigen. Zugleich organisieren sie die Suche stärker entlang vorgegebener Relevanzen. Wird Aufmerksamkeit

wiederholt an markierte Bereiche gebunden, steigt die Wahrscheinlichkeit, dass relevante, aber nicht hervorgehobene Befunde außerhalb des Suchfokus bleiben. Hinzu kommt, dass die visuelle Suche häufig endet, sobald ein plausibler Befund identifiziert ist; alternative Deutungen werden dann seltener weiterverfolgt, auch wenn zusätzliche Hinweise vorhanden wären (Berbaum et al., 1990; Drew et al., 2013).

Der Einsatz diagnostischer KI-Hinweise verschiebt die kognitive Aufgabe der Befundung. Statt eigenständig nach Auffälligkeiten zu suchen, richtet sich die Tätigkeit zunehmend auf die Prüfung eines bereits vorgeformten Vorschlags. Diese Umstellung begünstigt Automatisierungsverzerrungen, bei denen maschinelle Empfehlungen bevorzugt akzeptiert und eigenständige Suchbewegungen verkürzt werden (Parasuraman & Riley, 1997). Fehler entstehen dabei auf zwei Wegen. Zum einen können falsche Hinweise direkt übernommen werden; zum anderen wird das Ausbleiben eines Hinweises selbst als Information interpretiert und kann als stillschweigende Entwarnung wirken. Unter diesen Bedingungen werden schwache oder atypische Auffälligkeiten besonders anfällig fürs Übersehen, da sie weder visuell hervorstechen noch algorithmisch markiert sind. Gleichzeitig können hoch saliente Markierungen die Interpretation in eine bestimmte Richtung lenken und harmlose Strukturen als pathologisch erscheinen lassen. Studien zeigen, dass Zeitdruck, autoritative Ergebnisdarstellung und hohe Erwartungen an die Systemverlässlichkeit diese Effekte verstärken (Goddard et al., 2012; Dratsch et al., 2023). Der Bilddatensatz wird damit weniger als offener Suchraum behandelt, sondern als Grundlage zur Verifikation einer vorgegebenen Annahme. Diese Verschiebung der Verifikationspraxis verändert Art und Wahrscheinlichkeit von Fehlern systematisch und stellt ein zentrales kurzfristiges Sicherheitsrisiko dar.

Der Einsatz von KI verändert die Organisation radiologischer Arbeit entlang des diagnostischen Arbeitsgangs. Automatisierte Funktionen übernehmen vor allem wiederkehrende Prüfschritte und Vorstrukturierungen, wodurch sich die verbleibenden Aufgaben stärker auf Situationen konzentrieren, in denen Unsicherheit, Interpretation und Verantwortung zusammenwirken. Die diagnostische Arbeit verschiebt sich damit in Richtung selektiver Bearbeitung solcher Konstellationen, die eine Einordnung KI-basierter Hinweise erfordern. Besonders relevant sind Fälle, in denen Systemvorschläge

und eigene Einschätzungen auseinanderfallen. In diesen Situationen wird diagnostisches Entscheiden mit der Aufgabe verbunden, die eigene Bewertung nachvollziehbar zu begründen und gegenüber KI-gestützten Hinweisen einzuordnen. Diese Begründungstätigkeit gehört dann zum regulären Arbeitsprozess und prägt die diagnostische Praxis zunehmend. Parallel dazu erfordert der Einsatz von KI eine fortlaufende Aktualisierung der Einschätzung der Leistungsfähigkeit des KI-Systems, da die Aussagekraft KI-basierter Hinweise vom jeweiligen Kontext abhängt. Daraus ergeben sich veränderte Belastungsprofile im Arbeitsablauf. Unterschiedlich anspruchsvolle Falltypen folgen in dichter Abfolge aufeinander und machen wiederholte Übergänge und schnelle Wechsel zwischen Kontroll- und Analyseformen erforderlich. Abgesehen von lange bekannten Effizienzeinbußen durch häufige Aufgabenwechsel (Kiesel et al., 2010), richtet sich die ärztliche Aufmerksamkeit dabei möglicherweise stärker auf die Plausibilität, Konsistenz und Anschlussfähigkeit diagnostischer Ergebnisse. Wie sich diese Verschiebungen auswirken, hängt von der konkreten Einbindung der KI in bestehende Routinen ab (Dzindolet et al., 2003; Parasuraman & Manzey, 2010).

1.3 Langfristige kognitive Auswirkungen

Langfristig verändert der Einsatz von KI in der Radiologie nicht nur kognitive Prozesse im diagnostischen Arbeitsablauf, sondern die Bedingungen, unter denen diagnostische Kompetenz entsteht und stabil bleibt; er verändert also die ärztliche Expertise. Radiologische Expertise beruht wesentlich auf der fortgesetzten aktiven Auseinandersetzung mit Bildmaterial: visuelle Erkennungsleistungen im radiologischen Screening gehören zu den Bereichen, in denen die ärztliche Expertise ihre besondere Stärke entfaltet. Es handelt sich dabei nicht um das bloße Wiedererkennen bekannter Muster, sondern um eine aktive, suchende Exploration des Bildes mit dem Ziel, potenziell relevante Abweichungen überhaupt erst zu identifizieren. Diese Form der Leistung ist offen angelegt und auf Sensitivität gegenüber Unerwartetem ausgerichtet. Darin unterscheidet sie sich grundlegend von nachgelagerten Bewertungsprozessen, bei denen bereits identifizierte Befunde eingeordnet, relativiert und diagnostisch gewichtet werden. Visuelle Differenzierung, Mustererkennung und der Umgang mit

Unsicherheit werden dabei weniger durch abstraktes, explizites Faktenwissen als durch kontinuierliche praktische Anwendung aufrechterhalten. KI-gestützte Automatisierung verschiebt jedoch die funktionale Gliederung der Teiltätigkeiten im diagnostischen Prozess. Aufgaben, die KI zuverlässig übernimmt, fallen damit als regelmäßige Übungsgelegenheiten weg, die für den Erhalt diagnostischer Fertigkeiten erforderlich sind. Über längere Zeiträume kann KI somit zu einem Qualifikationsverlust (»deskilling«: Verlust von Fertigkeiten) führen: Kompetenzen, die nicht fortlaufend aktiviert werden, verlieren an Präzision, Geschwindigkeit und Robustheit. Langfristig entsteht so ein systemisches Risiko, da zunehmende KI-Abhängigkeit bei gleichzeitig abnehmender menschlicher Expertise die diagnostische Robustheit und damit die radiologische Versorgungssicherheit untergraben kann (Bainbridge, 1983; Endsley & Kiris, 1995).

Radiologische KI-Nutzung kann langfristig aber auch neue Formen fachlicher Kompetenz (»upskilling«) hervorbringen – jedoch nur unter der Voraussetzung, dass der Umgang mit der KI selbst zum expliziten Gegenstand des Lernens wird. Kompetenz verlagert sich dann von der isolierten Erkennung einzelner Auffälligkeiten hin zur sachgerechten Steuerung eines KI-gestützten diagnostischen Systems. Diese Form der Expertise besteht in konkreten operativen Fertigkeiten: der kontextsensitiven Einschätzung der Zuverlässigkeit von KI-Ausgaben, dem Wissen um typische Fehlmuster, dem reflektierten Umgang mit Abweichungen zwischen eigener Einschätzung und Systemvorschlag sowie der systematischen Integration von KI-Ergebnissen in weitere klinische Informationen. Diese Kompetenzverschiebung bleibt jedoch aus, wenn sie stillschweigend vorausgesetzt wird. Ohne explizite organisationale und didaktische Rahmung entstehen weder stabile Lernprozesse noch ein belastbarer Erhalt dieser Kompetenzen. Parallel dazu formt langfristige KI-Nutzung diagnostische Strategien selbst: Über die Zeit verändern sich Suchroutinen, Prüftiefen und Abbruchkriterien; Befundung kann sich von offener Exploration zu einer Praxis entwickeln, die primär auf Abgleich und Bestätigung ausgerichtet ist. Solche Anpassungen können in hochstrukturierten Kontexten mit hohem Durchsatz funktional sein. Problematisch werden sie dort, wo dieselben Routinen auf Situationen übertragen werden, in denen KI versagt oder nicht verfügbar ist. Ob diese Verschiebungen in erhöhte Robustheit oder in neue Formen der Fehleranfälligkeit im diagnostischen Prozess mün-

den, entscheidet die Qualität metakognitiver Überwachung (Lee & See, 2004).

Über längere Zeiträume lässt sich der Einsatz von KI als Einflussfaktor auf unterschiedliche Entwicklungspfade fachlicher Praxis analysieren. Im Zentrum steht dabei nicht die Annahme einer einheitlichen oder notwendigen Wirkung, sondern die Frage, unter welchen Bedingungen diagnostische Kontrolle stabilisiert oder geschwächt wird. Analytisch relevant ist dies auf der Ebene unterschiedlicher Praxisformen (verstanden als wiederkehrende Handlungs-, Aufmerksamkeits- und Überwachungsformen im diagnostischen Prozess), also der Weise, wie Urteil, Aufmerksamkeit und Überwachung im Zusammenspiel von Mensch und System organisiert sind. Unter bestimmten Bedingungen entstehen Routinen, in denen KI-Ausgaben fortlaufend auf Plausibilität geprüft und als vorläufige Hinweise behandelt werden. Die ärztliche Bildanalyse bleibt dann aktiv, Abweichungen werden aufgegriffen, und das Vertrauen in das System bleibt an eine kontinuierliche Überprüfung gekoppelt. Unter anderen Bedingungen kann sich hingegen eine Praxis herausbilden, in der hohe Systemleistung, Zeitdruck und seltene Fehlerereignisse die Aktivierung ärztlicher Prüfprozesse zunehmend reduzieren. Die Exploration des Bildmaterials tritt dann in den Hintergrund, während die fachliche Aufmerksamkeit auf Ausnahmen fokussiert wird. Diese Divergenz lässt sich als Ergebnis veränderter Kalibrierungs- und Überwachungsdynamiken rekonstruieren: Mit sinkender Häufigkeit korrektiver Eingriffe werden Überwachungsroutinen seltener aktiviert, während die Vigilanz (d. h. die Aufrechterhaltung der Aufmerksamkeit) unter Bedingungen niedriger Ereignisraten instabil wird. Aus epistemischer Sicht folgt daraus, dass langfristige Sicherheit weniger von der Qualität des KI-Systems allein abhängt als von organisationalen Bedingungen, die eigenständige Praxis, Rückmeldung und Überwachung der Systemperformanz dauerhaft absichern.

1.4 Maßnahmen gegen negative Auswirkungen der Automatisierung

Maßnahmen gegen negative Auswirkungen KI-basierter Automatisierung lassen sich nur dann angemessen bestimmen, wenn das zugrunde liegende Problem korrekt gefasst wird. Die relevanten Effekte

ergeben sich nicht primär aus individuellen Fehlhaltungen, sondern aus stabilen, strukturell bedingten Interaktionsmustern zwischen technischen Systemen, Arbeitsumgebung und Nutzungskontext. Entsprechend sind sie nicht als subjektive Kommunikationsdefizite zu analysieren, sondern als Folgen bestimmter Gestaltungs- und Organisationsentscheidungen. Appellative Hinweise wie »Vertrauen Sie der KI nicht zu sehr« setzen am bewussten Urteil an, während viele problematische Effekte aus Routinisierung, Zeitdruck, Aufgabenstruktur und systemseitigen Voreinstellungen hervorgehen. Zwar kann es situativ notwendig sein, Aufmerksamkeit gezielt zu bündeln, etwa durch Fokus-Instruktionen (Steinborn, Langner & Huestegge, 2017); dies bleibt jedoch punktuell. Wo Handeln überwiegend durch Arbeitsumgebung und Werkzeuglogik geprägt ist, ersetzen instruktionale Eingriffe keine dauerhafte Prävention. Aus arbeitspsychologischer Perspektive sind Fehlgebrauch, Nichtgebrauch und Missbrauch von Technologie zu erwartende Resultate spezifischer Kopplungen von Tätigkeit, Technik und Kontext (Parasuraman & Riley, 1997). Daraus folgt, dass wirksame Gegenmaßnahmen als Systemeigenschaften zu konzipieren sind. Sie müssen in Schnittstellendesign, Befundungsprotokolle, zeitliche Platzierung von KI-Ausgaben, Schulungsformate sowie Rückmelde- und Überwachungsschleifen eingebettet sein und kontinuierlich wirken, unabhängig von individueller Wachsamkeit.

