Masterstudium

Interprofessionelle Gesundheitswissenschaften

Interprofessionelles Capstone Projekt 2

Künstliche Intelligenz im Gesundheitswesen

Nützliche KI-Anwendungen für klinische Fragen

Wintersemester 2025/2026

Vortragende: Sen.Lecturer Priv.-Doz. Dr.rer.cur. Daniela Schoberer, BSc MSc

Inhaltsverzeichnis

Künstliche Intelligenz im Gesundheitswesen

Der Einsatz Künstlicher Intelligenz (KI) im Gesundheitswesen gewinnt zunehmend an Bedeutung und beeinflusst klinische Entscheidungsprozesse, Arbeitsabläufe sowie professionelle Rollen von Gesundheitsfachpersonen. Das vorliegende Dokument vertieft die im begleitenden Flyer angesprochenen Themen und bietet eine wissenschaftlich fundierte Auseinandersetzung mit Chancen, Herausforderungen und Voraussetzungen für den Einsatz von KI im klinischen Alltag.

Ziel dieses Projekts ist es, KI nicht als rein technische Innovation zu betrachten, sondern als Werkzeug, dessen Nutzen, Grenzen und Auswirkungen kritisch reflektiert werden müssen. Ergänzend zu den Literaturergebnissen werden in diesem Dokument auch die Ergebnisse einer im Rahmen des Projekts durchgeführten Umfrage unter Gesundheitspersonal zu Kenntnissen, Einstellungen und Akzeptanz von KI-Systemen dargestellt. Im Mittelpunkt stehen dabei die Perspektiven von Gesundheitsfachpersonen: ihr Vertrauen in KI-Systeme, ihr Verständnis der zugrunde liegenden Funktionsweisen sowie die ethischen, rechtlichen und organisatorischen Rahmenbedingungen, die für eine verantwortungsvolle Anwendung entscheidend sind.

Auf Grundlage aktueller Forschung werden zentrale Erkenntnisse zur Akzeptanz und Nutzung von KI-Systemen im Gesundheitswesen dargestellt. Darüber hinaus werden notwendige Kompetenzen, Schulungsbedarfe und ethische Prinzipien aufgezeigt, die für eine menschenzentrierte Integration von KI in klinische Arbeitsabläufe erforderlich sind. Das Dokument versteht sich als fachliche Vertiefung und Diskussionsgrundlage, um einen informierten, reflektierten und professionellen Umgang mit KI im Gesundheitswesen zu fördern.

Was ist Künstliche Intelligenz?

Künstliche Intelligenz (KI)-Systeme im medizinischen Kontext sind rechnergestützte Systeme, die menschliche kognitive Funktionen nachahmen, um Gesundheitsprofessionen in ihrem Aufgabengebiet zu unterstützen und deren Aufgaben zu automatisieren. KI- Systeme lernen aus großen Datensätzen, wie zum Beispiel aus Gesundheitsakten, Laborwerten oder Bildgebungen und erkennen daraus Muster. Die Anwendungsbereiche in der Medizin gliedern sich in zwei Hauptbereiche. Zum ersten Hauptbereich gehören virtuelle Systeme. Darunter versteht man Systeme, die auf maschinellem Lernen und Deep Learning basieren. Hierbei können in den gewonnenen Daten einerseits Muster erkannt werden, andererseits Klassifikationen und Strategien gebildet und schließlich Vorhersagen getroffen werden. Der zweite Hauptbereich sind physische Systeme. Dazu zählen medizinische Geräte und der Einsatz hochentwickelter Roboter (Hamet & Tremblay, 2017).

In der evidenzbasierten Medizin spielen Large Language Models (LLMs) und Retrieval-Augmented Generation (RAGs) eine zunehmende Rolle und sollen zukünftig in der Lage sein die Patientenaufklärung, klinische Entscheidungsfindung und klinische Prozesse erheblich verbessern (Woo et al., 2025). Diese werden durch große Datenmengen trainiert und zählen daher zu den virtuellen Systemen (Shool et al., 2025; Woo et al., 2025)

Large Language Models

LLMs arbeiten auf textuellen Daten und ermöglichen semantisches Verständnis sowie kontextbezogene Wissensgenerierung (Yu et al., 2025). LLMs sind in der Lage, Beziehungen zwischen einzelnen Wörtern und Satzteilen unabhängig von ihrer Position zu erkennen. Feedbackmechanismen wie *Reinforcement Learning from Human Feedback* verbessern die Verständlichkeit und den Zusammenhang medizinischer Texte. Das Sprachverständnis dieser Modelle ergibt sich aus umfangreichen Trainingsdaten und wird durch fachspezifisches Fine-Tuning weiter vertieft (Yu et al., 2025).

Die Einsatzgebiete medizinischer LLMs sind vielfältig: Sie reichen von der automatisierten Literaturauswertung über die Zusammenfassung klinischer Studien bis hin zur Unterstützung diagnostischer Entscheidungsprozesse (Yu et al., 2025). Im praktischen Einsatz finden sich spezialisierte Modelle wie ChatGPT-Med, Med-PaLM 2 oder BioGPT, die für medizinische Fragestellungen optimiert wurden.

Retrieval-Augmented Generation

RAG stellt eine Weiterentwicklung generativer Sprachmodelle dar. Die Textgenerierung wird um ein Modul zur evidenzbasierten Dokumentensuche erweitert (Xiong et al., 2024; Yu et al., 2025). In RAG-Systemen werden relevante Dokumente aus einer Wissensdatenbank abgerufen, semantisch verarbeitet und in eine zusammenhängende Antwort integriert. Dieser Mechanismus reduziert das Risiko fehlerhafter oder halluzinierter Antworten. Die generierten Inhalte beruhen auf überprüfbaren Quellen (Xiong et al., 2024).

Im medizinischen Kontext greifen RAG-Modelle auf wissenschaftliche Datenbanken (z. B. Pubmed) zurück (Xiong et al., 2024). Die Tatsache einer zugrundeliegenden Datenbank erleichtert die evidenzbasierte und zuverlässige Entscheidungsfindung (Xiong et al., 2024; Yu et al., 2025). Ein Beispiel dafür ist Elicit. Es ist ein für den medizinischen Alltag bekanntes RAG – System.

Hybride Architektur

Hybride Systeme werden nicht als eigene Systeme geführt. Eine hybride Architektur wird zu den RAG-Systemen gezählt. Das RAG-System bleibt das zentrale Element und wird mit weiteren technischen Mechanismen spezialisiert. Hierzu zählen das gezielte Trainieren der KI auf medizinische Fachinhalte und der Einsatz strukturierter Wissensnetzwerke. Unter strukturierten Wissensnetzwerkten versteht man Wissenssammlungen, in denen medizinische Begriffe, Krankheitsbilder, Behandlungen und ihre Zusammenhänge systematisch miteinander verknüpft sind. Die Erweiterung der hybriden Architektur hat das Ziel, das logische Denken der KI zu unterstützen und auf diese Weise medizinische Informationen verständlicher,

zuverlässiger und konkret auf die Anwendungssituation zugeschnitten, bereit zu stellen (Agrawal et al., 2025; Xiong et al., 2024; Yu et al. 2025).

