Fraunhofer IAIS

Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS

»Healthcare Agents« im Einsatz

Die nächste Generation der medizinischen KI-Assistenz

www.iais.fraunhofer.de/healthcare-agents

Autorinnen und Autoren

Sina Mackay ist Data Scientist am Fraunhofer IAIS und arbeitet im Bereich Healthcare Analytics. Sie leitet Projekte, die sich mit der Anwendung von Künstlicher Intelligenz in medizinischen Kontexten befassen. Ihr Arbeitsfokus liegt auf der Analyse und Nutzung medizinischer Daten zur Unterstützung von Forschungs- und Entwicklungsprojekten im Gesundheitswesen.

Julia Ebert ist Senior Consultant und Expertin für KI in Healthcare in der Business Line Healthcare bei adesso SE.

Dr. Alexander Schellinger ist Director Digital Transformation Consulting bei Siemens Healthineers Consulting.

Dario Antweiler ist Teamleiter des Geschäftsfelds Healthcare Analytics am Fraunhofer IAIS und betreut Projekte in den Themenbereichen Digitalisierung im Krankenhaus sowie KI in der Pharmakologie. Sein Forschungsfeld ist Machine Learning und Visual Analytics im Gesundheitswesen.

Dr. Stefan Rüping ist Leiter der Abteilung Knowledge Discovery am Fraunhofer IAIS. Er forscht in den Bereichen Explainable AI und Natural Language Processing.

Das Fraunhofer IAIS

Als Teil einer der führenden Organisationen für anwendungsorientierte Forschung ist das Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS mit Sitz in Sankt Augustin/Bonn und einem Standort in Dresden eines der führenden Wissenschaftsinstitute auf den Gebieten Künstliche Intelligenz (KI), Maschinelles Lernen und Big Data in Deutschland und Europa. Rund 380 Mitarbeitende unterstützen Unternehmen bei der Optimierung von Produkten, Dienstleistungen und Prozessen sowie bei der Entwicklung neuer digitaler Geschäftsmodelle. Das Fraunhofer IAIS gestaltet die digitale Transformation unserer Arbeits- und Lebenswelt: mit innovativen KI-Anwendungen für für Industrie, Gesundheit, Handel, Public Sector, Medien, Finance und Nachhaltigkeit, mit zukunftsweisenden Technologien wie großen KI-Sprachmodellen oder Quantum Machine Learning, mit Angeboten für die Aus- und Weiterbildung oder für die Prüfung von KI-Anwendungen auf Sicherheit und Vertrauenswürdigkeit.

Das Fraunhofer IAIS steht im Zentrum eines starken Forschungs- und Transfernetzwerks: Als einer von vier führenden Partnern betreibt das Fraunhofer IAIS KI-Spitzenforschung im »Lamarr-Institut für Maschinelles Lernen und Künstliche Intelligenz«, das als Teil der KI-Strategie der Bundesregierung dauerhaft vom Bund und Land Nordrhein-Westfalen gefördert wird. Den direkten Transfer der Forschungsergebnisse, insbesondere in mittelständische Unternehmen, sichert unter anderem die »Kompetenzplattform KI.NRW«, die das Fraunhofer IAIS als zentrale Vernetzungsinitiative des Landes Nordrhein-Westfalen zum Thema Künstliche Intelligenz leitet.

Weiterhin koordiniert das Fraunhofer IAIS als geschäftsführendes Institut die »Fraunhofer-Allianz Big Data und Künstliche Intelligenz«, welche die branchenübergreifende Expertise von über 30 Fraunhofer-Instituten bündelt und Fachkräfte aus Unternehmen mit einem etablierten und umfassenden Schulungsprogramm in Data Science und KI weiterbildet. Daneben bestehen langjährige enge Kooperationen in Forschung und Lehre mit der Universität Bonn und dem »Bonn-Aachen International Center for Information Technology« (b-it), das von der Fraunhofer-Gesellschaft, der Universität Bonn, der RWTH Aachen und der Hochschule Bonn-Rhein-Sieg getragen wird.