Welche Gegenmaßnahmen angemessen sind, lässt sich nur in Abhängigkeit davon bestimmen, auf welcher Ebene KI in den Arbeitsprozess eingreift. Unterschiedliche Eingriffsebenen verändern unterschiedliche kognitive und organisatorische Prozesse und erzeugen entsprechend verschiedene Risikomuster. Eine einheitliche Behandlung von Automatisierungseffekten wäre daher analytisch nicht haltbar. Systeme, die Aufmerksamkeit lenken, greifen nicht in Urteile selbst ein, sondern in die Such- und Wahrnehmungsdynamik, aus der Urteile hervorgehen. Aufmerksamkeit fungiert dabei als eine Art erkenntnisleitender Vorfilter: Informationen, die nicht in den Fokus gelangen, können später weder geprüft noch korrigiert werden. Vorstrukturierte Aufmerksamkeitslenkung kann daher dazu führen, dass relevante Aspekte systematisch unberücksichtigt bleiben. Der resultierende Fehler entsteht vor dem eigentlichen Urteil und bleibt häufig unbemerkt, da Sucharbeit implizit an das System delegiert wird. Entscheidungsunterstützende Systeme verlagern das Risiko

an eine andere Stelle. Hier steht nicht unterlassene Wahrnehmung im Vordergrund, sondern die Übernahme fehlerhafter Ergebnisse oder das Ausbleiben weiterer Prüfung. Der Fehlerort liegt in der Abschlussphase des Urteilsprozesses. Die Forschung zur Automatisierungsverzerrung zeigt, dass fehlerhafte KI-Empfehlungen handlungsleitend bleiben können, selbst wenn die fachlichen Kompetenzen für eine Korrektur prinzipiell gegeben sind (Goddard et al., 2012; Gaube et al., 2021). Fehleinschätzungen sind damit nicht allein als Mangel an Wissen oder Expertise zu erklären, sondern als Ergebnis spezifischer Prozesskonstellationen, in denen eine menschliche Korrekturleistung bezüglich der KI-Ausgabe nicht aktiviert oder nicht umgesetzt wird. KI-basierte Triage-Systeme greifen schließlich auf organisatorischer Ebene ein, indem sie klinische Priorisierung und Arbeitslast strukturieren. Unter Bedingungen hohen Durchsatzes können sie schleichende Nachlässigkeiten bei wenig salienten Abweichungen begünstigen, wie die Automatisierungsforschung zeigt (Parasuraman et al., 2000).

Aus dieser Differenzierung folgt, dass Gegenmaßnahmen jeweils systemspezifisch anzusetzen sind, da Interventionen ihre Wirksamkeit erst dann entfalten, wenn sie an die jeweilige Automatisierungsform und ihren funktionalen Eingriffspunkt gekoppelt sind. Entsprechend ist es erforderlich, zwischen diversen Vermittlungsmechanismen zu unterscheiden, anstatt Automatisierungseffekte als homogenen Problemtyp zu behandeln. Wirksame Gegenmaßnahmen müssen dabei gezielt an den jeweiligen Vermittlungsprozessen ansetzen, statt allgemein an Wachsamkeit oder Einstellung zu appellieren. Tabelle 1 zeigt eine exemplarische Liste konkreter Gestaltungsmöglichkeiten basierend auf der zuvor entwickelten Analyse; sie macht explizit, wie abstrakte Prinzipien in praktische Maßnahmen zur Reduktion unerwünschter kognitiver Konsequenzen der KI-Nutzung in der Radiologie übertragen werden können. Jede Maßnahme ist dabei einem spezifischen Risiko und einem klar benannten Wirkmechanismus zugeordnet, sodass der jeweilige Eingriffspfad nachvollziehbar bleibt. Auf diese Weise wird deutlich, dass Risikominderung nicht durch allgemeine Vorsicht oder erhöhte Aufmerksamkeit erreicht wird, sondern durch gezielte Eingriffe in kognitive und organisationale Prozesse. Die Tabelle folgt damit demselben Grundprinzip wie die vorausgehende Argumentation: Wirksamkeit entsteht aus der Passung zwischen Automatisierungsform, vermitteltem Pro-

zess und gewählter Intervention. Die ersten Maßnahmen adressieren Verzerrungen im unmittelbaren Umgang mit KI-Ergebnissen, etwa bei Ergebnisübernahme und Prüfungstiefe; weitere Maßnahmen zielen auf Aspekte der Arbeitsgestaltung und auf den langfristigen Erhalt fachlicher Kompetenz.

Tabelle 1: *Maßnahmen zur Risikominderung bei KI-basierter Automatisierung in der Radiologie*

	Maßnahme	**Zielt auf Risiko**	**Mechanismus**
1	Zweitleser-Prinzip: KI als unabhängige Kontrollinstanz bei ausgewählten Aufgaben	Verankerung: vorschnelle Fixierung auf erste Hypothesen	Unabhängigkeit: eigenständige Urteilsbildung vor KI-Konsultation
2	Verifikationsprotokoll: obligatorischer KI-Check vor Befundabschluss	Auslassungsfehler: übersehene oder fälschlich ergänzte Befunde	Aktivierung: bewusste Prüfung statt passiver Übernahme
3	Transparenz: Anzeige aufgabenspezifischer Unsicherheiten und Grenzen	Übervertrauen: überhöhte Glaubwürdigkeit durch scheinbare Objektivität	Kalibrierung: realistisches mentales Modell der KI-Fehlermodi
4	Dezenz: zurückhaltende Gestaltung ohne alarmartige Hervorhebungen	Aufmerksamkeitsbindung: dominante KI-Markierungen lenken ab	Balance: Erhalt der eigenständigen visuellen Exploration
5	Integration: gebündelte Ergebnisdarstellung ohne Systemwechsel	Fragmentierung: kognitive Last durch Multi-Tool-Management	Effizienz: reduzierter Aufwand für Zusammenführung der Information
6	Fehlertraining: Kalibrierung anhand lokaler Fehlerfälle und Audits	Abhängigkeit: schleichende Gewöhnung an KI-Unterstützung	Sensibilisierung: Kenntnis der Randbedingungen und Grenzen
7	Kompetenzerhalt: regelmäßige KI-freie Befundung als Übung	Dequalifizierung: Erosion diagnostischer Fertigkeiten	Praxis: Erhaltung der Kompetenz für seltene Befundmuster

Anmerkung. KI = Künstliche Intelligenz. Die Maßnahmen 1–3 adressieren kognitive Verzerrungen; Maßnahmen 4–7 betreffen strukturell-organisatorische Aspekte sowie den Kompetenzerhalt.

2. Motivationale und emotionale Aspekte KI-basierter Automatisierung

2.1 Identität, Autonomie und Bedeutung der Arbeit

Der Einsatz von KI in der Radiologie verändert nicht nur Abläufe, sondern die Bedeutungsstruktur professioneller Arbeit. Motivational relevant wird diese Veränderung dort, wo sich verschiebt, was als zentrale Leistung gilt und woran professionelle Anerkennung gebunden ist. KI-basierte Automatisierung wirkt damit nicht nur auf Effizienz, sondern auf das Verständnis dessen, was fachliche Leistung ausmacht. Diese Bedeutungsverschiebung lässt sich funktional an der veränderten Tätigkeitslogik festmachen. Arbeit verlagert sich von unmittelbarer Bildauswertung hin zur Steuerung eines diagnostischen Systems. Entsprechend gewinnen Kompetenzen an Gewicht, die auf Integration, Priorisierung und Beurteilung algorithmischer Ergebnisse zielen, während klassische Wahrnehmungs- und Klassifikationsleistungen an Sichtbarkeit verlieren. Der Wandel betrifft damit nicht nur einzelne kognitive Fertigkeiten, sondern auch die symbolische Ordnung professioneller Kompetenz. Diese Neuordnung bleibt nicht äußerlich, sondern wird aktiv in das berufliche Selbstverständnis eingebaut. Identität ist hier nicht als stabile Eigenschaft zu verstehen, sondern als fortlaufender Aushandlungsprozess im Umgang mit sich verändernden Anforderungen. Empirische Arbeiten zeigen, dass diese Anpassung ambivalent ausfallen kann (Perez et al., 2024): Einerseits kann die Reduktion von Routine und die Betonung komplexer Koordination als Aufwertung erlebt werden; andererseits entstehen Spannungen, wenn unabhängige Wahrnehmung und eigenständiges ärztliches Urteil als zentrale Marker professioneller Anerkennung an Bedeutung verlieren. Theorien beruflicher Identitätsbildung verorten solche Spannungen in der Aushandlung von Normen, sozialer Anerkennung und symbolischen Kompetenzmarkern (Cruess et al., 2014; Shonhe & Min, 2025).

Status und Autorität ergeben sich im Arbeitsvollzug nicht allein aus formaler Verantwortung, sondern aus spezifischen Kompetenzmarkern, die im praktischen Handeln sichtbar und erfahrbar werden. KI-basierte Automatisierung kann diese Marker verschieben, indem neue Referenzpunkte für Richtigkeit und Verlässlichkeit etabliert werden. Werden im Arbeitsteam algorithmische Ergebnisse

zur primären Vergleichsfolie, verändert sich die Zuschreibung von Autorität – mitunter auch dann, wenn menschliche Expertise weiterhin funktional erforderlich bleibt. Diese Verschiebung betrifft weniger Kompetenz im engeren Sinn als ihre Wahrnehmbarkeit und soziale Zuschreibung; Expertise wird damit nicht aufgehoben, sondern anders verankert. Automatisierte oder vorlagenbasierte Berichte verstärken diesen Effekt, indem sie die radiologische Tätigkeit als Überwachung, Bestätigung oder Abweichungskontrolle rahmen und sie weniger als eigenständige Problemlösung erscheinen lassen. Dadurch verändert sich die performative Darstellung fachlicher Kompetenz im Arbeitsprozess und damit auch ihre soziale Sichtbarkeit. Qualitative Studien zeigen, dass auf solche Verschiebungen mit diskursiven Strategien reagiert wird, die Kompetenz jenseits reiner Bildinterpretation oder Ergebnisübernahme markieren (Lombi & Rossero, 2024). Autorität wird dabei neu begründet, etwa über Kontextwissen, Koordinationsleistung oder die Fähigkeit, algorithmische Ergebnisse situativ einzuordnen. KI verändert damit nicht nur Arbeitsprozesse, sondern auch die Bedingungen, unter denen fachliche Autorität hergestellt und anerkannt wird.

Diese Verschiebung lässt sich präzisieren, wenn Status und Autorität nicht als formale Zuschreibungen, sondern als relationale und erlebte Größen gefasst werden. Autorität ist dann weniger an Zuständigkeit oder Haftung gebunden als an die Erfahrung, im Handlungsvollzug wirksam beteiligt zu sein. Maßgeblich dabei ist, ob das eigene Handeln als Quelle relevanter Effekte erfahren wird. Ein solches Erleben ist agentisch, insofern es Situationen markiert, in denen Eingriffe nicht nur ausgeführt, sondern als wirksam erlebt werden. Autorität hängt in dieser Perspektive eng mit Selbstwirksamkeit zusammen, verstanden als Erwartung, durch eigenes Handeln kompetent Einfluss nehmen zu können. Sie entsteht nicht aus beruflicher Position und organisationaler Hierarchie, sondern aus der wiederholten Erfahrung, dass eigenes Tun Bedeutung für den Verlauf der Tätigkeit hat. Vor diesem Hintergrund wird deutlich, dass KI-basierte Automatisierung in der Radiologie nicht nur diagnostische Arbeitsabläufe verändert, sondern die Verteilung solcher Wirksamkeitsmarker. Wenn algorithmische Outputs zur maßgeblichen Referenz für Richtigkeit werden, verschiebt sich die Zuschreibung von Wirksamkeit vom handelnden Subjekt zum System, selbst bei fortbestehender menschlicher Verantwortung. Der zentrale Ef-

fekt liegt dabei nicht im Verlust formaler Zuständigkeit, sondern in der Umdeutung der Tätigkeit. Sichtbare Spuren eigener Kompetenz werden seltener, agentische Marker abgeschwächt. Automatisierung wirkt damit direkt auf das Erleben von Handlungsmacht, indem sie die Erfahrungsbedingungen wirksamen Handelns verändert.