Bewertung und Vergleich der Modellansätze

KI-Systeme, wie LLMs und RAGs eröffnen im Bereich der Medizin und Gesundheitsversorgung erhebliches Potential. Dem gegenüber steht die wesentliche Herausforderung die Qualität der generierten Informationen zu beurteilen. Standardisierte Bewertungsinstrumente, die eingesetzt werden können, um die Zuverlässigkeit und schließlich die sichere und effektive Anwendung im klinischen Bereich gewährleisten zu können, sind hierbei unerlässlich. Die Leistungs- und Qualitätsbewertung stellt methodisch eine große Herausforderung dar. Es müssen sowohl quantitative Kriterien *(Accuracy, Precision und Recall)*, als auch qualitative Kriterien (Transparenz, Interpretierbarkeit und klinischer Nutzen) berücksichtigt werden (Shool et al., 2025).

LLMs greifen bei der Informationsgenerierung auf die zugrunde liegenden Trainingsdaten zurück. Die Qualität und Zuverlässigkeit der Antworten sind daher von den Trainingsdaten, deren Inhalt teils nicht nachvollziehbar ist, abhängig. RAGs beziehen externe, aktuelle und evidenzbasierte Wissensquellen, wie klinische Leitlinien, in die Antwortgenerierung mit ein. Somit verbessert sich die Transparenz, Nachvollziehbarkeit und faktische Genauigkeit in RAGs (Neha et al., 2025).

Bewertung LLMs, RAGs und Hybridmodelle: Güte, Transparenz und Erklärbarkeit und Klinischen Eignung

In der hier vorliegenden Arbeit wird ein Überblick über verschiedene KI-Modelle gegeben und untersucht, inwieweit sich diese zur Unterstützung der Entscheidungsfindung von Gesundheitsprofessionen eignen. Zu diesem Zweck wurden drei unterschiedliche KI-Modelle ausgewählt, die für den Einsatz im medizinischen Kontext relevant sind. Die Modelle werden systematisch hinsichtlich ihrer Qualität, Zuverlässigkeit und klinischen Nutzbarkeit geprüft. In dem Systematischen Review von Shool et al. (2025) wurden 16 Parameter identifiziert, die zur Bewertung großer Sprachmodelle (LLMs), insbesondere im medizinisch/ klinischen

Bereich verwendet wurden. Es zeigte sich eine starke Fokussierung auf den Parameter Genauigkeit (Accuracy), er war der am häufigsten verwendete Parameter. Weitere häufig bewertete Parameter waren die Vollständigkeit, Lesbarkeit, Qualität, Relevanz und Zuverlässigkeit (Shool et al., 2025).

Hierbei ist jedoch anzumerken, dass überwiegend General Domain-LLMs (93,55%), wie zum Beispiel Chatgpt, evaluiert wurden. Medizinisch spezialisierte LLMs, wie zum Beispiel Meditron, BioGPT oder BioMistral wurden in der Studie von Shool et al. (2025) mit 6,45 % deutlich seltener evaluiert (Shool et al., 2025).

Nach einer orientierenden Recherche in ChatGPT und OpenEvidence zu Bewertungsmöglichkeiten von LLMs und RAGs erfolgte eine Recherche in Pubmed und GoogleScholar. Nach Durchsicht relevanter Publikationen zu Bewertungen der Güte, Transparenz, Erklärbarkeit und klinischen Eignung wurden relevante Daten extrahiert. Dabei wurde versucht Parameter wie Transparenz, Genauigkeit, Lesbarkeit, Einsatz für die klinische Entscheidungsfindung, Vorhandensein von Halluzinationen und die praktische Anwendung sowohl für LLMs, RAGs und Hybridmodelle zu ermitteln. Die Gesamteinschätzung der Parameter basierte, auf dem jeweils schlechtesten in den Publikationen berichteten Wert. Die Ergebnisse sind in den nachfolgenden Tabellen 1, 2 und 3 ersichtlich.

Tabelle 1: Bewertung der Güte von Large Language Models für klinische Fragen/Informationen

Publikation	Bewertung Large Language Modelle					
	Transparenz	**Genauigkeit (Accuracy)**	**Lesbarkeit**	**Einsatz für klinische Entscheidungsunterstützung**	**Halluzination (Fehlerrate)**	**Anwendung**
Scaff et al., 2025		Mittel	Mittel			
Xiong et al., 2024		Mittel bis hoch (60-73%)				
Agrawal et al., 2025		Hoch		Unzuverlässige Ergebnisse		
Liu et al., 2025						Erklärungen, Beispiele
Wang et al., 2024		Mittel	Variabel	Halluzinationen, Informationsverlust und Inkonsistenzen Schwach – Wissen auf Trainingszeitpunkt begrenzt	Hoch	Allgemeine Textgenerierung
Gesamteinschätzung	**Keine Angaben**	**Mittel**	**Mittel**	**Eingeschränkt**	**Hoch**	**Allgemeine Textgenerierung**

Tabelle 2: Bewertung der Güte von RAG Models für klinische Fragen/Informationen

Publikation	Bewertung RAG-Modelle						
	Genauigkeit (Accuracy)	**Lesbarkeit**	**Einsatz für klinische Entscheidungsunterstützung**	**Leistungsfähigkeit**	**Transparenz (Nachvollziehbarkeit) Quellenangabe**	**Halluzination**	**Anwendbarkeit**
Pingua et al., 2025	Sehr hoch		Sehr gut	Hoch	Hoch		
Xiong et al., 2024	Mittel bis hoch (rund 70%)						
Liu et al., 2025	Hoch	Hoch	Faktenbasiert, sehr gut geeignet	Hoch	Faktenbasiert, Hoch (kontextspezifisches Wissen aus verifizierten Quellen)		
Wang et al., 2024	Hoch, Verbessern Genauigkeit deutlich im vgl. zu LLM	Hoch		Sehr gut – aktuelle Leitlinien und Literatur abrufbar		Deutlich reduziert im vgl. zu LLM	Medizinische QA, Wissensabfragen, Dokumentanalyse
Ke et al., 2025	Sehr hoch		Sehr gut geeignet, weil effizient und zuverlässig	Hoch			
Xu et al., 2024	Kombi RAG und LLM Genauigkeit und Leistung deutlich erhöht						
Gesamteinschätzung	**Hoch**	**Hoch**	**Sehr gut geeignet**	**Hoch**	**Hoch**	**Gering**	**Medizinische QA, Wissensabfragen, Dokumentanalyse**