www.iais.fraunhofer.de

Inhalt

1. Executive Summary

Demografischer Wandel, steigende Kosten und akuter Fachkräftemangel stellen das Gesundheitswesen und insbesondere die Patientenversorgung vor erhebliche Herausforderungen. In diesem Kontext wird der Einsatz von KI-Agenten als vielversprechende Lösung diskutiert. Diese intelligenten Systeme gehen über die Funktionen klassischer Künstlicher Intelligenz (KI) hinaus, indem sie proaktive Entscheidungen treffen und in Echtzeit auf komplexe Situationen reagieren können. KI-Agenten verbinden verschiedene Technologien zur Optimierung medizinischer und administrativer Prozesse miteinander, was zu einer verbesserten Effizienz und Qualität der Patientenversorgung führt.

Die Anwendung von KI im Gesundheitswesen erfordert jedoch eine sorgfältige Planung, insbesondere im Hinblick auf Datenschutz und Informationssicherheit. Benutzerfreundliche Schnittstellen und die Förderung der digitalen Bildung sind entscheidend für die Akzeptanz bei Anwenderinnen und Anwendern dieser Systeme. Um die Herausforderungen der modernen Patientenversorgung erfolgreich zu meistern, sollten sich technische Entwicklungen künftig auf die Integration umfassender Agentensysteme konzentrieren, die verschiedene KI-Technologien verwenden. So kann eine proaktive Versorgung ermöglicht werden. Dabei ist eine interdisziplinäre Zusammenarbeit zwischen Fachleuten und die iterative Entwicklung von Lösungen unerlässlich.

In diesem Whitepaper erläutern wir die Unterschiede zwischen KI-Agenten und Großen KI-Sprachmodellen (engl.: Large Language Models, LLMs), zeigen praxisnahe Anwendungsbeispiele und diskutieren potenzielle Risiken, um eine fundierte Wissensgrundlage für den erfolgreichen Einsatz dieser Technologien im Gesundheitswesen zu schaffen.

2. Hintergrund

2.1. Künstliche Intelligenz im Gesundheitswesen

Die Patientenversorgung steht vor immer größer werdenden Herausforderungen. Der demografische Wandel [1] ist mit steigenden Kosten für die Gesundheitsversorgung verbunden [2], während zunehmend Fachkräftemangel in vielen Bereichen herrscht [3]. Diese Faktoren verdeutlichen den dringenden Bedarf an innovativen Lösungen zur Prozessautomatisierung und Effizienzsteigerung im Gesundheitssektor.

Die Organisation von Leistungen im Gesundheitswesen folgt Kernprozessen wie Eintritt, Anamnese, Diagnose, Therapie, Pflege und Entlassung und umfasst verschiedene unterstützende Prozesse (siehe Abbildung 1, primäre und sekundäre Wertschöpfungsprozesse). KI kann beispielsweise bei patientenbezogenen Prozessen wie der Auswertung von Anamnesebögen oder bei patientennahen Prozessen wie der Unterstützung bei Diagnosen und der Therapieplanung helfen. Dadurch kann die Effizienz und Genauigkeit der medizinischen Dokumentation gesteigert werden. KI kann darüber hinaus auch administrative Aufgaben übernehmen, etwa das Verfassen von Entlassbriefen (patientennahe Prozesse) oder die Abrechnungscodierung (patientenferne Prozesse). Sorgfältige Planung, insbesondere im Hinblick auf Datenschutz und Informationssicherheit, sind bei der Implementierung von KI im Gesundheitswesen jedoch unerlässlich. Es ist entscheidend, dass diese Systeme sowohl die gesetzlichen Vorgaben einhalten als auch das Vertrauen der Nutzenden gewinnen.

Zusammenfassend stellt die Anwendung von KI in der Patientenversorgung eine vielversprechende Lösung dar, um den Herausforderungen des Gesundheitswesens effektiv zu begegnen. KI-Technologien können als Grundlage für umfassendere Systeme dienen, die eine qualitativ hochwertige und effiziente Versorgung gewährleisten.