Ob jedoch die KI-Nutzung in der Radiologie die wahrgenommene ärztliche Handlungsmacht untergräbt oder stabilisiert, lässt sich nur im Zusammenhang mit ihrem institutionellen Status bestimmen. Entscheidend ist, ob KI als unterstützende Ressource oder als autoritative Instanz in die Organisation eingebettet ist. Diese Unterscheidung ist nicht graduell, sondern betrifft die Art der Zuschreibung von Beiträgen und Kontrolle. Als unterstützende Ressource kann KI Handeln erweitern, ohne die Zuschreibung von Kompetenz grundlegend zu verschieben: Das eigene Urteil bleibt der Referenzpunkt, während algorithmische Ausgaben zusätzliche Orientierung liefern. Wird KI hingegen zur autoritativen Instanz, verändert sich, wem Beiträge zur Richtigkeit der Diagnostik zugerechnet werden. In dieser Konstellation bleibt fachliche Expertise zwar funktional notwendig, verliert jedoch an Sichtbarkeit und Anerkennungswert. Die beobachteten diskursiven Strategien lassen sich vor diesem Hintergrund als aktive Versuche verstehen, unter veränderten Zuschreibungsbedingungen agentische Selbstverortung zu bewahren und Kompetenz jenseits algorithmischer Referenzpunkte sichtbar zu halten. Autonomie zeigt sich dabei nicht in regelkonformer Ausführung, sondern dort, wo eigenes Handeln als wirksam erlebt wird. Konfigurierbare KI-Systeme, die dem klinischen Urteil eindeutig nachgeordnet bleiben, können Handlungsspielräume sichern. Wird der Einsatz von KI dagegen verpflichtend und Workflow-bestimmend (z. B. erforderliche Eingabeaufforderungen, erzwungene Triage, Audits), verschiebt sich das Erleben von Autonomie hin zu externer Kontrolle und Überwachung, was wiederum die intrinsische Motivation verringert und zu konformitätsorientiertem, unkritischem Verhalten im Umgang mit KI-Ergebnissen führen kann. Autonomie erweist sich damit als Ergebnis konkreter Design- und Organisationsentscheidungen bezüglich der Einbettung von KI-Anwendungen, nicht als stabile Eigenschaft professioneller Rollen.

Die Einführung von KI-Anwendungen in der Radiologie kann also die Bedeutung professioneller Arbeit verändern und damit unmittelbar auf die Arbeitsmotivation wirken. Diese Veränderung ist

jedoch über diverse KI-Anwendungen nicht einheitlich, sondern teils sogar durch gegenläufige Effekte gekennzeichnet: Der Wegfall von Routinetätigkeiten kann Arbeit als anspruchsvoller, interessanter und wertiger erscheinen lassen, während zugleich der Verlust von Eigenverantwortung und sichtbarer Urheberschaft dazu führen kann, dass Arbeit an Bedeutung verliert. Motivation entsteht aus dem Zusammenspiel dieser beiden Effekte, nicht aus einem davon isoliert. Die Bedeutungszuschreibung ist dabei handlungsrelevant: Sie beeinflusst, ob zusätzliche Prüfungen vorgenommen, Zweitbegutachtungen eingeholt oder Diskrepanzen aktiv als Lerngelegenheiten genutzt werden. Motivation fungiert hier als vermittelnde Größe zwischen der Wahrnehmung von Wertigkeit und Verantwortung einerseits und konkretem sicherheitsrelevantem Verhalten andererseits. Qualitative Interviews zeigen, dass radiologische KI-Tools bislang selten formale Entscheidungsautonomie entziehen, jedoch die Stellung ärztlicher (menschlicher) Kompetenz und Autorität als Referenz für Richtigkeit in Frage stellen (Lombi & Rossero, 2024). Der Eingriff erfolgt damit nicht über explizite Restriktionen, sondern über eine veränderte symbolische Ordnung der Tätigkeit. In dieser Konstellation können Stress und Bedrohungserleben jene kognitiven Muster (z. B. Automatisierungsverzerrungen) verstärken, die in Kapitel 1 als sicherheitsrelevant beschrieben wurden. Motivation wird so zu einem Faktor, der bestehende Risikodynamiken abschwächen oder verstärken kann, abhängig davon, wie Bedeutung und Verantwortung im Arbeitsvollzug erlebt werden.

2.2 Emotionale Reaktionen: Unsicherheit, Sorgen und Erleichterung

Emotionale Konsequenzen des KI-Einsatzes in der Radiologie speisen sich nicht ausschließlich aus konkreten Nutzungserfahrungen, sondern in erheblichem Maße aus Erzählungen über die ärztliche Ersetzbarkeit durch KI. Solche Narrative entfalten ihre affektive Wirkung auch dann, wenn sie nicht als realistische Prognosen akzeptiert werden, denn sie erzeugen eine anhaltende Unsicherheit darüber, welchen Wert fachliche Expertise künftig besitzt, wie sich Karriereverläufe entwickeln und welche Bedeutung die erfahrene Aus- und Weiterbildung behält (Coppola et al., 2021). Diese Unsicherheit ist

nicht punktuell, sondern strukturell, da sie Erwartungen über zukünftige Handlungs- und Anerkennungsbedingungen betrifft. Die Verteilung dieser Unsicherheit ist dabei sozial und biografisch differenziert: In frühen Karrierestufen richtet sie sich vor allem auf Fragen zur (vermeintlichen) Automatisierbarkeit des radiologischen Tätigkeitsfeldes, während sie in späteren Phasen die Neubewertung bereits aufgebauter Expertise betrifft. Damit verschiebt sich der affektive Fokus von antizipierter Eintrittsunsicherheit zu retrospektiver Sinn- und Wertprüfung. Internationale Umfragen und qualitative Arbeiten zeigen, dass solche Sorgen Einstellungen zur Einführung von KI systematisch prägen (Gong et al., 2019; Huisman et al., 2021; Rony et al., 2024; Dang & Li, 2025). Die Ausprägung dieser Effekte variiert dabei mit radiologischer Erfahrung, Wissen über KI und der wahrgenommenen institutionellen Ausrichtung. Emotionale Reaktionen sind damit nicht hauptsächlich Ausdruck individueller Dispositionen, sondern Ergebnis spezifischer Erwartungs- und Deutungsrahmen, in denen die KI-Nutzung verortet wird.

Emotionale Reaktionen im praktischen Umgang mit KI-Anwendungen entstehen bevorzugt dann, wenn die eigene (ärztliche) Einschätzung mit einem KI-basierten Ergebnis divergiert. Eine solche Abweichung markiert nicht automatisch einen Fehlerhinweis, sondern zunächst einen offenen Bewertungsmoment, der eine Positionierung erforderlich macht. Aus dieser Situation können unterschiedliche affektive Dynamiken hervorgehen: Die Abweichung kann zum Anlass werden, die eigene Begründung zu prüfen, Evidenz erneut zu sichten und den Fall aktiv weiterzubearbeiten, was emotional als positiv erlebt wird. Sie kann jedoch auch als Hinweis auf die Unzulänglichkeit der eigenen Einschätzung interpretiert werden, sodass sich Kompetenzängste herausbilden und sich die weitere Bearbeitung zunehmend an der algorithmischen Vorgabe orientiert. In der ersten, emotional positiven Dynamik bleibt die ärztliche Urteilspraxis aktiv und die KI fungiert als Kontrastfolie für Präzisierung und Lernen; in der zweiten, emotional eher negativen Dynamik zieht sich die handelnde menschliche Instanz aus der initiativen Beurteilung zurück. Welche dieser Dynamiken einsetzt, ergibt sich vor allem aus lokalen Bewertungs- und Rückmeldestrukturen. Entscheidend dabei ist, ob Abweichungen zwischen eigener ärztlicher Einschätzung und KI-Systemausgabe als legitimer Bestandteil professioneller Urteilsbildung behandelt werden oder als erklärungsbedürfti-

ge Fehler. Affektive Reaktionen fungieren in diesem Sinne nicht als Störgrößen, sondern als Umschaltpunkte erhöhter Sensitivität, an denen sich entweder aktive Urteilspraxis und Lernorientierung stabilisieren oder Abhängigkeit und Passivierung ausbilden.

Emotionale Reaktionen im Umgang mit KI lassen sich auch als Indikatoren dafür verstehen, wie gut KI funktional in den radiologischen Arbeitsvollzug integriert ist, also ob die ärztliche Interaktion mit dem jeweiligen KI-Tool als unterstützend und entlastend oder als reibungsreich und anstrengend erlebt wird. Negativer Affekt entsteht im praktischen Umgang nämlich häufig aus dem Erleben operativer Schwierigkeiten im Arbeitsprozess. Wiederholte Fehlalarme, schwer integrierbare Darstellungen oder zusätzlicher Dokumentationsaufwand erhöhen die Interaktionskosten und verkomplizieren den Arbeitsprozess, auch dann, wenn das KI-System in anderen Situationen hilfreiche Beiträge liefert. Daneben lassen sich aber auch klar konturierte positive Affekte bei der KI-Nutzung beobachten: Erleichterung entsteht dort, wo KI monotone, zeitaufwendige und gleichzeitig aufmerksamkeitsfordernde Teiloperationen übernimmt und dadurch im selben Handlungskontext ein spürbarer Übergang zu geringerer Beanspruchung erfahren wird. Die affektive Qualität resultiert dabei aus der wahrgenommenen Diskontinuität zwischen einer Phase hoher Belastung und einem nachfolgenden Zustand erweiterter Handlungsspielräume. Darüber hinaus kann die KI-Nutzung auch affektiv positiv bewertete Selbstzuschreibungen wie Stolz erzeugen, z. B. wenn Abweichungen zwischen eigener (korrekter) Einschätzung und algorithmischem Output begründet und der eigenen Leistungsfähigkeit zugerechnet werden können. In diesem Moment wird KI als begrenztes System erfahrbar, an dessen Grenzen menschliche Kompetenz sichtbar wird. Erleichterung markiert damit gelingende funktionale Entlastung im zeitlichen Vollzug der Arbeit, Stolz eine gelingende Selbstzuschreibung von Kompetenz im Urteilsvollzug.

Aktuelle theoretische Ansätze gehen davon aus, dass emotionale Reaktionen keine epiphänomenalen Begleiterscheinungen sind, sondern integraler Bestandteil kognitiver Regulation mit eigener funktionaler Rolle (Damasio, 1994; Cosmides & Tooby, 2000; Schwarz & Clore, 2007). Sie sind systematisch an Situationen gebunden, in denen Handlungen und Entscheidungen als folgenreich antizipiert werden; in diesem Sinne tragen Emotionen zur Gewichtung noch

nicht vollständig spezifizierter Konsequenzen bei. Sie wirken als regulatorische Signale, die Relevanz strukturieren, Prioritäten setzen und Orientierung unter Unsicherheit ermöglichen (Loewenstein et al., 2001; Lerner et al., 2015). Im diagnostischen Arbeiten mit KI greifen emotionale Reaktionen früh in Wahrnehmung, Aufmerksamkeitsverteilung, Entscheidungsfindung und Handlungssteuerung ein. Sie modulieren, wie breit oder fokussiert die Aufmerksamkeit ausgerichtet wird, wie lange Ambiguität toleriert wird und wie Abweichungen zwischen eigener Einschätzung und algorithmischem Ergebnis gewichtet werden. Diese Wirkungen sind kontextabhängig differenziert: Emotionale Reaktionen entfalten keine einheitliche Wirkungsrichtung, sondern modulieren kognitive Prozesse abhängig von Intensität, Belastung und verfügbarer Handlungsfreiheit. Zustände wie Angst, Unsicherheit oder Überforderung können Aufmerksamkeit verengen und Prüfprozesse verkürzen; moderate Anspannung kann dagegen die Sensitivität für Inkonsistenzen erhöhen und eine sorgfältigere Prüfung unterstützen. Man kann also sagen, Emotionen fungieren als ein Bestandteil der Prozesssicherheit, allein schon deswegen, weil sie mitbestimmen, wie ärztliche Urteile gebildet werden und wie stabil diese Urteile gegenüber den Grenzen, aber auch Verheißungen technischer Systeme bleiben (Croskerry et al., 2013).

2.3 Angst vor Fehlern und moralischer Distress

Die Frage der Verantwortung für diagnostische Entscheidungen (s. Abschnitt 3.1) bildet einen zentralen emotionalen Bezugspunkt, an dem Technologieeinsatz, Vertrauen und wahrgenommene moralische Verpflichtung zusammenlaufen. Nach geltender Gesetzgebung verbleibt die Verantwortung für klinische Entscheidungen bei der Ärzteschaft, aber im praktischen Einsatz kann diese Zuordnung de facto verschoben werden, wenn KI-Empfehlungen zugleich als maßgeblich und in ihrer Entstehung nicht nachvollziehbar wahrgenommen werden. Unter solchen Bedingungen kann Unsicherheit darüber entstehen, welche Handlungsoptionen als legitim gelten. Dies ist insbesondere dann der Fall, wenn davon ausgegangen wird, dass jene ärztlichen Einschätzungen, die vom KI-Ergebnis abweichen, besonders kritisch geprüft werden oder dass die Anpassung

an algorithmische Empfehlungen administrativ sicherer ist als ein begründeter Widerspruch. Die formale Verantwortung bleibt dabei bestehen, während die erlebte und erwünschte Handlungsfähigkeit eingeschränkt wird.