Anmerkungen: Abkürzung QA = Fragen

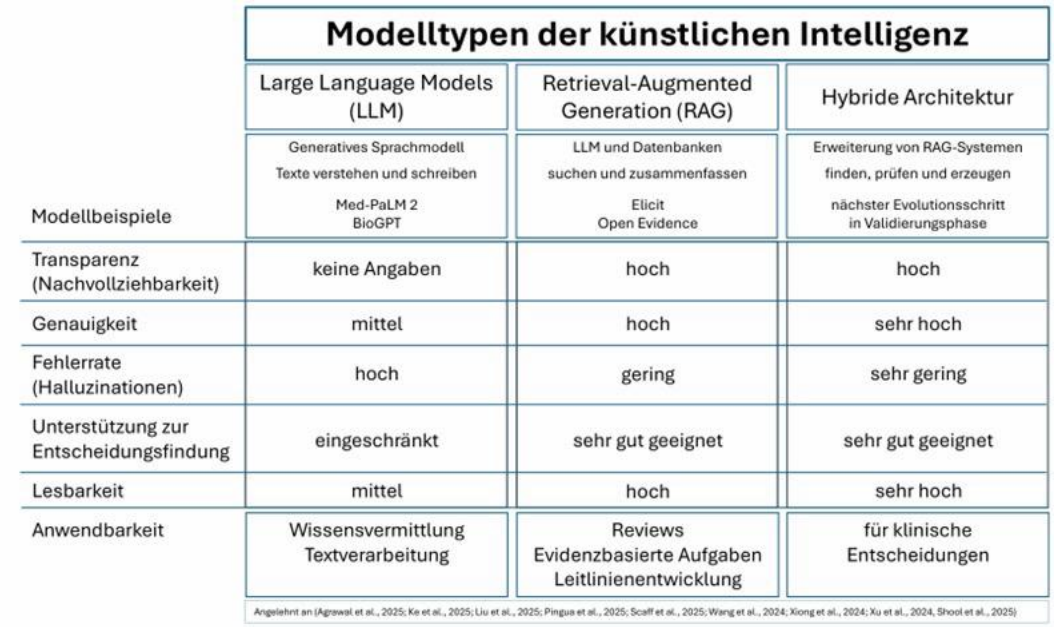
Tabelle 3: Bewertung der Güte von Hybride Architektur Models für klinische Fragen/Information

Publikation	Bewertung Hybride Architektur Modelle						
	Genauigkeit (Accuracy)	**Lesbarkeit**	**Einsatz für klinische Entscheidungsunterstützung**	**Leistungsfähigkeit**	**Transparenz (Nachvollziehbarkeit) Quellenangabe**	**Halluzination**	**Anwendbarkeit**
Wang et al., 2024	Sehr hoch, Hybride Ansätze am stabilsten	Sehr hoch	Sehr gut, Wissen aktuell		Hoch	Sehr gering	Hochkritische klinische Anwendungen, präzise Extraktion
Gesamteinschätzung	**Sehr hoch**	**Sehr hoch**	**Sehr gut geeignet**		**Hoch**	**Sehr gering**	**Hochkritische klinische Anwendungen, präzise Extraktion**

Zusammenfassung der Bewertung

Künstliche Intelligenz, insbesondere Large Language Models (LLMs) und Retrieval-Augmented Generation (RAG), gewinnt in der Medizin zunehmend an Bedeutung. Während LLMs medizinische Texte verstehen und generieren können, sind sie aufgrund statischer Trainingsdaten anfällig für Halluzinationen und veraltete Informationen. RAG-Systeme erweitern LLMs durch die Einbindung externer, evidenzbasierter Wissensquellen und verbessern dadurch Transparenz, Genauigkeit und klinische Nachvollziehbarkeit. Hybride Architekturen verstärken diese Effekte durch zusätzliche Mechanismen wie medizinisches Fine-Tuning und Wissensnetzwerke. Die Leistungs- und Qualitätsbewertung stellt methodisch eine große Herausforderung dar. Es müssen sowohl quantitative Kriterien (*Accuracy, Precision und Recall*), als auch qualitative Kriterien (Transparenz, Interpretierbarkeit und klinischer Nutzen) berücksichtigt werden. Häufige Parameter, die zur Evaluierung von KI-Modellen verwendet wurden, waren Genauigkeit, Vollständigkeit, Lesbarkeit, Qualität, Relevanz und Zuverlässigkeit.

Die Bewertung zeigt, dass reine LLMs nur eingeschränkt für die klinische Entscheidungsunterstützung geeignet sind, da sie eine mittlere Genauigkeit und eine hohe Halluzinationsrate aufweisen. RAG-Systeme erzielen eine hohe Genauigkeit, Transparenz und geringe Fehlerraten und eignen sich gut für medizinische Wissensabfragen und Entscheidungsprozesse. Hybride Modelle erreichen die beste Gesamtleistung mit sehr hoher Genauigkeit, minimalen Halluzinationen und besonderer Eignung für hochkritische klinische Anwendungen. Eine genauere Übersicht liefert Abbildung 1.



Modelltypen der künstlichen Intelligenz			
	Large Language Models (LLM)	Retrieval-Augmented Generation (RAG)	Hybride Architektur
Modellbeispiele	Generatives Sprachmodell Texte verstehen und schreiben Med-PaLM 2 BioGPT	LLM und Datenbanken suchen und zusammenfassen Elicit Open Evidence	Erweiterung von RAG-Systemen finden, prüfen und erzeugen nächster Evolutionsschritt in Validierungsphase
Transparenz (Nachvollziehbarkeit)	keine Angaben	hoch	hoch
Genauigkeit	mittel	hoch	sehr hoch
Fehlerrate (Halluzinationen)	hoch	gering	sehr gering
Unterstützung zur Entscheidungsfindung	eingeschränkt	sehr gut geeignet	sehr gut geeignet
Lesbarkeit	mittel	hoch	sehr hoch
Anwendbarkeit	Wissensvermittlung Textverarbeitung	Reviews Evidenzbasierte Aufgaben Leitlinienentwicklung	für klinische Entscheidungen

Angelehnt an (Agrawal et al., 2025; Ke et al., 2025; Liu et al., 2025; Pingua et al., 2025; Scaff et al., 2025; Wang et al., 2024; Xiong et al., 2024; Xu et al., 2024, Shool et al., 2025)

Abbildung 1: Modelltypen der künstlichen Intelligenz. Eigene Darstellung.