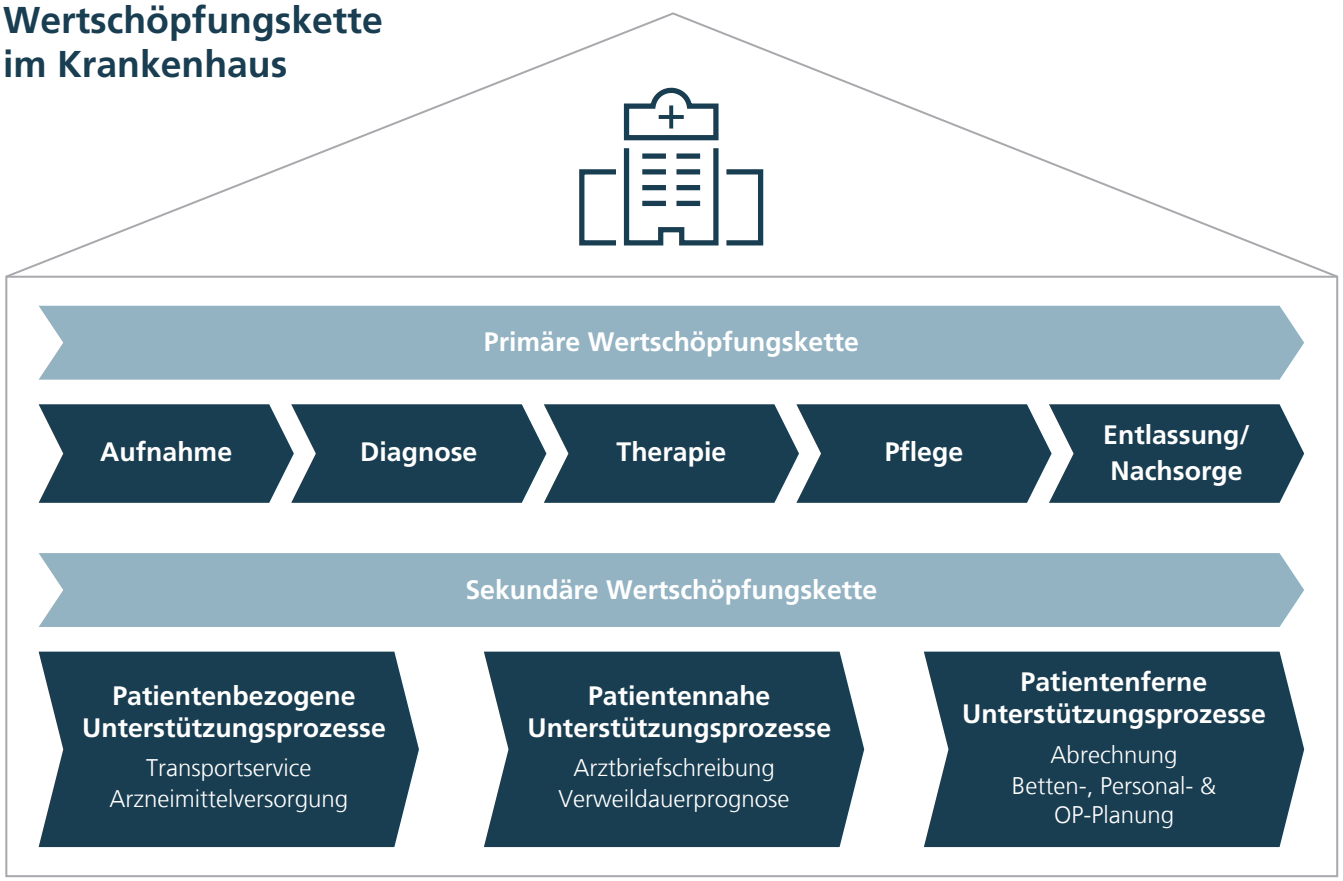

Abbildung 1: Darstellung der primären und sekundären Wertschöpfungskette im Krankenhaus

2.2. Agentenbasierte Systeme

Vor dem Hintergrund der steigenden Anforderungen gewinnen KI-Agentensysteme an Bedeutung. KI-Agenten sind intelligente Systeme, die über die Funktionen von Large Language Models (LLMs) hinausgehen, indem sie nicht nur Texte und Bilder verarbeiten, sondern auch komplexere Ziele selbstständig verfolgen können. Beispielsweise treffen sie proaktive Entscheidungen oder reagieren in Echtzeit auf Ereignisse in komplexen Umgebungen und setzen hierfür externe Werkzeuge ein (siehe Tabelle 1). Diese Systeme bestehen aus verschiedenen Komponenten, die synergistisch zusammenarbeiten, um die Effizienz und Qualität der Patientenversorgung zu verbessern. Sie integrieren verschiedene KI-Technologien, um medizinische und administrative Prozesse im Gesundheitswesen zu optimieren.