Aus dieser Konstellation heraus kann die Sorge vor Fehlern Verhaltensweisen begünstigen, die primär auf Absicherung und formale Korrektheit zielen, etwa ein übermäßig detailliertes Überprüfen wenig relevanter Befunde, ausgedehnte Dokumentationspraktiken zur haftungsbezogenen Absicherung oder auch ein unverhältnismäßiges Übernehmen technischer Hinweise. Die ärztliche Kontrollfunktion verschiebt sich damit vom patientenfokussierten Prüfen von Ergebnissen hin zum Erfüllen formaler Anforderungen, womit die Technologie ihren operativen Zweck im radiologischen Arbeitsprozess verändert: Was als Unterstützung eingeführt wird, fungiert als Auslöser ärztlicher Verhaltensanpassungen (d. h. Absicherungsstrategien und »Compliance-Rituale«), die die beabsichtigten Effekte abschwächen oder umkehren (Lebovitz et al., 2021). Moralisches Belastungserleben (»moral distress«; Jameton, 1984) entsteht dabei, wenn die verantwortungsorientierte ärztliche Praxis dauerhaft mit institutionellen Erwartungen oder systemseitigen Vorgaben kollidiert, wenn also die moralische Verpflichtung zu patientenbezogener Vorsicht und Umsicht in Konflikt zu Anforderungen wie Durchsatz (also Arbeitstempo), Standardisierung oder Nutzung KI-gestützter Abläufe gerät (Kherbache et al., 2022). Diese Lage wird besonders belastend, wenn Zweifel an der Eignung des KI-Systems für bestimmte Kontexte bestehen. Persistiert die Spannung, kann sie zermürbend wirken und die Arbeitsleistung beeinträchtigen: Arbeitszufriedenheit und -motivation nehmen ab, Frustrations- und Schuldgefühle können entstehen, und professionelle Haltung kann in Zynismus kippen (Dave et al., 2023). Moralischer Distress ist damit weniger ein individuelles Problem als das Resultat anhaltender organisational-struktureller Inkongruenzen mit dem ärztlichen Selbstbild und Arbeitsethos.

Die zuvor beschriebenen Konflikte machen deutlich, dass ärztliche Verantwortung und moralische Integrität unter Bedingungen technischer Intransparenz und organisationaler Spannung ohne strukturelle Absicherung kaum tragfähig bleiben. Schutzfaktoren müssen folglich institutionell, also im soziotechnischen System selbst, verankert sein und dürfen nicht an persönliche Belastbarkeit

oder situative Selbstregulation delegiert werden. Zentrale Elemente einer solchen Absicherung sind klare Verantwortungsnormen, die Zurechnungsunsicherheit reduzieren, sowie psychologisch sichere Wege zur Eskalation KI-bezogener Bedenken, die Zweifel artikulierbar machen, ohne Sanktionen oder auch nur Sanktionserwartungen auszulösen. Partizipative Formen der KI-Implementierung in der Radiologie dabei die Kontrolle über Einsatzbedingungen stärken, während kollektive Diskussionsformate Abweichungen als gemeinsame Bearbeitungsgegenstände rahmen, statt sie zu individualisieren. Auf diese Weise bleiben Konflikte bearbeitbar und die ärztliche Aufsicht wird stabilisiert, anstatt in defensive Routinen abzugleiten. Diese Strukturen verweisen zugleich auf weiterführende Implementierungsfragen, die im weiteren Verlauf systematisch aufgegriffen werden. Zunächst richtet sich der Fokus jedoch auf Verantwortung und Erklärbarkeit als zentrale Bedingungen eines angemessenen Umgangs mit KI.

3. Verantwortung und Erklärbarkeit

3.1 Verantwortung und Rechenschaft bei hybriden Entscheidungsprozessen

Im Kontext klinischer KI-Nutzung ist Verantwortung nicht allein als rechtliche Haftungszuschreibung zu verstehen, sondern als psychologisches und organisatorisches Moment, das Handeln unter Unsicherheit strukturiert. Wenn KI in operative radiologisch-diagnostische Arbeitsschritte wie Fallvorsortierung, Mustererkennung oder Prioritätensetzung eingreift, entsteht ein hybrider Entscheidungsprozess, in dem menschliche und algorithmische Beiträge funktional verschränkt sind. Verantwortung lässt sich in solchen Konstellationen nicht mehr eindeutig lokalisieren, sondern wird entlang unterschiedlicher Ebenen gerahmt. Auf der Ebene des Entscheidungsergebnisses bleibt Verantwortung (bislang) an die ärztliche Befundung gebunden, während sie organisational häufig als verteilt konzipiert wird und die Technologie als unterstützende Instanz erscheint. Diese Gleichzeitigkeit inkonsistenter Rahmungen erzeugt Spannungen, die das Verantwortungserleben im Arbeitsvollzug beeinflussen. Aus der Sozialpsychologie ist seit langem bekannt, dass in Situationen, in

denen mehrere Betroffene oder Instanzen beteiligt sind, die vom Einzelnen wahrgenommene Verantwortung »diffundieren«, sich also ausdünnen, kann (Darley & Latané, 1968). Im Kontext der radiologischen KI-Nutzung kann dabei die Annahme vorherrschen, die Sicherheit der eingesetzten Technologie sei bereits an vorgelagerten Stellen hergestellt worden, etwa durch wissenschaftliche Validierung, technische Zertifizierung, institutionelle Beschaffung oder formale Freigabeprozesse. Diese Vorverlagerung von Zuständigkeit und Diffusion von Verantwortlichkeiten kann dann auf ärztlicher Seite zu nachlassender Wachsamkeit führen, da Risiken als bereits abgefedert oder abgesichert gelten. Zugleich können »moralische Knautschzonen« entstehen, in denen die Verantwortung für KI-gestützte Fehlentscheidungen der Ärzteschaft zugeschrieben wird, obwohl deren tatsächlicher Einfluss auf KI-Systemdesign, Trainingsdaten oder Schnittstellengestaltung begrenzt ist (Elish, 2019).

Die Rechenschaftspflicht im Kontext radiologischer KI-Nutzung ist nicht allein organisationsintern bestimmt, sondern ergibt sich aus regulatorischen Anforderungen an Risikomanagement, diagnostische Transparenz und ärztliche Aufsicht. Diese Anforderungen verleihen einer moralisch verankerten ärztlichen Verantwortung einen rechtlich verbindlichen Charakter. Um diesen Pflichten gerecht zu werden, müssen in der klinischen Praxis Kontrollprozesse konkret definiert und organisatorisch verankert werden, etwa für eine lokale Validierung und Aktualisierung von KI-basierten Anwendungen, die Dokumentation von Abweichungen sowie die Überwachung von Leistungsänderungen des KI-Systems über die Zeit. Die Umsetzung dieser Prozesse erzeugt zusätzlichen Lern- und Koordinationsaufwand und verändert etablierte Routinen, Zeitbudgets und Kompetenzanforderungen. Rechenschaftspflicht fungiert damit nicht als bloße formale Vorgabe, sondern greift unmittelbar in den radiologischen Arbeitsalltag ein. Neben der Berücksichtigung des zusätzlichen Aufwands ist dabei entscheidend, wie klar die Zuständigkeiten geregelt sind: Bleibt unbestimmt, wer welche Kontrollaufgaben trägt, entsteht Unsicherheit darüber, was erwartet wird und welche Handlungen als angemessen gelten. Diese Erwartungsunsicherheit begünstigt wiederum Stress und Abwehrreaktionen. Ob Rechenschaftspflichten bezüglich der Nutzung von KI in der Radiologie zur Stärkung von Sicherheit beitragen oder primär die ärztliche Belastung erhöhen, hängt also wesentlich vom zusätzlichen Arbeits-

aufwand sowie der Klarheit der Prozess- und Zuständigkeitsstruktur ab.

Eine weitere Herausforderung für ärztliche Verantwortlichkeit und Rechenschaft in hybriden KI-gestützten Entscheidungsprozessen ist die Schwierigkeit, Ergebnisse von KI-Systemen nachzuvollziehen und ggf. zu hinterfragen. Ohne die Möglichkeit, KI-Ergebnisse anzufechten, ist eine verantwortliche Nutzung nicht denkbar. Opake (d. h. technisch wenig oder nicht durchschaubare) Systeme, wie es typischerweise KI-Systeme sind (»Black Box«-Charakter), verschieben diese Möglichkeit, indem sie begründeten Widerspruch aufwendig machen. Denn dort, wo Erklärungen für Ergebnisse des technischen Systems nicht zugänglich sind, steigt der Aufwand für Prüfung und Begründung erheblich. Die ärztliche Letztverantwortung bleibt also formal bestehen, wird praktisch jedoch schwer einlösbar. Aus diesem Konflikt ergeben sich soziotechnische Gestaltungsfragen, etwa zur Erklärbarkeit von KI oder zur Zuständigkeit für die Widerspruchsprüfung.

3.2 Von der Erklärbarkeit zur Interpretierbarkeit

Forderungen nach erklärbarer KI beruhen häufig auf der impliziten Annahme, dass mehr Transparenz den klinischen Einsatz grundsätzlich verbessert. Diese Annahme ist nicht grundsätzlich falsch, greift jedoch zu kurz, da sie Kriterien aus der technischen Entwicklung, Überprüfung oder Regulierung unreflektiert auf situative radiologische Entscheidungskontexte überträgt. Im klinischen Alltag erfüllen Erklärungen keinen Selbstzweck, sondern sie sollen der Ärzteschaft helfen, sich *angemessen* auf KI-Tools zu verlassen, auch ohne jedes technische Detail zu kennen. D. h., sie sollen die KI-Nutzung im diagnostischen Urteilsprozess steuern, indem sie zwei symmetrische Fehlerformen begrenzen: die unkritische Übernahme von Ergebnissen wie auch ihre reflexhafte Zurückweisung. Diese Steuerung durch Erklärbarkeit erfüllt dabei in erster Linie die folgenden drei Funktionen: Sie stellt Anfechtbarkeit her (sodass KI-Ergebnisse sinnvoll hinterfragt werden können), bewirkt Lernen (sodass Fehlermodi im Laufe der Zeit besser verstanden werden) und vermittelt Vertrauenskalibrierung (sodass klar wird, wann das KI-System wahrscheinlich richtig oder falsch liegt) (Doshi-Velez & Kim, 2017; Miller, 2019).

Eine weitere wichtige Differenzierung im Kontext der Erklärbarkeit von KI-gestützten Ergebnissen in der Radiologie ist, dass verschiedene Arten von Erklärungen unterschiedlichen Zielgruppen und Zwecken dienen: Einige zielen darauf ab, einen allgemeinen Eindruck davon zu vermitteln, wie das betreffende KI-System insgesamt funktioniert; andere konzentrieren sich darauf, warum das Modell in einem bestimmten Fall ein bestimmtes Ergebnis geliefert hat; und wieder andere vermitteln das Ausmaß der Ergebnisunsicherheit oder »Was-wäre-wenn«-Szenarien. Dementsprechend bemisst sich die Relevanz von KI-Erklärbarkeit an ihrem konkreten Zweck und Nutzen im klinischen Alltag – aber auch an ihren Kosten. Denn diese diversen Erklärungstypen unterscheiden sich in kognitiver Belastung, Zeitbedarf und vorausgesetzter Expertise. Detaillierte technische Begründungen können z. B. eine zeitkritische Befundung belasten, während (über)vereinfachte Darstellungen relevante Einschränkungen der Ergebnisgültigkeit verdecken können. Aus diesen Zielkonflikten folgt, dass nicht maximale (und damit oft überbordend informationsreiche) Transparenz leitend sein kann, sondern zweckmäßige Interpretierbarkeit. Gemeint ist die Passung zwischen Erklärung, Nutzungskontext und Entscheidungsrisiko, d. h. Erklärungen, die gerade so detailliert sind, dass sie das Fehlerrisiko im spezifischen klinischen Kontext verringern oder das spezifische Kommunikationsziel optimal erreichen (Amann et al., 2020; Lipton, 2018, Rudin, 2019). Unter- und Überinformation stellen dabei symmetrische Fehlformen dar, da beide falsche Sicherheit erzeugen und eine angemessene KI-Nutzung unterminieren.

Beachtenswert für die ärztliche Kommunikation aber auch das eigene ärztliche Verständnis von KI-Ergebnissen ist, dass (vermeintliche) Erklärungen epistemische Risiken erzeugen können, wenn sie Verstehen suggerieren, ohne die Fähigkeit zur Bewertung oder Anfechtung von Ergebnissen zu erweitern. Anschauliche Visualisierungen und vereinfachte Darstellungen erzeugen oft subjektive Plausibilität und stärken Vertrauen, ohne eine wirkliche Erklärung (d. h., einen Einblick in das tatsächliche Verhalten des KI-Modells) bereitzustellen (Lipton, 2018; Miller, 2019). Solche oberflächlichen Scheinerklärungen wirken dann nicht klärend, sondern vermindern die Sicherheit, wenn sie als zuverlässige Belege und nicht als fehlbare Entscheidungshilfen betrachtet werden und sich (überhöhtes) Vertrauen vom tatsächlichen Systemverhalten zunehmend entkoppelt

(Ghassemi et al., 2021). Verantwortliche Nutzung verlangt daher weniger fallbezogene Illustration als Wissen darüber, unter welchen Bedingungen KI-Ergebnisse Geltung im radiologisch-diagnostischen Entscheidungszusammenhang insgesamt besitzen oder verlieren, z B. Wissen darüber, auf welchen Populationen und Scannern das KI-Modell trainiert wurde, für welche Indikationen es vorgesehen ist und an welchen Grenzen seine Leistung bekanntermaßen nachlässt (Amann et al., 2020; Rudin, 2019). Entscheidend ist, ob diese Bedingungen und dieses Wissen im klinisch-radiologischen Arbeitsprozess vorhanden bleiben und praktisch berücksichtigt werden. Nur so bleiben verantwortliche diagnostische Entscheidungen in der Radiologie auch bei algorithmischer Unterstützung durch technisch opake KI-Systeme dauerhaft möglich (Ghassemi et al., 2021; Miller, 2019; London, 2019).