Wichtige Erkenntnisse zur Akzeptanz und Nutzung von KI-Systemen durch medizinisches Fachpersonal

Studien zeigen, dass häufig Unsicherheit über die Funktionsweise dieser Systeme und deren Algorithmen besteht. Dieses fehlende Verständnis beeinträchtigt das Vertrauen und führt zu Zweifeln an der Zuverlässigkeit der KI-Systeme und deren Empfehlungen. Medizinisches Fachpersonal hat Schwierigkeiten, KI-Systeme zu verstehen, insbesondere regelbasierte KI-Systeme, die auf maschinellem Lernen basieren (Ayorinde et al., 2024). Das medizinische Fachpersonal bewertete den Nutzen von KI sehr unterschiedlich. Einige sehen in KI einen Mehrwert für die Entscheidungsfindung, andere betrachten sie vor allem als Unterstützung zur Bestätigung ihrer klinischen Einschätzung, während wiederum einige keinen Nutzen erkennen. Solche unterschiedlichen Einschätzungen stellen ein häufig beschriebenes Thema in der Forschung dar (Ayorinde et al., 2024). Aktuelle Forschung zeigt, dass das Vertrauen ein entscheidender Faktor für die erfolgreiche Integration von KI-basierten klinischen Entscheidungsunterstützungssystemen in die klinische Praxis bleibt. Für die Verbesserung der Akzeptanz und Anwendung von KI-Systemen im Gesundheitswesen ist Gesundheitsfachpersonen die Transparenz der Anwendung am wichtigsten, was die Notwendigkeit einer klaren und interpretierbaren KI-Anwendung unterstreicht (Tun et al., 2025).

Wichtige Faktoren, die die Akzeptanz beeinflussen

In der Literatur lassen sich drei übergeordnete Kategorien von Faktoren identifizieren, die die Akzeptanz von KI beeinflussen. Dazu zählen erstens nutzerbezogene Faktoren wie Vertrauen, Systemverständnis, KI-Kompetenz und Technologieakzeptanz. Zweitens umfassen systembezogene Faktoren der Nutzung Aspekte wie die wahrgenommene Selbstwirksamkeit, Belastung sowie die Integration in bestehende Arbeitsabläufe. Drittens spielen soziokulturelle und organisatorische Faktoren eine Rolle, darunter sozialer Einfluss, organisatorische Bereitschaft, ethische Aspekte und die wahrgenommene Bedrohung der beruflichen Identität (Hua et al., 2023).

Hindernisse für die Einführung

Systematische Übersichtsarbeiten identifizierten mehrere zentrale Hindernisse. Zu den Faktoren, die das Vertrauen in KI beeinträchtigen, zählen die fehlende Transparenz der Algorithmen, unzureichende Schulungen sowie ethische Herausforderungen (Tun et al., 2025). Zudem äußerte medizinisches Fachpersonal Bedenken hinsichtlich eines möglichen Arbeitsplatzverlustes, einer übermäßigen Abhängigkeit von KI, rechtlicher Konsequenzen und des Datenschutzes (Sahoo et al., 2025). Darüber hinaus zeigen Studien, dass dialogorientierte große Sprachmodelle wie ChatGPT nicht immer zuverlässige Antworten bei komplexen gesundheitsbezogenen Fragestellungen liefern, die spezifisches Fachwissen erfordern (Wang et al., 2024).

Schulungs- und Ausbildungsbedarf

Aktuelle Forschung zeigt, dass Schulung und Vertrautheit mit KI entscheidende Faktoren für die Akzeptanz sind und hebt die Bedeutung von Wissensaustausch sowie Benutzerschulungen hervor (Tun et al., 2025). Zudem sind die Ausbildung und der Aufbau von Kompetenzen beim Gesundheitspersonal zentral für die Einführung, Umsetzung und Weiterentwicklung von KI im Gesundheitswesen (Saho et al., 2025). Darüber hinaus weisen Studien darauf hin, dass wichtige Aspekte medizinischer KI-Systeme häufig nicht berücksichtigt werden, insbesondere wenn theoretische Modelle verwendet werden, die gesundheitsbezogene Besonderheiten nicht ausreichend einbeziehen (Hua et al., 2023).

Integration in klinische Arbeitsabläufe

Die Benutzerfreundlichkeit des Systems stellte ein zentrales Thema dar, insbesondere im Hinblick auf die effektive Integration in klinische Arbeitsabläufe (Tun et al., 2025). Allerdings haben viele Studien die Auswirkungen auf diese Arbeitsabläufe vernachlässigt, obwohl sie für den Einsatz medizinischer KI-Systeme entscheidend sind. Die Forschung deutet darauf hin, dass eine höhere Akzeptanz von KI die Entwicklung menschenzentrierter Systeme erfordert, die sowohl klinische Prozesse als auch den institutionellen Anwendungskontext berücksichtigen (Hua et al., 2023).

Bedenken hinsichtlich Anwendbarkeit und Zuverlässigkeit

Aktuelle KI-Systeme zeigen unterschiedliche Ergebnisse hinsichtlich ihrer Anwendbarkeit und Zuverlässigkeit. Konversationsbasierte große Sprachmodelle erzielten positive Ergebnisse bei der Zusammenfassung und Vermittlung allgemeiner medizinischer Informationen für Patientinnen und Patienten (Wang et al., 2024). Die derzeitige Version von ChatGPT erreichte in verschiedenen Tests jedoch nur mäßige Ergebnisse und gilt für den direkten klinischen Einsatz als nicht ausreichend zuverlässig (Li et al., 2024).

Notwendige Kompetenzen von Gesundheitspersonal im Umgang mit KI

Mit der zunehmenden Nutzung von KI im Gesundheitswesen müssen Gesundheitsfachpersonen neue Kenntnisse und Fähigkeiten (siehe Tabelle 4) entwickeln, um KI sicher, effektiv und ethisch einsetzen zu können.

Tabelle 4: Notwendige Kompetenzen von Gesundheitspersonal im Umgang mit KI

Kompetenzbereich	Beschreibung
Technologisches Grundverständnis	Basiswissen über KI-gestützte Systeme und Anwendungsmöglichkeiten im klinischen Kontext (Russell et al., 2023; Charow et al., 2021).
Kritisches Denken	– Ergebnisse kritisch hinterfragen – Bias und Unsicherheiten in Betracht ziehen – Ergebnisse mit klinischer Erfahrung abgleichen – Ergebnisse nicht unreflektiert übernehmen (Russell et al., 2023; Choudhury & Chaudhry, 2024).
Ethik und Recht	Bewusstsein für Datenschutz, Transparenz und Haftung (Garvey et al., 2022; Svedberg et al., 2022).
Klinische Kommunikation	Fähigkeit, Ergebnisse von KI-Algorithmen verständlich und zielgruppengerecht erklären (Charow et al., 2021).
Lernen und Anpassen	– Flexibel auf neue KI-Entwicklungen reagieren – relevante Weiterbildungen wahrnehmen – Klinische Entscheidungsprozesse an neue KI-Entwicklungen anpassen (Russell et al., 2023)

Der Einsatz von KI in der Medizin birgt das Risiko einer Überabhängigkeit im Sinne eines „Deskilling “, da zentrale klinische Kompetenzen wie diagnostisches Denken und eigenständige Entscheidungsfindung zunehmend an algorithmische Systeme delegiert werden, während zugleich der Mangel an strukturierten Bildungsprogrammen den Erwerb notwendiger KI-Kompetenzen bei medizinischem Personal erschwert (Choudhury & Chaudhry, 2024; Charow et al., 2021; Garvey et al., 2022).