Im Bereich der Künstlichen Intelligenz existieren intelligente Agenten schon seit längerer Zeit, jedoch waren diese traditionell oft aufgabenspezifisch und funktionierten nach dem Paradigma »Wahrnehmen, Denken, Handeln«. Das bedeutet konkret, dass sie ihre Umgebung in der Vergangenheit mithilfe von Sensoren wahrnahmen, Informationen verarbeiteten und dann durch bestimmte Aktionen interagierten [4]. Im Gegensatz dazu bieten aktuelle KI-Agenten eine breitere Anwendbarkeit (siehe Tabelle 1).

KI-Agenten können Aufgaben automatisieren, die über einfache Datenverarbeitung hinausgehen, wie etwa die Analyse von Patientendaten, die Durchführung von Diagnosen und die Koordination von Behandlungsabläufen. Sie sind besonders nützlich in Szenarien, in denen schnelle Entscheidungen getroffen und die Konsequenzen dieser Entscheidungen evaluiert und an zukünftige Handlungen angepasst werden müssen. Durch ihre Fähigkeit, aus unterschiedlichen Datenquellen zu lernen und sich an wechselnde Bedingungen anzupassen, tragen KI-Agenten zur Verbesserung der Effizienz und Qualität der Patientenversorgung bei.

2.2.1. Planung

Ein zentraler Unterschied von KI-Agenten gegenüber klassischen Large Language Models (LLMs) liegt in ihrer Fähigkeit zur autonomen Handlungsplanung. Während LLMs primär darauf ausgelegt sind, Texte zu generieren und Fragen basierend auf vorgegebenen Eingaben zu beantworten, können KI-Agenten ihr Vorgehen beim Problemlösen selbstständig strukturieren, Teilziele definieren und Strategien zur Zielerreichung adaptiv anpassen.

Der Vorteil liegt in der erhöhten Effizienz und Flexibilität: KI-Agenten können medizinische Diagnosen, Therapieoptionen oder Forschungsstrategien iterativ entwickeln und optimieren, anstatt lediglich statische Empfehlungen auszugeben. Allerdings ergeben sich auch Herausforderungen, insbesondere hinsichtlich der Nachvollziehbarkeit und Sicherheit solcher autonomen Entscheidungen.

Die Fähigkeit zur Selbstplanung erfordert robuste Mechanismen zur Validierung und Kontrolle, um Fehlschlüsse oder unerwartete Entscheidungswege frühzeitig zu erkennen und zu korrigieren. Die Entwicklung transparenter, regulierbarer und ethisch vertretbarer Planungsalgorithmen ist daher eine essenzielle Voraussetzung für den erfolgreichen Einsatz von KI-Agenten im Gesundheitswesen.

Large Language Models (LLMs)	KI-Agenten
interagieren »gedächtnislos«	besitzen ein Langzeitgedächtnis für kontextbezogene Interaktionen
sind beschränkt auf Wissen aus eigenem Trainingsdatensatz	nutzen Werkzeuge, um auf externes Wissen zuzugreifen
erledigen klar definierte Aufgaben	zerlegen komplexe Ziele selbstständig in Unterschritte
sind in hohem Maße abhängig von menschlichen Prompts	sind nur geringfügig abhängig von menschlichen Prompts
sind typischerweise in der Lage, Texte und Bilder zu verarbeiten	verarbeiten multimodale Inputdaten

Tabelle 1: Unterschiede zwischen Large Language Models (LLMs) und KI-Agenten [5]

2.2.2. Verwendung externer Werkzeuge

KI-Agenten im Gesundheitswesen können externe Werkzeuge wie medizinische Datenbanken, diagnostische Algorithmen, Bildverarbeitungssoftware oder Schnittstellen sowie andere Agentensysteme nutzen, um ihre Funktionalität zu erweitern (siehe Abbildung 2).