4. Vertrauen in KI-basierte Technologien in der Radiologie

4.1 Vertrauen und Zuverlässigkeit

Der Einsatz von KI in der Radiologie setzt ein Mindestmaß an Vertrauen in Leistungsfähigkeit und Sicherheit der jeweiligen Anwendung voraus. Ziel klinischer KI-Nutzung ist jedoch nicht die Maximierung dieses Vertrauens, sondern dessen Kalibrierung: KI-basierte Verfahren sollen dort eingesetzt werden, wo sie im jeweiligen Kontext mit hoher Wahrscheinlichkeit zu besseren Entscheidungen beitragen, und dort unterbleiben, wo systematische Fehlentscheidungen zu erwarten sind. Vertrauen wird damit zu einer steuerungsrelevanten Größe, die in konkrete Einsatz- und Nutzungsentscheidungen übersetzt werden muss. Um diese Steuerungsfunktion analytisch fassen zu können, ist eine begriffliche Trennung zweier häufig vermischter Konstrukte erforderlich. Vertrauen bezeichnet eine Einstellung, verstanden als Erwartung an Kompetenz, Vorhersagbarkeit und Intentionen eines Systems. Es beschreibt, was von einer Anwendung angenommen wird, bevor gehandelt wird. Verlässlichkeit bezieht sich demgegenüber auf tatsächliches Nutzungsverhalten, also darauf, wie KI-Ergebnisse im klinischen Entscheiden und Handeln aufgegriffen, geprüft oder ignoriert werden, wie also mit einer KI-Anwendung faktisch umgegangen wird. Einstellung und Verhalten

können dabei auseinanderfallen: Ein hohes Maß an Vertrauen kann mit zurückhaltender Nutzung einhergehen und umgekehrt. Für die Analyse klinischer KI-Nutzung in der Radiologie ist es daher notwendig, Vertrauen und Verlässlichkeit getrennt zu betrachten und ihre Beziehung nicht vorauszusetzen, sondern empirisch und kontextbezogen zu bestimmen.

Empirische Arbeiten zeigen, dass ein hohes Maß an Vertrauen in KI-basierte Technologien dort problematisch wird, wo es zu einer erwartungsbasierten Reduktion von Überwachung und aktiver Kontrolle der KI-Systeme führt. In solchen Konstellationen werden KI-Ergebnisse weniger geprüft, nicht weil Bedenken fehlen, sondern weil Zuverlässigkeit (Verlässlichkeit) antizipiert wird: Kontrolle unterbleibt, da Fehler als unwahrscheinlich gelten (Parasuraman & Riley, 1997; Lee & See, 2004). Eine gegenläufige Fehlerform entsteht bei zurückhaltendem Vertrauen. In diesem Fall werden Systeme selbst dann nicht genutzt, wenn sie unter gegebenen Bedingungen verlässlich sind und objektiv Leistungsgewinne ermöglichen. Fehlanpassung zeigt sich hier nicht als Übernahme falscher Ergebnisse, sondern als Unterausnutzung verfügbarer Unterstützung. Beide Konstellationen verdeutlichen, dass Vertrauen weder maximal noch minimal sein sollte, sondern aufgabenspezifisch ausgerichtet (d. h. kalibriert) werden muss. Kalibriertes Vertrauen bezeichnet vor diesem Hintergrund den aufgaben- und kontextsensitiven Einsatz KI-basierter Verfahren. Ein System kann für Screening-nahe Erkennungsaufgaben gut geeignet sein, unter artefaktreichen Bildprotokollen, bei selten eingesetzten Geräten oder veränderten Prävalenzen jedoch an Verlässlichkeit verlieren. Zusätzlich ist zu berücksichtigen, dass Verzerrungen in den Trainingsdaten die Aussagekraft für bestimmte Patientengruppen einschränken können, was – bei Nichtbeachtung – zu einer unangemessene Einordnung KI-basierter Hinweise im klinischen Alltag führen kann (Kocak et al., 2025).

Vertrauen in KI-Systeme speisen sich im klinischen Alltag parallel aus mehreren Quellen: aus beobachteter Leistungshistorie, aus systemseitigen Transparenz- und Statushinweisen sowie aus sozialen und institutionellen Signalen, die den Einsatz der Technologie rahmen (Lee & See, 2004; Hoff & Bashir, 2015). Diese Quellen unterscheiden sich funktional und sind nicht gleich verlässlich, wirken jedoch gemeinsam auf die Vertrauensbildung ein. Die tatsächliche Zuverlässigkeit eines Systems ist im klinischen Alltag kaum direkt

erfahrbar. Einschätzungen stützen sich daher auf einzelne Nutzungsepisoden statt auf systematische Leistungsprofile, wobei Ersteindrücke, besonders markante Erfolge oder auffällige Fehler sowie Hinweise auf einen hohen »Automatisierungsstatus« des KI-Systems das Vertrauen überproportional prägen und als heuristische Anker wirken können. Gestalterische Merkmale fungieren in diesem Zusammenhang als Glaubwürdigkeitsindikatoren: Benutzeroberfläche, Sprachstil, Visualisierungsformen oder infrastrukturelle Einbettung beeinflussen, wie kompetent und zuverlässig ein System erscheint, ohne notwendigerweise mit seiner tatsächlichen Leistungsfähigkeit zu korrespondieren. Vertrauenskalibrierung ist daher weniger das Ergebnis individueller Beurteilungsfähigkeit, als das Produkt eines soziotechnischen Arrangements, das bestimmt, welche Leistungsbelege sichtbar sind und wie organisationale Reputationssignale vermittelt werden. Fehlkalibriertes Vertrauen ist in diesem Sinne primär ein Gestaltungs- und Organisationsproblem.

4.2 (Fehl-)Kalibrierung des Vertrauens: Folgen, Dynamik und Modifikation

Fehlkalibriertes Vertrauen in KI-Anwendungen in der Radiologie äußert sich in qualitativ unterschiedlichen problematischen Nutzungsmustern, nämlich Fehlgebrauch, Nichtgebrauch und Missbrauch. Fehlgebrauch wird durch Übervertrauen begünstigt, indem fehlerhafte KI-Systemvorschläge oder -ergebnisse akzeptiert und unabhängige Such- und Prüfaktivitäten eingeschränkt werden. Der zentrale Effekt liegt dabei weniger in der Übernahme falscher KI-basierter Hinweise, als in der Reduktion der ärztlichen Suchbreite, wodurch von der KI unbemerkte und nicht markierte, aber potenziell relevante Befunde häufiger übersehen werden, weil ärztliche Vigilanz und Exploration zugunsten systemseitiger Hinweise zurücktreten. Untervertrauen führt demgegenüber zu Nichtgebrauch: KI-Ergebnisse bleiben trotz nachgewiesenem Nutzen unberücksichtigt, etwa infolge von Skepsis, Frustration oder mangelnder KI-Integration in den Arbeitsablauf. Nichtgebrauch ist damit kein rationales Abwägen im Einzelfall, sondern ein organisations- und erwartungsbedingter Nutzungsausfall. Missbrauch schließlich stellt ein drittes, eigenständiges Muster dar, das den Einsatz von KI außerhalb validier-

ter Bedingungen, etwa bei anderen Modalitäten, Populationen oder Erfassungssettings. Häufig liegt dem keine bewusste Regelverletzung zugrunde, sondern eine inadäquate Übergeneralisierung, also eine spezifische Form des Übervertrauens, oft unter Zeit-, Kosten- oder Leistungsdruck. Fehlgebrauch, Nichtgebrauch und Missbrauch sind daher als unterschiedliche Fehlerformen im Umgang mit KI-Technologien in der Radiologie zu behandeln, die jeweils spezifische Gegenmaßnahmen erfordern und nicht durch eine einheitliche Vertrauensstrategie adressiert werden können (Parasuraman & Riley, 1997).

Die Kalibrierung von Vertrauen in radiologische KI-Anwendungen ist kein einmaliger Akt, sondern ein fortlaufender Anpassungsprozess, denn das Vertrauen verändert sich mit KI-Nutzungserfahrungen, Systemaktualisierungen und Aufgabenwechseln und bleibt damit prinzipiell reversibel. Diese Dynamik ist jedoch unter klinischen Bedingungen systematisch verzerrt, da die Voraussetzungen für verlässliches Lernen nur eingeschränkt gegeben sind. Zentral ist dabei die typische Struktur des Feedbacks in der Radiologie: Rückmeldungen über die tatsächliche Leistung eines KI-Systems sind häufig verzögert, unvollständig oder verrauscht. Diagnostische Konsequenzen werden oft erst nach Wochen sichtbar oder lassen sich nicht eindeutig einer einzelnen KI-Systemausgabe zuordnen. Unter diesen Bedingungen wird die Anpassung bestehender Überzeugungen unsicher, da eine einzelne Erfahrung kein stabiles Lernsignal liefert. Hinzu kommt eine asymmetrische Gewichtung von Ereignissen, wobei einzelne auffällige Fehler überproportional vertrauenssenkend wirken und abrupte Nichtnutzung auslösen können, obwohl die mittlere Systemleistung weiterhin hoch ist. Umgekehrt können wiederholte unauffällige Erfolge oder plausible, aber oberflächliche Fehlererklärungen Vertrauen schleichend und unangemessen normalisieren und dadurch langfristig eine nachlassende Wachsamkeit und Sorglosigkeit begünstigen (Dzindolet et al., 2003; Hoff & Bashir, 2015; Lee & See, 2004). Ein weiterer Mechanismus der Fehlkalibrierung betrifft die inadäquate Übertragung von Vertrauen zwischen verschiedenen Aufgaben einer KI-Anwendung (z. B. von der Triage mittels Brust-CT auf die subtilere onkologische Stadieneinteilung). Diese Generalisierung ist epistemisch unbegründet, da Leistungsprofile aufgaben- und kontextspezifisch sind. Sie stellt damit eine Form des Missbrauchsrisikos dar, wobei einheitliche Benut-

zeroberflächen, konsistenter Sprachstil oder gemeinsame Herstelleridentitäten dieses Risiko noch verstärken können, indem Vertrauen an Design oder Marke gebunden wird statt an nachgewiesene Aufgabenleistung. In der Summe zeigt sich, dass die Dynamik von Vertrauen weniger durch individuelle Urteilskraft bestimmt ist als durch Lernbedingungen, Ereignissalienz und gestalterische Rahmungen. Fehlkalibrierung von Vertrauen in radiologische KI-Anwendungen ist damit kein zufälliges Nebenprodukt, sondern ein Ergebnis der Art und Weise, wie Rückmeldung, Sichtbarkeit und Aufgabenbezug im klinischen Alltag organisiert sind.

Ein ausgewogenes Vertrauen in KI-basierte Systeme erfordert eine fortlaufende Kalibrierung, die als Sicherheitsaufgabe zu verstehen ist. Zentrale Voraussetzungen sind kontinuierliche Rückkopplung, lokale Validierung sowie eine explizite Kommunikation von Anwendungsbereich und typischen Fehlermodi. Kalibrierung ist dabei kein einmaliger Schulungseffekt, sondern ein Prozess, der dauerhaft durch geeignete Strukturen gestützt werden muss. Wirksame Maßnahmen greifen auf mehreren Ebenen: Auf individueller Ebene sollte Nutzungsschulung als Kalibrierungstraining angelegt sein, das mit repräsentativen Grenzfällen arbeitet, typische Fehlermodi sichtbar macht und Diskrepanzen zwischen KI-System- und ärztlichem Fachurteil systematisch reflektiert. Auf der Ebene des Arbeitsablaufs unterstützen Audit-Schleifen die Stabilisierung von Überwachungsstrategien, etwa durch regelmäßige Auswertung von Abweichungen getrennt nach KI-positiven und KI-negativen Fällen. Leistungs-Dashboards ermöglichen zusätzlich die Beobachtung von driftenden Abweichungen, Prävalenzverschiebungen und Untergruppenleistungen. Auf Governance-Ebene beruht Kalibrierung auf operativer Transparenz: Klare Angaben zu Verwendungszweck, Versionierung, bekannten Einschränkungen und Update-Management können unangemessene Vertrauensübertragungen begrenzen und eine kontextangemessene Einordnung von Ergebnissen erleichtern (Lee & See, 2004; Parasuraman & Riley, 1997). Entscheidend ist dabei ein Perspektivwechsel: Nicht die Steigerung von Vertrauen in KI-basierte Automatisierung per se ist relevant, sondern die Reduktion sicherheitsrelevanter Fehlhandlungen und Unterlassungen. Da sich Vertrauensdynamiken zwischen professioneller Praxis und Patientenseite unterscheiden, wird Letztere im folgenden Abschnitt gesondert betrachtet.