Ethische und rechtliche Aspekte bei der Anwendung von KI im klinischen Alltag

Der Einsatz Künstlicher Intelligenz (KI) im klinischen Alltag verändert zunehmend, die Art und Weise wie Gesundheitsfachpersonen Daten analysieren und klinische Entscheidungen treffen. KI-Systeme werden dabei nicht nur zur Diagnosestellung, sondern auch zur Prozessoptimierung, Dokumentation und Auswertung von Forschungsdaten genutzt (Naik et al., 2022; Stöger, 2024). Diese Anwendungen werfen komplexe ethische Fragen zu Verantwortung, Transparenz und menschlicher Kontrolle auf, die sowohl in der Forschung als auch in der klinischen Praxis intensiv diskutiert werden (Stöger, 2024; Jeyaraman et al., 2023).

Erklärbarkeit, Transparenz und menschliche Verantwortung

Ein zentraler ethischer Aspekt ist die Erklärbarkeit und Nachvollziehbarkeit von KI-Systemen. Viele Modelle sind aufgrund ihrer algorithmischen Komplexität schwer durchschaubar („Black-Box-Phänomen“), was die ethische und rechtliche Nachvollziehbarkeit von Entscheidungen einschränken kann (Stöger, 2024; Naik et al., 2022). Daher bleibt die Plausibilitätsprüfung durch medizinisches Fachpersonal – der sogenannte *human-in-the-loop*-Ansatz – unerlässlich. Die Verantwortung für Interpretation und Anwendung von KI-Ergebnissen liegt weiterhin bei den behandelnden Fachpersonen (Wallner, 2024; Naik et al., 2022).

Fairness, Bias und Datenschutz

Ethische Kernfragen betreffen insbesondere Fairness, Bias und Datenschutz. Verzerrte oder unausgewogene Trainingsdaten können zu struktureller Ungerechtigkeit führen und bestehende gesundheitliche Ungleichheiten verstärken (Stöger, 2024; Jeyaraman et al., 2023). Daher sind Datenqualität, Repräsentativität und ethische Prüfung der Trainingsdaten zentrale Voraussetzungen für den vertrauenswürdigen Einsatz von KI. Ebenso entscheidend ist der Schutz personenbezogener Daten sowie das Recht von Patient*innen, nicht ausschließlich

automatisierten Entscheidungen unterworfen zu werden (Wallner, 2024; Yuste et al., 2023).

Ethische Prinzipien bei der Entwicklung sowie Autonomie und Würde der Patient*innen

Darüber hinaus betonen Yuste et al. (2023) die Bedeutung von Ethical-by-Design-Prinzipien, die sicherstellen, dass KI-Systeme bereits in der Entwicklungsphase ethischen Standards und Werten der Gesundheitsversorgung entsprechen. Dazu zählen auch Transparenz, Nachvollziehbarkeit und das Prinzip der Verantwortlichkeit. Nur wenn KI-Systeme diese Anforderungen erfüllen, können sie Vertrauen schaffen und klinische Prozesse wirksam unterstützen.

Schließlich bleibt der Respekt vor der Autonomie und Würde der Patient*innen ein zentrales ethisches Prinzip. Patient*innen müssen über den Einsatz, die Funktionsweise und die Grenzen von KI-Systemen informiert werden, um Behandlungsentscheidungen bewusst mittragen zu können (Stöger, 2024; Rigby, 2019).

KI als Unterstützung, nicht als Ersatz menschlicher Urteilskraft

Zusammenfassend lässt sich sagen, dass KI klinische Entscheidungsprozesse erheblich unterstützen kann, darf sie jedoch nicht entmenschlichen. Ein ethisch verantwortlicher Einsatz bedeutet, dass KI die menschliche Urteilskraft ergänzt – nicht ersetzt (Wallner, 2024; Naik et al., 2022; Yuste et al., 2023). Nur durch interdisziplinäre Zusammenarbeit zwischen Entwickler*innen, Gesundheitsfachpersonen und Entscheidungsträgerinnen lassen sich technische, ethische und klinische Anforderungen erfolgreich miteinander verbinden (Jeyaraman et al., 2023).

Nutzung von KI-Anwendungen von Beschäftigten im Gesundheitswesen

Im Rahmen dieser Arbeit wurde eine Online-Befragung mittels vorab entwickelten Fragebogen unter Mitarbeiter*innen im Gesundheitswesen durchgeführt, um den aktuellen Einsatz von KI-Anwendungen im beruflichen Alltag zu erfassen. Ziel war es, einen Überblick über Nutzungshäufigkeit, Einsatzbereiche sowie bestehende Unterstützungs- und Informationsbedarfe im Umgang mit KI zu gewinnen. Die Ergebnisse sollen dazu beitragen, Ansatzpunkte für eine sichere, verantwortungsvolle und praxisnahe Integration von KI-Anwendungen im Gesundheitswesen aufzuzeigen.

Methodik

Die Datenerhebung wurde als nicht-standardisierte, anonyme Online-Befragung mittels Lime Survey durchgeführt. Zuvor wurde die Zustimmung der Medizinischen Universität Graz zur Durchführung der Erhebung eingeholt. Der Erhebungszeitraum erstreckte sich vom 25. November bis zum 14. Dezember, in welchem der Fragebogen für die Teilnehmenden mittels Link zur Befragungswebseite zugänglich war.

Die Verbreitung des Fragebogens erfolgte im Schneeballsystem, z.B. über WhatsApp und innerhalb von beruflichen Gruppenchats. Die Rekrutierung der Teilnehmenden erfolgte somit über eine Gelegenheitsstichprobe, da die Teilnahme nicht zielgerichtet erfolgte und über bestehende berufliche Kommunikationsstrukturen ermöglicht wurde. Zielgruppe der Befragung waren Angehörige medizinischer Berufsgruppen. Die Teilnahme war freiwillig und anonym. Es wurden keine personenbezogenen Daten erhoben, die eine Identifikation der Teilnehmenden zulassen. Die erhobenen Daten wurden ausschließlich zu wissenschaftlichen Zwecken verwendet und in aggregierter Form ausgewertet.

Insgesamt wurden 136 Fragebögen beantwortet. Davon waren 118 vollständig ausgefüllt, während 18 Fragebögen unvollständig waren. Unvollständige Datensätze wurden in der Analyse nicht erfasst.

Ergebnisse

Demographische Daten

Nachfolgend werden relevante Charakteristika der Teilnehmenden dargestellt. Erhoben wurden die Berufsgruppen, das Alter und das Geschlecht.