Ein Beispiel für die Nutzung externer Werkzeuge durch KI-Agenten im Gesundheitswesen ist die Erstellung von Arztbriefen. Ein KI-Agent analysiert zunächst die Informationen über den Patientenaufenthalt, um eine grobe Struktur des Briefes festzulegen. Dabei wird auf interne Richtlinien zur Erstellung von Briefen geachtet. Aus externen Berichten und einem Anamnese-Protokoll werden Vorerkrankungen und Allergien extrahiert und in den Arztbrief integriert. Anschließend werden relevante Befunde mittels eines Werkzeugs zusammengefasst und als Abschnitte in den Brief eingefügt. Falls notwendig, werden fremdsprachige Notizen übersetzt und Grammatikfehler korrigiert. Danach wird das Codierungs-Werkzeug genutzt, um passende ICD- und OPS-Codes zu erstellen und mit dem Behandlungsverlauf abzugleichen. Am Ende wird der finale Arztbrief exportiert und dem Patienten sowie den nachbehandelnden Ärztinnen und Ärzten zur Verfügung gestellt.

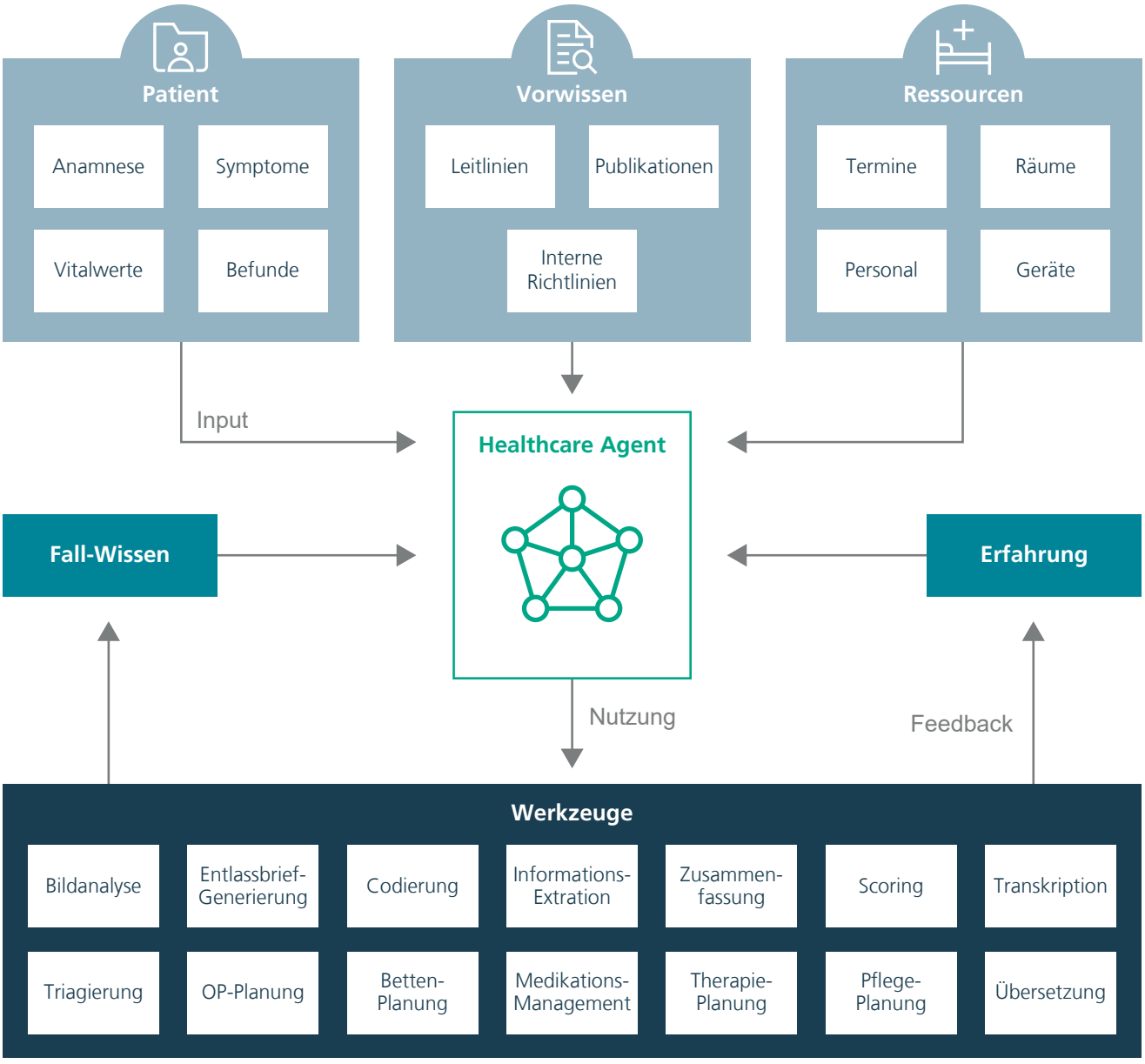

Abbildung 2: Schematischer Aufbau eines KI-Agenten im Gesundheitsbereich

Die Zusammenarbeit von mehreren KI-Agenten spiegelt insofern genau die medizinische Praxis von interdisziplinärer Entscheidungsfindung wider, wie sie beispielsweise in einem Tumorboard zu finden ist. Darüber hinaus können KI-Agenten auch externe Datenbanken verwenden, die Informationen zu Medikamenten oder Genmutationen enthalten – um noch weitere Beispiele der Anwendung zu nennen.

2.2.3. Gedächtnis

KI-Agenten können ein Langzeitgedächtnis aufbauen und kontinuierlich aus Patientenfällen lernen. Während herkömmliche LLMs zwar auf umfangreichen Trainingsdaten basieren, fehlt ihnen die Möglichkeit, Informationen aus vergangenen Interaktionen kontinuierlich zu speichern und sofort in zukünftige Entscheidungen einfließen zu lassen [4]. Ihre Kontextfenster sind begrenzt, sodass Kontinuität über mehrere Interaktionen hinweg nicht immer ausreichend gegeben ist.