5. Patientenperspektiven zum Einsatz von KI in der Radiologie

5.1 Erwartungen und wahrgenommene Risiken

Die patientenseitige Einordnung der KI-Nutzung in der Radiologie erfolgt auf der Grundlage unterschiedlicher mentaler Modelle, die als kognitive Rahmen fungieren und Erwartungen an Rolle, Nutzen und Risiko der Technologie strukturieren. Je nachdem, ob KI als autonomer Diagnostiker, Ko-Detektor oder als untergeordnetes technisches Hilfsmittel verstanden wird, verschieben sich Annahmen über Kompetenz, Kontrollbedarf und Verantwortlichkeit. Mit den mentalen Modellen verändern sich zugleich Nutzen- und Risikozuschreibungen: Erwartungen an höhere Genauigkeit, beschleunigte Abläufe und gleichbleibende Qualität werden mit Befürchtungen wie Fehleranfälligkeit, Entmenschlichung der Versorgung, Verlust ärztlicher Aufmerksamkeit, Datenschutzproblemen, sekundärer Datennutzung oder unklaren Verantwortlichkeiten verknüpft. Akzeptanz entsteht dabei nicht aus einer allgemeinen Technikbefürwortung, sondern unter spezifischen Bedingungen. Empirische Arbeiten zeigen, dass Zustimmung maßgeblich davon abhängt, ob ärztliche Aufsicht und Entscheidungsverantwortung im Behandlungsprozess erkennbar bleiben (Promberger & Baron, 2006; Nadarzynski et al., 2019; Karger, 2026). Der patientenseitige Informationsbedarf in der Radiologie richtet sich entsprechend auf konkrete, handlungsnahe Fragen: Welche Rolle KI im diagnostischen Prozess einnimmt, ob radiologische Bilder unabhängig geprüft werden, wie mit Meinungsverschiedenheiten zwischen Mensch und Maschine umgegangen wird und welche Konsequenzen Fehler haben. Demgegenüber kommunizieren Organisationen den KI-Einsatz häufig in allgemeinen Formeln, die Transparenz signalisieren, ohne Informationsautonomie tatsächlich zu ermöglichen (z. B. »es werden fortschrittliche Tools verwendet«). Mentale Modelle werden so weniger durch formale Offenlegung opaker KI-Systeme als durch die Art geprägt, in der ärztliche Verantwortlichkeit, Aufsicht und Entscheidungsprozesse im Umgang mit solchen Systemen nachvollziehbar gemacht werden.

Darüber hinaus ist die patientenseitige Bewertung von KI-Algorithmen in medizinischen Settings stark kontextabhängig und folgt

den Anforderungen der jeweiligen Entscheidungssituation. Es existiert keine einheitliche Akzeptanzlage, vielmehr variiert die Einschätzung algorithmischer Entscheidungen systematisch mit dem Aufgabencharakter. Dort, wo Genauigkeit, Konsistenz und Objektivität im Vordergrund stehen, werden KI-basierte Verfahren häufig positiv bewertet. In Situationen hingegen, in denen individuelles Urteilsvermögen, Empathie oder ein situationssensitives Eingehen als zentral gelten, stoßen algorithmische Entscheidungen auf Zurückhaltung (Dietvorst et al., 2015; Logg et al., 2019; Longoni et al., 2019). Diese Zurückhaltung ist dabei eben nicht als pauschale Technikablehnung zu verstehen, sondern als Ausdruck unterschiedlicher normativer Erwartungen daran, was in bestimmten medizinischen Kontexten als legitime Entscheidungsgrundlage gilt. Patientenseitig ist Akzeptanz damit funktional gebunden und nicht global zu erzielen. Unter diesen Bedingungen übernimmt Transparenz eine regulierende Rolle. Sie soll Orientierung bieten und informierte Zustimmung ermöglichen, ohne Akzeptanz erzwingen zu wollen oder Vorbehalte zu delegitimieren. Transparenz darf daher nicht als persuasive, ärztliche Kommunikationsstrategie eingesetzt werden, die patientenseitige Zweifel als irrational erscheinen lässt oder implizit nahelegt, Zurückhaltung oder Ablehnung gegenüber dem Einsatz von KI-Tools sei unerwünscht. Nur so kann Transparenz dazu beitragen, kontextangemessene Bewertungen von KI-Einsatz zu unterstützen, anstatt soziale oder moralische Druckeffekte zu erzeugen.

Über individuelle Risiko-Nutzen-Abwägungen hinaus wird der medizinische Einsatz von KI auch nach Fairnesskriterien beurteilt. Besonders in Bevölkerungs-Screenings und Triage-Situationen, in denen KI die Priorisierung, Reihenfolge oder Zugang zur medizinischen Versorgung beeinflusst, rücken Verteilungswirkungen in den Vordergrund. Fairness wird hier zu einer eigenständigen Bewertungsdimension neben Genauigkeit und Effizienz. Das öffentliche Bewusstsein für algorithmische Verzerrungen, verstärkt durch prominente Befunde zur Vorhersage von Gesundheitsrisiken, hat Fairness zu einem zentralen Vertrauensfaktor gemacht, auch in Anwendungsfeldern wie der (radiologischen) Bildinterpretation, in denen systematische Urteilsverzerrungen zunächst weniger offensichtlich erscheinen (Obermeyer et al., 2019; Rajkomar et al., 2018). Wahrgenommene Gerechtigkeit hängt dabei entscheidend davon ab, ob ein System als angemessen für Personen »wie mich« gilt und ob

Fehler als gerecht verteilt erscheinen, statt systematisch bestimmte Gruppen stärker zu betreffen. Hinzu kommt die Bedeutung der Anfechtbarkeit KI-basierter Ergebnisse: Die Möglichkeit, KI-gestützte Entscheidungen ohne substantielle Hürden zu hinterfragen und Eskalationswege zu nutzen, beeinflusst maßgeblich, ob der KI-Einsatz als fair wahrgenommen wird, unabhängig von der formalen Leistungsfähigkeit des KI-Systems. Studien zur algorithmischen Entscheidungsfindung zeigen, dass Wahrnehmungen von Gerechtigkeit stark durch Erklärungstypen, Beteiligungsformate, Rechtsmittel und Ergebniskommunikation geprägt werden (Binns et al., 2018). Diese Faktoren sind direkt anschlussfähig an radiologische Befundungs- und Eskalationsprozesse und zeigen auf, wie die ärztliche KI-Nutzung kommuniziert werden sollte, damit auf Patientenseite angemessenes Vertrauen entsteht.

5.2 KI-Nutzung und Vertrauen in asymmetrischen Arzt-Patient-Beziehungen

Vertrauen in den Einsatz von KI im medizinischen Kontext entsteht unter Bedingungen typischerweise asymmetrischer Arzt-Patient-Beziehungen und wird von diesem Vertrauensverhältnis vermittelt oder moduliert. In den meisten radiologischen Untersuchungsabläufen besteht patientenseitig kein direkter Kontakt mit der KI; ihre Wirkung wird vielmehr über institutionelle Strukturen und professionelles Handeln vermittelt. Vertrauen bezieht sich damit auf die Weise, in der der KI-Einsatz in diagnostische Prozesse eingebettet und nach außen sichtbar gemacht wird. Die Effekte des Systems werden indirekt erfahrbar, etwa über Wartezeiten, die Formulierung von Befunden, nachfolgende klinische Entscheidungen und die kommunikative Haltung der behandelnden Seite. Diese beobachtbaren Konsequenzen fungieren als Anhaltspunkte, anhand derer Vertrauen aufgebaut oder revidiert wird. Aufgrund der ungleichen Verteilung von Wissen und Entscheidungsmacht ist Vertrauen dabei besonders sensibel gegenüber Signalen der Verantwortungsübernahme. Diese Vermittlung wirkt in zwei Richtungen: Vertrauen kann gestärkt werden, wenn KI als überwachte Unterstützung erscheint, die diagnostische Konsistenz erhöht und in das ärztliche Urteil sichtbar integriert wird; es kann jedoch untergraben werden, wenn KI als Ersatz für

ärztliche Einordnung, Entscheidung oder Verantwortung kommuniziert wird oder zumindest so erscheint (Karger, 2026).

Eine spezifische Herausforderung der Radiologie liegt in der begrenzten ärztlichen Sichtbarkeit im klinischen Alltag. Radiologische Expertise ist häufig räumlich, zeitlich und kommunikativ vom unmittelbaren Behandlungskontakt getrennt, sodass die Zuschreibung von Verantwortung nicht selbstverständlich entsteht. Diese strukturelle Distanz kann durch den Einsatz von KI verstärkt werden, denn durch KI tritt eine zusätzliche Vermittlungsebene hinzu, die Zuständigkeiten, diagnostische Entscheidungspfade und Verantwortungszuschreibung weiter verschleiern kann. Erscheint ein Befundbericht als technisches Artefakt einer undurchsichtigen Prozesskette (»Black Box«), etwa als Ergebnis von »KI-System plus Institution«, kann Vertrauen in die Behandlungsbeziehung sinken, selbst wenn sich die diagnostische Leistung objektiv verbessert. Vertrauen reagiert hier weniger auf Leistungskennzahlen als auf Zuschreibbarkeit von Verantwortung und Ansprechbarkeit (Chiou & Lee, 2023). Demgegenüber können strukturierte und dem Zweck angemessene Erklärungsangebote – etwa durch die überweisende Ärzteschaft, sorgfältig kuratierte Patientenportale oder klar zugängliche Rückfragekanäle – die patientenseitige Wahrnehmung ärztlicher Präsenz und Verantwortung in der KI-gestützten Radiologie stabilisieren. Wie bereits erwähnt, wirkt Transparenz in solchen Informationsangeboten kontextabhängig: Hinweise auf Unsicherheit können beruhigen, wenn sie Aufsicht und Sorgfalt signalisieren, oder aber verunsichern, wenn sie primär die Fehlbarkeit betonen. Ziel ist daher eine zweckmäßig kontextualisierte Kommunikation der Unsicherheiten im KI-gestützten diagnostischen Prozess, die klar macht, was bekannt ist, was offen bleibt und welche Sicherungen greifen.

Die Umsetzung dieser Einsichten ist möglich, ohne die radiologische Arbeit in permanente Einzelkonsultationen zu überführen. Vertrauensbildung im Kontext der radiologischen KI-Nutzung erfordert keine kompensatorische Verdichtung persönlicher Arzt-Patient-Interaktionen, sondern eine strukturierte Kalibrierung von patientenseitigen Erwartungen gegenüber hybriden Entscheidungsprozessen. Wie bereits bezüglich der Ärzteschaft erwähnt (s. Abschnitt 4), zielt Vertrauenskalibrierung darauf, Vertrauen in KI-basierte Technologien weder zu maximieren noch zu minimieren, sondern es an deren tatsächliche Zuverlässigkeit und Grenzen anzupassen. Drei Leitlini-

en sind dabei zentral: (1) KI sollte, soweit zutreffend, konsequent als ärztlich überwachte Unterstützung kommuniziert werden, die Priorisierung und Befundung erleichtert, während die diagnostische Verantwortung eindeutig außerhalb des KI-Systems verankert bleibt. (2) Systemseitige Unsicherheiten und Grenzen sollten möglichst *konkret* dargestellt werden, wie z B. der validierte Anwendungsbereich, typische Fehlerquellen sowie das Vorgehen bei mehrdeutigen Ergebnissen. (3) Verantwortlichkeiten, Anfechtungs- und Regressmöglichkeiten bezüglich der KI-basierten Ergebnisse sollten *klar* benannt werden, um sichtbar zu machen, wie mit abweichenden Interpretationen umgegangen und aus etwaigen Fehlern gelernt wird. Diese Prinzipien können helfen, patientenseitige Erwartungen mit hybriden Entscheidungsrealitäten in der Radiologie in Einklang zu bringen und das Risiko unkritischer Überhöhung ebenso wie pauschaler Ablehnung zu senken. Da Vertrauen und Akzeptanz auf Patientenseite die Einführung und Verstetigung neuer Technologien wesentlich mitbestimmen, werden sie im nächsten Abschnitt gemeinsam mit der ärztlichen Perspektive adressiert.