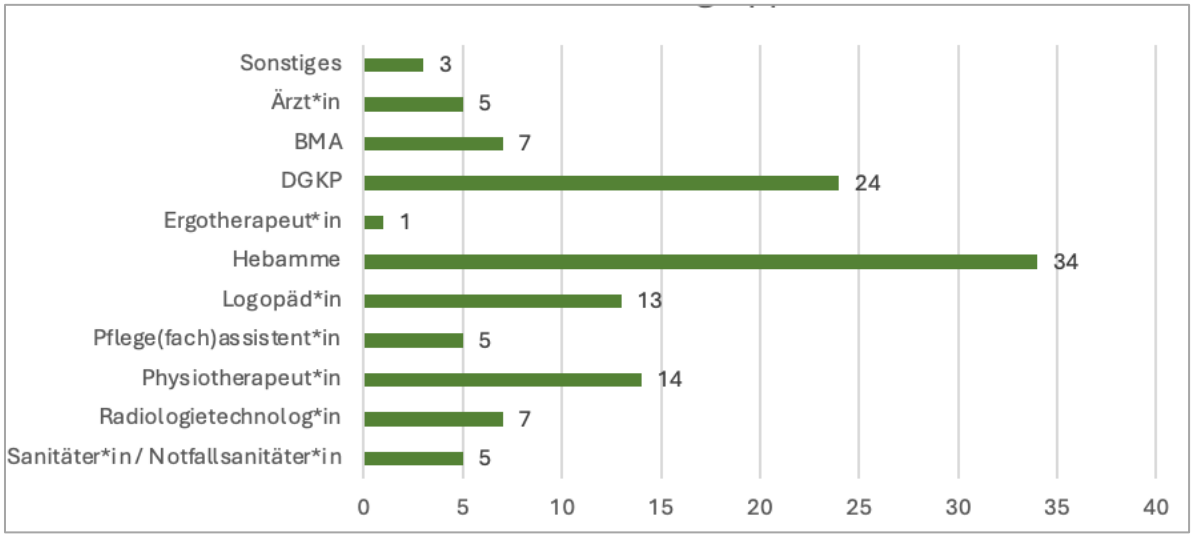


Abbildung 2: An der Umfrage teilnehmende Berufsgruppen

Die größte Teilnehmergruppe bilden Hebammen mit 28,8 % (n = 34), gefolgt von DGKP mit 20,3 % (n = 24) sowie Physiotherapeut*innen (11,9 %; n = 14) und Logopäd*innen (11,0 %; n = 13). Die übrigen Berufsgruppen sind in geringem Ausmaß vertreten (siehe Abbildung 2).

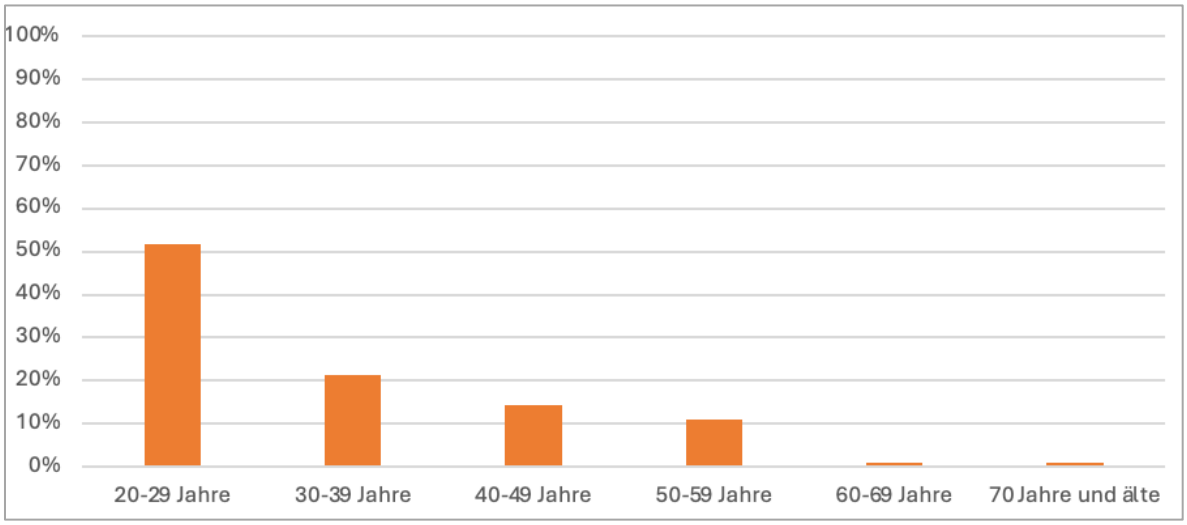


Abbildung 3: Altersverteilung in %

Die Altersverteilung zeigt einen hohen Anteil von jüngeren Altersgruppen in der Befragung: Über die Hälfte aller Teilnehmenden (N= 118, 100 %) ist 20–29 Jahre alt (ca. 51 %), gefolgt von den 30–39-Jährigen (ca. 21 %). Mit zunehmendem Alter nimmt der Anteil kontinuierlich ab, wobei Personen ab 60 Jahren nur einen sehr geringen Anteil (unter 2 %) der Stichprobe ausmachen.

87,3% der befragten Personen fühlen sich dem weiblichen Geschlecht zugehörig, während 12,7% sich als männlich einordnen.

Verwendung von KI-Anwendungen

KI-Tools werden am häufigsten für die Texterstellung genutzt werden, während die Nutzung für Fachwissen und spezifische Fragestellungen etwas geringer ausfällt. Insgesamt verdeutlicht das Diagramm eine fast ausgeglichene Verteilung zwischen Nutzern und Nicht-Nutzern, wobei in zwei von drei Kategorien die Nicht-Nutzung noch leicht überwiegt. Eine Übersicht mit Prozentzahlen bietet Abbildung 4.

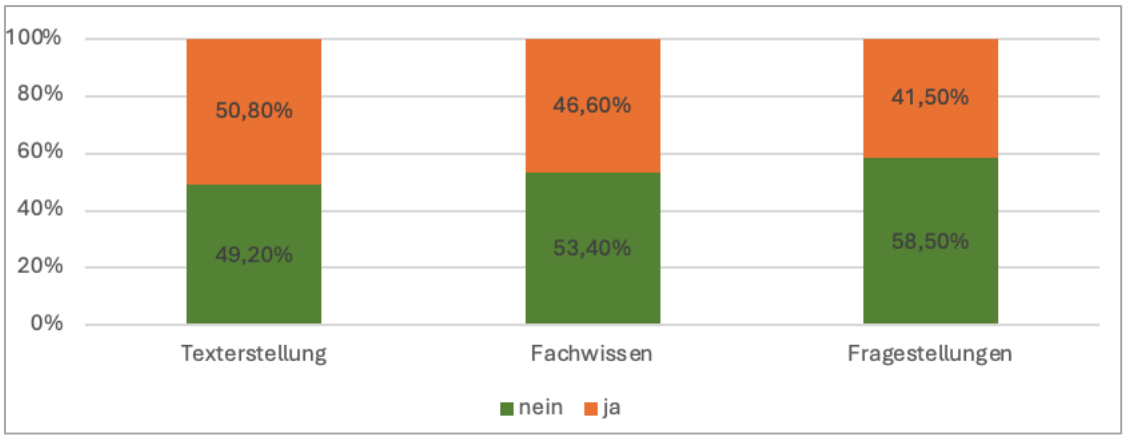


Abbildung 4: Verwendung von KI-Tools in %

Verwendung von KI-Anwendungen zur Texterstellung

Etwa die Hälfte der Befragten (50,8 % von 118 Teilnehmer*innen) setzt KI-Anwendungen zur Texterstellung ein, etwa für Dokumentationen, E-Mails oder Infomaterial. Die andere Hälfte (49,2 %) verzichtet darauf. Von den 60 Teilnehmenden,

die KI für diese Zwecke nutzen, verwenden 58,3 % die Technologie selten, 28,3 % wöchentlich und 13,3 % sogar täglich.

Verwendung von KI-Anwendungen zum Finden von aktuellem Fachwissen

Etwas weniger als die Hälfte der Befragten (46,6%) setzen KI-Anwendungen ein, um sich über aktuelles Fachwissen, wie Leitlinien oder Studien, zu informieren. Von den 55 Teilnehmenden, die KI für die Fachwissensrecherche verwenden, greifen 49,1 % selten, 45,5 % wöchentlich und 5,5 % täglich darauf zurück.

Verwendung von KI-Anwendung zum Lösen von praxisrelevanten Fragestellungen

Weniger als die Hälfte, 41,5 % der Befragten, nutzen KI-Anwendungen, um praxisrelevante Fragen zu beantworten, etwa zur Auswahl eines geeigneten Medikaments bei chronischem Husten. Von den 49 Teilnehmenden, die KI für solche Fragestellungen einsetzen, verwenden 55,1 % die Technologie selten, 40,8 % wöchentlich und nur 4,1 % täglich.

Wünsche im beruflichen Alltag zum Umgang mit KI-Anwendungen

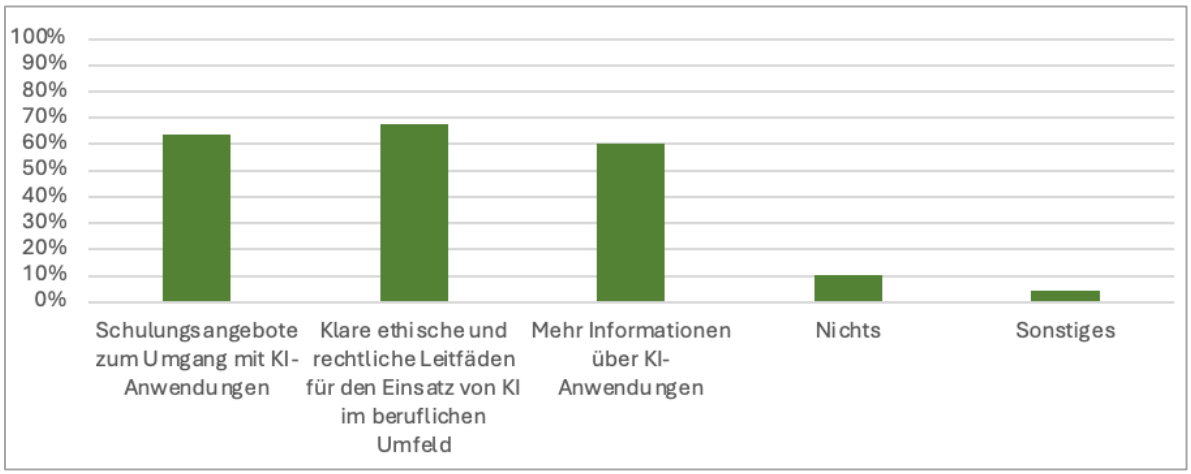


Abbildung 5: Wünsche im beruflichen Alltag zum Umgang mit KI-Anwendungen

Die Ergebnisse zeigen einen klaren Wunsch nach Unterstützungsbedarf im beruflichen Umgang mit KI-Anwendungen: Am häufigsten wünschen sich die Teilnehmenden klare ethische und rechtliche Leitfäden (rund 67 %), gefolgt von Schulungsangeboten (ca.

63 %) und mehr Informationen über KI-Anwendungen (ca. 60 %). Nur ein kleiner Anteil gibt an, keinen Bedarf zu haben (etwa 10 %) oder nennt sonstige Wünsche (rund 4 %), was auf ein insgesamt hohes Interesse an Orientierung und Qualifizierung hinweist.

Fazit

Die Ergebnisse zeigen, dass KI-Anwendungen im Gesundheitswesen bereits von einem Teil der Beschäftigten genutzt werden, jedoch noch keine breite Anwendung im beruflichen Alltag gefunden haben. Etwa die Hälfte der Befragten setzt KI zur Texterstellung ein, während die Nutzung zur Recherche von Fachwissen und zur Lösung praxisrelevanter Fragestellungen geringer ausfällt. Insgesamt überwiegt in allen Bereichen der Anteil der Nicht-Nutzer*innen.

Von den Teilnehmenden wird ein hoher Unterstützungsbedarf im Umgang mit KI gefordert: Die Mehrheit der Teilnehmenden wünscht sich klare ethische und rechtliche Leitlinien, Schulungsangebote sowie mehr Informationen zu KI-Anwendungen im Gesundheitswesen. Dies unterstreicht, dass für eine sichere und sinnvolle Integration von KI neben technischen Möglichkeiten vor allem Orientierung, Wissen und Qualifizierung notwendig sind.

Als erste Orientierungshilfe kann hierbei die begleitende Broschüre (siehe Abbildung 6) dienen. Sie klärt praxisnah über Einsatzmöglichkeiten, Chancen und Grenzen von KI im Gesundheitswesen auf.

Abbildung 6: Auszug Brochüre

Literaturverzeichnis

Agrawal, M., Chen, I. Y., Gulamali, F., & Joshi, S. (2025). The evaluation illusion of large language models in medicine. npj Digital Medicine, 8(1), 600. https://doi.org/10.1038/s41746-025-01963-x

Ayorinde, A., et al. (2023). Health care professionals' experience of using AI: Systematic review with narrative synthesis. Journal of Medical Internet Research, 26, e55766. https://doi.org/10.2196/55766

Charow, R., Jeyakumar, T., Younus, S. et al. (2021). Artificial intelligence education programs for health care professionals: Scoping review. *JMIR Medical Education, 7*(4), e31043. https://doi.org/10.2196/31043

Choudhury, A. and Chaudhry, Z. (2024). Large language models and user trust: Consequence of self-referential learning loop and the deskilling of health care professionals. *Journal of Medical Internet Research, 26*, e56764. https://doi.org/10.2196/56764

Garvey, K.V., Thomas Craig, K.J., Russell, R.G. et al. (2022). Considering clinician competencies for the implementation of artificial intelligence–based tools in health care: Findings from a scoping review. *JMIR Medical Informatics, 10*(11), e37478. https://doi.org/10.2196/37478

Hamet, P., & Tremblay, J. (2017). Artificial intelligence in medicine. *Metabolism, 69s*, S36–s40. https://doi.org/10.1016/j.metabol.2017.01.011

Hua, D., Petrina, N., Young, N., Cho, J. G., & Poon, S. K. (2024). Understanding the factors influencing acceptability of AI in medical imaging domains among healthcare professionals: A scoping review. *Artificial intelligence in medicine*, *147*, 102698. https://doi.org/10.1016/j.artmed.2023.102698

Jeyaraman, M., Balaji, S., Jeyaraman, N. and Yadav, S. (2023). Unraveling the ethical enigma: Artificial intelligence in healthcare. *Cureus, 15*(8), e43262. https://doi.org/10.7759/cureus.43262

Ke, Y. H., Jin, L., Elangovan, K., Abdullah, H. R., Liu, N., Sia, A. T. H., Soh, C. R., Tung, J. Y. M., Ong, J. C. L., Kuo, C. F., Wu, S. C., Kovacheva, V. P., & Ting, D. S. W.