6. Technologieakzeptanz und organisatorische Bedingungen nachhaltiger KI-Nutzung in der Radiologie

6.1 Einstellungen und Akzeptanz gegenüber KI-basierten Technologien

Der großflächige und nachhaltige Einsatz von KI in der Radiologie hängt von dessen Akzeptanz in den beiden zentralen Akteursgruppen, also der Ärzte- und Patientenschaft, ab (Caspers et al., 2025). Aus ärztlicher Perspektive steht die Frage im Vordergrund, ob ein gegebenes KI-Tool als zweckmäßiges Automatisierungselement in bestehende Arbeitsabläufe passt. Forschung zur Technologieakzeptanz zeigt, dass vor allem Nutzenwahrnehmung, Bedienbarkeit und unterstützende/hemmende Rahmenbedingungen die Adoption neuer Technologien prägen (Davis, 1989; Venkatesh et al., 2003). Aus Patientensicht muss KI-gestützte Versorgung vor allem als legitim, angemessen, verlässlich und fair gelten, wie im vorangegangenen Kapitel ausgeführt. Werden diese unterschiedlichen Determinanten

der Akzeptanz nicht beachtet, laufen auch objektive, leistungsfähige und potenziell hilfreiche KI-Systeme Gefahr, im klinischen Alltag nicht nachhaltig Fuß zu fassen. Einstellungen wirken als Schwelle für eine Skalierung: Ohne Zuversicht und Akzeptanz auf ärztlicher Ebene bleibt eine Nutzung begrenzt; ohne Akzeptanz auf Patientenebene entstehen Legitimations- und Reputationsprobleme, die eine Ausweitung verhindern (Greenhalgh et al., 2004; Rogers, 2003).

Die ärztliche Haltung gegenüber KI-basierten Technologien in der Radiologie lässt sich als Ergebnis eines Spannungsverhältnisses zwischen wahrgenommenem klinischem Nutzen und wahrgenommenen beruflichen Kosten rekonstruieren. Akzeptanz entsteht dabei nicht aus der Leistungsfähigkeit der Technologie allein, sondern aus der Relation zwischen erwarteten entlastenden Effekten der Automatisierung und erwartetem zusätzlichen Aufwand im konkreten Arbeitsprozess. Ein zentraler Faktor ist die Nützlichkeit in der lokalen Aufgabenumgebung. KI-Akzeptanz wird gefördert, wenn KI-Systeme Fehler reduzieren, Priorisierung verbessern oder Auswertungszeiten verkürzen, ohne nachgelagerte Prozesse nennenswert zu belasten. Ebenso entscheidend ist die Passung zum Arbeitsablauf. Systeme können auch bei hoher Genauigkeit abgelehnt werden, wenn sie den Arbeitsfluss stören, umständlich zu bedienen sind oder eine Vielzahl manuell zu prüfender Ausnahmen erzeugen. In solchen Fällen ist Mehrbelastung unmittelbar erfahrbar, während der Nutzen unsicher oder zeitlich verzögert bleibt. Hinzu kommt die Frage professioneller Autonomie: Wird KI als Ersatz für Expertise positioniert, entsteht Widerstand; wird sie als Erweiterung bestehender Praxis kommuniziert, ist sie mit beruflichen Normen vereinbar (s. Abschnitt 2). Eine weitere Einflussgröße betrifft die ärztliche Verantwortlichkeit und Haftung: Gelten Abweichungen von KI-Empfehlungen als riskant, begünstigt dies eine Ablehnung solcher Technologien. Empirische Befunde zu Akzeptanzfaktoren in der Ärzteschaft zeigen entsprechend eine Mischung aus Leistungsoptimismus und Sorge um Verantwortung, Qualifikationsverlust und Arbeitsbelastung, was die nachhaltige Adoption von KI-Technologien in der Radiologie als soziotechnisches Integrationsproblem ausweist (Waymel et al., 2019; Huisman et al., 2021). Die aus diesen und anderen Überzeugungen resultierende Einstellung zur KI-Nutzung bestimmt am Ende nicht nur, ob KI überhaupt genutzt wird, sondern auch, ob

sie zur überprüfenden Unterstützung, zu diagnostischen Abkürzungen oder zur rein formalen Absicherung eingesetzt wird.

Die Akzeptanz von radiologischen KI-Anwendungen auf Patientenseite bestimmt sich aus einer Abwägung zwischen wahrgenommenem Nutzen und wahrgenommenen Risiken. Vorteile wie Geschwindigkeit, Konsistenz und zugeschriebene Objektivität stehen dabei Befürchtungen gegenüber, die sich weniger auf technische Leistungsparameter als auf normative und relationale Aspekte beziehen, etwa Fehlerfolgen, Entpersonalisierung, unklare Verantwortlichkeiten und Fairness. Diese Bewertungen und Haltungen haben nicht nur individuelle, sondern auch gesellschaftliche Konsequenzen, denn sie bestimmen die Zustimmung zu, die Zufriedenheit mit und die öffentliche Legitimität von großangelegten medizinischen KI-Programmen, insbesondere in Szenarien, in denen eine KI den Zugang zu oder die Priorisierung von ärztlicher Betreuung steuert, etwa in Screening- und Triage-Situationen. Wie bereits erwähnt, legt die Forschung zur algorithmischen Entscheidungsfindung eine konditionale Akzeptanzdynamik nahe: Algorithmische Urteilsfindung erhält Zustimmung, wenn sie als genauer wahrgenommen wird, löst jedoch Ablehnung aus, sobald sie als Bedrohung von Handlungsfähigkeit, menschlicher Würde oder individueller Versorgung interpretiert wird oder wenn KI-systemseitig Fehler salient werden (Dietvorst et al., 2015; Logg et al., 2019; Longoni et al., 2019; Karger, 2026). Die Akzeptanz von KI in der Radiologie auf Patientenseite ist somit weder stabil noch selbstverständlich, sondern kontextabhängig und revidierbar. Entscheidend für die Akzeptanz wird daher das Vertrauen in die radiologische Einrichtung, in die die Technologie eingebettet ist, sowie die Art und Weise, in der Aufsicht, Unsicherheit und Anfechtungswege von ärztlicher Seite aus kommuniziert werden. Neben gesellschaftlicher Akzeptanz spielen allerdings auch förderliche oder hemmende organisational-strukturelle Bedingungen eine Rolle für die nachhaltige Adoption von KI-Technologien in der Radiologie; psychologische Aspekte dieser Bedingungen werden im Folgenden kurz umrissen.

6.2 Organisationale Bedingungen einer nachhaltigen KI-Adoption in der Radiologie

Die Einführung von KI in der Radiologie ist als sozio-organisatorisches Projekt zu analysieren und nicht als bloße Beschaffungsentscheidung. Ob ein KI-basiertes System zwecks Automatisierung in der radiologischen Diagnostik sicher und wirksam eingesetzt wird, hängt – wie in den vorherigen Abschnitten erörtert – nicht allein von seiner technischen Leistungsfähigkeit ab, sondern von der Weise, in der es in bestehende und ggf. anzupassende Strukturen eingebettet wird. Die Implementierungsforschung zeigt, dass die erfolgreiche Einführung und Adoption von Technologien durch ein Bündel von Faktoren bestimmt werden. Dazu zählen Merkmale der Intervention selbst, etwa ihre Komplexität, ebenso wie Eigenschaften des internen professionellen Umfelds, darunter Kultur, Ressourcen und Routinen. Hinzu kommen externe Einflüsse wie regulatorischer oder reputationsbezogener Druck sowie die organisationale Fähigkeit zur fortlaufenden Bewertung und Anpassung (Damschroder et al., 2009; Greenhalgh et al., 2004). Aus soziotechnischer Perspektive entstehen Risiken nicht aus der Technologie isoliert, sondern aus dem Zusammenspiel von Technik, Aufgabenstruktur und organisationalem Kontext. Daraus folgt, dass Gestaltung und Überwachung radiologischer KI-Anwendungen nicht als einmalige oder bei Einführung fixierte Maßnahmen angelegt sein dürfen. Sicherheit und Wirksamkeit von KI-Systemen im diagnostischen Prozess ergeben sich aus kontinuierlicher Abstimmung, laufender Rückmeldung und wiederholter Anpassung der Arbeitsabläufe (Sittig & Singh, 2010; Carayon et al., 2014). Für einen nachhaltigen und skalierbaren KI-Einsatz lassen sich vor diesem Hintergrund mehrere relevante organisationale Facetten identifizieren, die im Folgenden kurz betrachtet werden.

Vertrauen in KI lässt sich nur dann stabilisieren, wenn Einsatz, Leistung und Abweichungen fortlaufend beobachtbar und überprüfbar gemacht werden. Dies erfordert Praktiken, die wiederholt vor Ort nachprüfen, ob ein System im vorgesehenen Anwendungsbereich zuverlässig arbeitet, die im laufenden Betrieb regelmäßig nachsteuern, die Fehler- und Beinahefehler-Ereignisse systematisch erfassen und die KI-Systemleistung kontinuierlich beobachten. Solche Verfahren bestimmen, ob die breite KI-Nutzung im radiologischen Alltag als nachhaltig tragfähig erscheint und ob KI-gestütz-

te Versorgung als legitim bewertet wird. Wird nachvollziehbar gemacht, wofür ein System eingesetzt wird, wann es aktualisiert wurde und wie mit Abweichungen umgegangen wird, kann Vertrauen an überprüfbare Routinen gebunden werden. Bleiben diese Prozesse undurchsichtig, reagieren Organisationen häufig mit Rückzug aus der Nutzung oder mit defensiven Anpassungen, die letztlich aber einen Fehlgebrauch begünstigen. Damit Kontrolle lernfähig wird, müssen Abweichungen und Beinahefehler-Ereignisse ohne Sorge vor Schuldzuweisung thematisiert werden können. Psychologische Sicherheit ist daher eine operative Voraussetzung belastbarer Rückmeldung und Verbesserung (Edmondson, 1999). Ein »Just Culture«-Ansatz behandelt KI-bezogene Abweichungen als systemische Lernanlässe und verhindert, dass strukturelle Fehler an der operativen Ebene personalisiert abgefangen werden (Reason, 1997).

Die Nutzung von KI in der radiologischen Routine unterliegt auch wirtschaftlichen Gegebenheiten, die sich in Anreizsystemen und Leistungskennzahlen niederschlagen können. Werden dabei Durchsatzziele und Produktivitätsmetriken in den Vordergrund gestellt, kann dies unbeabsichtigt zu reduzierter Überprüfung und erhöhter Abhängigkeit von KI-gestützten Ergebnissen führen, insbesondere wenn KI primär als Instrument zur Beschleunigung radiologischer Arbeitsabläufe positioniert ist. Wird Erfolg überwiegend an Durchlaufzeiten gemessen, verengt sich der Bewertungsraum. Fehlertypen, Fehlerfolgen und nachgelagerte Effekte geraten aus dem Blick, obwohl sie für die Sicherheit und Versorgungsqualität zentral sind. Allgemeiner gilt, dass Kennzahlen, sobald sie zu Zielgrößen werden, Verhalten verzerren und Belastungen verlagern können. Effizienzgewinne an einer Stelle gehen dann mit zusätzlichen Kosten an anderer Stelle einher. So erhöhen falsch-positive Befunde die Zahl von Folgeuntersuchungen und die Belastung auf Versorgungsebene, auch wenn interne Produktivitätskennzahlen steigen (Muller, 2018). Einführung und Skalierung von KI erfordern daher Bewertungssysteme, die Effizienz nicht isoliert und auf wenige Leistungsparameter verengt erfassen, sondern Sicherheit, Arbeitsbelastung und nachgelagerte Nutzung systematisch mitberücksichtigen.

Nachhaltiger Nutzen von KI setzt organisationale Lernmechanismen voraus, die über die einmalige Einführung hinausgehen. Abweichungen und Leistungsänderungen des KI-Systems müssen systematisch in eine veränderte Praxis übersetzt werden, etwa durch

angepasste Arbeitsroutinen, aktualisierte Richtlinien oder gezielte Auslöser für die ärztliche Weiterbildung und Revalidierung des Systems. Studien zu Audits und Feedback zeigen, dass Rückmeldungen Verhalten nur dann verändern, wenn sie zeitnah erfolgen, spezifisch sind und klare Handlungsimplikationen enthalten (Ivers et al., 2012). Diese Bedingungen markieren zugleich Anforderungen an die Interpretierbarkeit der KI-bezogenen Überwachung, denn Überwachung wird erst wirksam, wenn sie nicht nur Abweichungen fortlaufend misst, sondern relevante Abweichungen so aufbereitet, dass Anpassung möglich wird. Organisational verankerte Lernschleifen müssen zudem mit Heterogenität umgehen können, denn Leistungs- und Vertrauensmuster unterscheiden sich zwischen Standorten, eingesetzten Technologien und Prävalenzkontexten, sodass einheitliche Vorgaben hier zu kurz greifen. Die organisationale Steuerung muss daher lokale Anpassung ermöglichen, ohne eine unkontrollierte und potenziell invalide Ausweitung von Einsatzbereichen zuzulassen.