(2025). Retrieval augmented generation for 10 large language models and its generalizability in assessing medical fitness. *NPJ Digit Med, 8*(1), 187. https://doi.org/10.1038/s41746-025-01519-z

Li, J., Dada, A., Puladi, B., Kleesiek, J., & Egger, J. (2024). ChatGPT in healthcare: A taxonomy and systematic review. *Computer methods and programs in biomedicine*, *245*, 108013. https://doi.org/10.1016/j.cmpb.2024.108013

Liu, Z., Agrawal, P., Singhal, S., Madaan, V., Kumar, M., & Verma, P. K. (2025). LPITutor: an LLM based personalized intelligent tutoring system using RAG and prompt engineering. *PeerJ Comput Sci, 11*, e2991. https://doi.org/10.7717/peerj-cs.2991

Naik, N., Hameed, B.M.Z., Shetty, D.K. et al. (2022). Legal and ethical considerations in artificial intelligence in healthcare: Who takes responsibility? *Frontiers in Surgery, 9*, 862322. https://doi.org/10.3389/fsurg.2022.862322

Neha, F., Bhati, D., & Shukla, D. K. (2025). Retrieval-augmented generation (RAG) in healthcare: A comprehensive review. *AI, 6*, 226. https://doi.org/10.3390/ai6090226

Pingua, B., Sahoo, A., Kandpal, M., Murmu, D., Rautaray, J., Barik, R. K., & Saikia, M. J. (2025). Medical LLMs: Fine-Tuning vs. Retrieval-Augmented Generation. *Bioengineering (Basel), 12*(7). https://doi.org/10.3390/bioengineering12070687

Rigby, M.J. (2019). Ethical dimensions of using artificial intelligence in health care. *AMA Journal of Ethics, 21*(2), E121–E124. https://doi.org/10.1001/amajethics.2019.121

Russell, R.G., Novak, L.L., Patel, M. et al. (2023). Competencies for the use of artificial intelligence–based tools by health care professionals. *Academic Medicine, 98*(3), pp. 348–356. https://doi.org/10.1097/ACM.0000000000004963

Sahoo, R. K., Sahoo, K. C., Negi, S., Baliarsingh, S. K., Panda, B., & Pati, S. (2025). Health professionals' perspectives on the use of Artificial Intelligence in healthcare: A systematic review. Patient education and counseling, 134, 108680. https://doi.org/10.1016/j.pec.2025.108680

Scaff, S. P. S., Reis, F. J. J., Ferreira, G. E., Jacob, M. F., & Saragiotto, B. T. (2025). Assessing the performance of AI chatbots in answering patients' common questions about low back pain. *Ann Rheum Dis, 84*(1), 143–149. https://doi.org/10.1136/ard-2024-226202

Shool, S., Adimi, S., Saboori Amleshi, R., Bitaraf, E., Golpira, R., & Tara, M. (2025). A systematic review of large language model (LLM) evaluations in clinical medicine. *BMC Med Inform Decis Mak, 25*(1), 117. https://doi.org/10.1186/s12911-025-02954-4

Stöger, K. (2024). Künstliche Intelligenz in der Medizin: Rechtliche Aspekte. Vortrag vor der Vorarlberger Juristischen Gesellschaft, pp. 1–3.

Svedberg, P., Reed, J., Nilsen, P. et al. (2022). Toward successful implementation of artificial intelligence in health care practice: Protocol for a research program. *JMIR Research Protocols, 11*(3), e34920. https://doi.org/10.2196/34920

Tun, H. M., Rahman, H. A., Naing, L., & Malik, O. A. (2025). Trust in Artificial Intelligence-Based Clinical Decision Support Systems Among Health Care Workers: Systematic Review. Journal of medical Internet research, 27, e69678. https://doi.org/10.2196/69678

Wallner, F. (2024). Rechtliche Risiken durch den Einsatz von KI in der Medizin. *Zeitschrift für Gesundheitspolitik, 1/2024*, pp. 81–98.

Wang, L., Wan, Z., Ni, C., Song, Q., Li, Y., Clayton, E., Malin, B., & Yin, Z. (2024). Applications and Concerns of ChatGPT and Other Conversational Large Language Models in Health Care: Systematic Review. Journal of medical Internet research, 26, e22769. https://doi.org/10.2196/22769

Wang, Y., Leutner, S., Ingrisch, M., Klein, C., Hinske, L. C., & Danhauser, K. (2024). Optimizing Data Extraction: Harnessing RAG and LLMs for German Medical Documents. *Stud Health Technol Inform, 316*, 949–950. https://doi.org/10.3233/shti240567

Woo, J. J., Yang, A. J., Olsen, R. J., Hasan, S. S., Nawabi, D. H., Nwachukwu, B. U., Williams, R. J., 3rd, & Ramkumar, P. N. (2025). Custom Large Language Models Improve Accuracy: Comparing Retrieval Augmented Generation and Artificial

Intelligence Agents to Noncustom Models for Evidence-Based Medicine. *Arthroscopy, 41*(3), 565–573.e566. https://doi.org/10.1016/j.arthro.2024.10.042

Xiong, G., Jin, Q., Lu, Z., & Zhang, A. (2024). Benchmarking retrieval-augmented generation for medicine. *Findings of the Association for Computational Linguistics ACL 2024*, pages 6233–6251, Bangkok, Thailand. Association for Computational Linguistics

Xu, R., Hong, Y., Zhang, F., & Xu, H. (2024). Evaluation of the integration of retrieval-augmented generation in large language model for breast cancer nursing care responses. *Sci Rep, 14*(1), 30794. https://doi.org/10.1038/s41598-024-81052-3

Yu, E., Chu, X., Zhang, W., Meng, X., Yang, Y., Ji, X., & Wu, C. (2025). Large language models in medicine: Applications, challenges, and future directions. *International Journal of Medical Sciences, 22*(11), 2792–2801. https://doi.org/10.7150/ijms.111780

Yuste, R., Goering, S., Arcas, B.A. et al. (2023). Ethical priorities for artificial intelligence in the clinical context. *Frontiers in Medicine, 11*, 1343456. https://doi.org/10.3389/fmed.2023.1343456

Abbildungsverzeichnis

Tabellenverzeichnis