Zusammenfassend zeigt unsere Analyse, dass die psychologischen Wirkungen KI-basierter Automatisierung nicht an der Technologie selbst entschieden werden, sondern an den Bedingungen ihres Einsatzes. Dieselben Systeme können die diagnostische Sicherheit erhöhen oder neue Fehlerquellen erzeugen, abhängig davon, wie sie in Arbeitsabläufe, Verantwortungsstrukturen und Rückmeldemechanismen eingebettet sind. Automatisierungsverzerrung, Fehlkalibrierung von Vertrauen, Kompetenzverschiebungen oder moralischer Distress entstehen nicht zufällig und auch nicht primär aus individuellen Einstellungen, sondern aus stabilen Kopplungen zwischen Systemlogik, Aufgabenstruktur und organisationalen Erwartungen. Damit verschiebt sich der Fokus von der Frage, ob KI leistungsfähig ist, zu der Frage, unter welchen Bedingungen ihre Leistungsfähigkeit wirksam und kontrollierbar wird. Nachhaltiger Nutzen setzt voraus, dass Kontrolle, Lernen und Verantwortung nicht an individuelle Wachsamkeit und Sorgfalt delegiert werden, sondern strukturell abgesichert sind. Transparente Einsatzgrenzen, lernfähige Rückkopplung, ausbalancierte Anreizsysteme und psychologisch sichere Umgangsformen mit Abweichungen bestimmen, ob KI als unterstützende Ressource wirkt oder Abhängigkeit, Passivierung und defensive Routinen begünstigt. Erst auf dieser Grundlage wird eine Implementierung möglich, die über punktuelle Effizienzgewinne hinausreicht

und auf langfristige Robustheit, Sicherheit und Skalierbarkeit ausgerichtet ist.

Literaturverzeichnis

Amann, J., Blasimme, A., Vayena, E., & Frey, D. (2020). Explainability for artificial intelligence in health care: A multidisciplinary perspective. *BMC Medical Informatics and Decision Making*, *20*(1), 310. https://doi.org/10.1186/s12911-020-01332-6

Bainbridge, L. (1983). Ironies of automation. *Automatica*, *19*(6), 775–779. https://doi.org/10.1016/0005-1098(83)90046-8

Carayon, P., Wetterneck, T. B., Rivera-Rodriguez, A. J., Hundt, A. S., Hoonakker, P., Holden, R., & Gurses, A. P. (2014). Human factors systems approach to healthcare quality and patient safety. *Applied Ergonomics*, *45*(1), 14–25. https://doi.org/10.1016/j.apergo.2013.04.023

Caspers, J., Karger, C., Langner, R., Weißenfels, S., Günther, J., Spranger, T.M., Wagner, R., Lanzerath, D., Eickhoff, S.B., & Heinrichs, B. (2025). Künstliche Intelligenz in der Radiologie – Von der experimentellen Phase zur produktiven Anwendung. *Deutsches Ärzteblatt International, 122*(Sonderausgabe KI), 24–27.

Chiou, E. K., & Lee, J. D. (2023). Trusting automation: Designing for responsivity and resilience. *Human Factors*, *65*(1), 137–165. https://doi.org/10.1177/00187208211009995

Cook, R. I., & Woods, D. D. (1996). Adapting to new technology in the operating room. *Human Factors*, *38*(4), 593–613. https://doi.org/10.1518/001872096778827224

Cosmides, L., & Tooby, J. (2000). Evolutionary psychology and the emotions. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (2nd ed., pp. 91-115). New York, NY: Guilford Press

Croskerry, P., Singhal, G., & Mamede, S. (2013). Cognitive debiasing 1: Origins of bias and theory of debiasing. *BMJ Quality & Safety*, *22*(Suppl 2), ii58–ii64. https://doi.org/10.1136/bmjqs-2012-001712

Cruess, R. L., Cruess, S. R., Boudreau, J. D., Snell, L., & Steinert, Y. (2014). Reframing medical education to support professional identity formation. *Academic Medicine*, *89*(11), 1446–1451. https://doi.org/10.1097/ACM.0000000000000427

Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York, NY: Putnam

Damschroder, L. J., Aron, D. C., Keith, R. E., Kirsh, S. R., Alexander, J. A., & Lowery, J. C. (2009). Fostering implementation of health services research findings into practice: A consolidated framework for advancing implementation science. *Implementation Science*, *4*(1), 50. https://doi.org/10.1186/1748-5908-4-50

Dang, Q., & Li, G. (2025). Unveiling trust in AI: The interplay of antecedents, consequences, and cultural dynamics. *AI & Society*. https://doi.org/10.1007/s00146-025-02477-6

Darley, J. M., & Latané, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology*, *8*(4 Pt. 1), 377–383. https://doi.org/10.1037/h0025589

Dratsch, T., Chen, X., Rezazade Mehrizi, M., Kloeckner, R., Mahringer-Kunz, A., Pusken, M., ... Pinto Dos Santos, D. (2023). Automation bias in mammography: The impact of artificial intelligence BI-RADS suggestions on reader performance. *Radiology*, *307*(4), e222176. https://doi.org/10.1148/radiol.222176

Doshi-Velez, F., & Kim, B. (2017). *Towards a rigorous science of interpretable machine learning*. arXiv preprint arXiv:1702.08608. https://arxiv.org/abs/1702.08608

Drew, T., Vo, M. L.-H., & Wolfe, J. M. (2013). The invisible gorilla strikes again: Sustained inattentional blindness in expert observers. *Psychological Science*, *24*(9), 1848–1853. https://doi.org/10.1177/0956797613479386

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Human-Computer Studies*, *58*(6), 697–718. https://doi.org/10.1016/S1071-5819(03)00038-7

Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, *44*(2), 350–383. https://doi.org/10.2307/2666999

Elish, M. C. (2019). Moral crumple zones: Cautionary tales in human–robot interaction. *Engaging Science, Technology, and Society*, *5*, 40–60. https://doi.org/10.17351/ests2019.260

Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors*, *37*(2), 381–394. https://doi.org/10.1518/001872095779064555

Ghassemi, M., Oakden-Rayner, L., & Beam, A. L. (2021). The false hope of current approaches to explainable artificial intelligence in health care. *The Lancet Digital Health*, *3*(11), e745–e750. https://doi.org/10.1016/S2589-7500(21)00208-9

Goddard, K., Roudsari, A., & Wyatt, J. C. (2012). Automation bias: A systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association*, *19*(1), 121–127. https://doi.org/10.1136/amiajnl-2011-000089

Gong, B., Nugent, J. P., Guest, W., Parker, W., Chang, P. J., Khosa, F., & Nicolaou, S. (2019). Influence of artificial intelligence on Canadian medical students' preference for radiology specialty: A national survey study. *Academic Radiology*, *26*(4), 566–577. https://doi.org/10.1016/j.acra.2018.10.007

Greenhalgh, T., Robert, G., Macfarlane, F., Bate, P., & Kyriakidou, O. (2004). Diffusion of innovations in service organizations: Systematic review and recommendations. *Milbank Quarterly*, *82*(4), 581–629. https://doi.org/10.1111/j.0887-378X.2004.00325.x

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, *57*(3), 407–434. https://doi.org/10.1177/0018720814547570

Huisman, M., Ranschaert, E., Parker, W., Mastrodicasa, D., Koci, M., Pinto de Santos, D., ... Willemink, M. J. (2021). An international survey on AI in radiology: Fear of replacement, knowledge, and attitude. *European Radiology*, *31*(9), 7058–7066. https://doi.org/10.1007/s00330-021-07781-5

Ivers, N., Jamtvedt, G., Flottorp, S., Young, J. M., Odgaard-Jensen, J., French, S. D., ... Oxman, A. D. (2012). Audit and feedback: Effects on professional practice and healthcare outcomes. *Cochrane Database of Systematic Reviews*, *2012*(6), CD000259. https://doi.org/10.1002/14651858.CD000259.pub3

Jameton, A. (1984). *Nursing practice: The ethical issues*. Englewood Cliffs, NJ: Prentice Hall

Karger, C. R. (2026). Patients' perspectives on the implementation of artificial intelligence in radiological diagnostics: A focus group study. *Journal of Medical Internet Research, 30/03/2026 (im Druck)*, 89178. https://doi.org/10.2196/89178

Kherbache, A., Mertens, E., & Denier, Y. (2022). Moral distress in medicine: An ethical analysis. *Journal of Health Psychology*, *27*(8), 1971–1990. https://doi.org/10.1177/13591053211014586

Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., & Koch, I. (2010). Control and interference in task switching—A review. *Psychological Bulletin*, *136*(5), 849–874. https://doi.org/10.1037/a0019842

Lebovitz, S., Lifshitz-Assaf, H., & Levina, N. (2022). To engage or not to engage with AI for critical judgments: How professionals deal with opacity when using AI for medical diagnosis. *Organization Science*, *33*(1), 126–148. https://doi.org/10.1287/orsc.2021.1549

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392

Lerner, J. S., Li, Y., Valdesolo, P., & Kassam, K. S. (2015). Emotion and decision making. *Annual Review of Psychology*, *66*, 799–823. https://doi.org/10.1146/annurev-psych-010213-115043

Lipton, Z. C. (2018). The mythos of model interpretability. *Communications of the ACM*, *61*(10), 36–43. https://doi.org/10.1145/3233231

Loewenstein, G. F., Weber, E. U., Hsee, C. K., & Welch, N. (2001). Risk as feelings. *Psychological Bulletin, 127*(2), 267–286. https://doi.org/10.1037/0033-2909.127.2.267

Lombi, L., & Rossero, E. (2024). How artificial intelligence is reshaping the autonomy and boundary work of radiologists: A qualitative study. *Sociology of Health & Illness*, *46*(2), 200–218. https://doi.org/10.1111/1467-9566.13702

London, A. J. (2019). Artificial intelligence and black-box medical decisions: Accuracy versus explainability. *Hastings Center Report*, *49*(1), 15–21. https://doi.org/10.1002/hast.973

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, *267*, 1–38. https://doi.org/10.1016/j.artint.2018.07.007

Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors*, *52*(3), 381–410. https://doi.org/10.1177/0018720810376055

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, *39*(2), 230–253. https://doi.org/10.1518/001872097778543886

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, *30*(3), 286–297. https://doi.org/10.1109/3468.844354

Perez, F., Conway, N., Peterson, J., & Roques, O. (2024). Me, my work and AI: How radiologists craft their work and identity. *Journal of Vocational Behavior*, *155*, 104042. https://doi.org/10.1016/j.jvb.2024.104042

Reason, J. T. (1990). *Human error*. Cambridge, UK: Cambridge University Press.

Reason, J. T. (1998). Achieving a safe culture: Theory and practice. *Work & Stress*, *12*(3), 293–306. https://doi.org/10.1080/02678379808256868

Rony, M. K. K., Parvin, M. R., Wahiduzzaman, M., Debnath, M., Bala, S. D., & Kayesh, I. (2024). »I wonder if my years of training and expertise will be devalued by machines«: Concerns about the replacement of medical professionals by artificial intelligence. *SAGE Open Nursing*, *10*, 23779608241245220. https://doi.org/10.1177/23779608241245220

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*, 206–215. https://doi.org/10.1038/s42256-019-0048-x

Schwarz, N., & Clore, G. L. (2007). Feelings and phenomenal experiences. In A. W. Kruglanski & E. T. Higgins (Eds.), *Social psychology: Handbook of basic principles* (2nd ed., pp. 385–407). New York, NY: The Guilford Press

Shonhe, L., & Min, Q. (2025). Mitigating AI-induced professional identity threat and fostering adoption in the workplace. *AI & Society*, *40*(5), 4079–4092. https://doi.org/10.1007/s00146-024-02170-0

Sittig, D. F., & Singh, H. (2010). A new sociotechnical model for studying health information technology in complex adaptive healthcare systems. *Quality and Safety in Health Care*, *19*(Suppl 3), i68–i74. https://doi.org/10.1136/qshc.2010.042085

Steinborn, M. B., Langner, R., & Huestegge, L. (2017). Mobilizing cognition for speeded action: Try-harder instructions promote motivated readiness in the constant-foreperiod paradigm. *Psychological Research*, *81*, 1135–1151. https://doi.org/10.1007/s00426-016-0810-1

Wiens, J., Saria, S., Sendak, M., Ghassemi, M., Liu, V. X., Doshi-Velez, F., ... Goldenberg, A. (2019). Do no harm: A roadmap for responsible machine learning for health care. *Nature Medicine*, *25*(9), 1337–1340. https://doi.org/10.1038/s41591-019-0548-6

Wolfe, J. M., Horowitz, T. S., & Kenner, N. M. (2005). Cognitive psychology: Rare items often missed in visual searches. *Nature*, *435*(7041), 439–440. https://doi.org/10.1038/435439a