KI-Agenten mit Langzeitgedächtnis können hingegen patientenspezifische Verlaufsdaten, Behandlungsstrategien und klinische Muster langfristig als Kontext erfassen. Dies ermöglicht personalisierte Empfehlungen, präzisere Diagnosen und eine verbesserte Kontinuität der Versorgung (siehe Abbildung 3).

2.2.4. Arbeitsweise von KI-Agenten

KI-Agenten im Gesundheitswesen operieren nach einem iterativen Schema: *Nutzereingabe* › *Planung* › *Handlung* › *Beobachtung*. Ein Beispiel ist ein KI-gestützter Gesundheitsassistent, der eine Medikamenteninteraktion überprüfen soll. Der Nutzer fragt das System per Eingabe: »Verträgt sich Medikament A mit Medikament B?« Der Agent analysiert zunächst medizinische Datenbanken und Leitlinien, um mögliche Wechselwirkungen zu identifizieren (Planung). Anschließend gibt er eine fundierte Empfehlung oder Warnung aus (Handlung). Falls der Nutzer Rückfragen hat oder zusätzliche Informationen liefert, passt der Agent seine Antwort entsprechend an (Beobachtung) [6].

2.3. Herausforderungen beim Einsatz von KI-Agenten im Gesundheitswesen

Ein erhebliches Hindernis für die Einführung von KI im Gesundheitswesen ist bisher noch die Zuverlässigkeit der Systeme, da Fehler schwerwiegende Folgen haben können (siehe auch Kapitel 4.2). Hier bieten KI-Agenten Verbesserungspotenzial: Sie können Werkzeuge nutzen, um so genannte »Halluzinationen«– das sind sprachlich und syntaktisch korrekte aber faktisch falsche Ausgaben eines KI-Sprachmodells – zu reduzieren

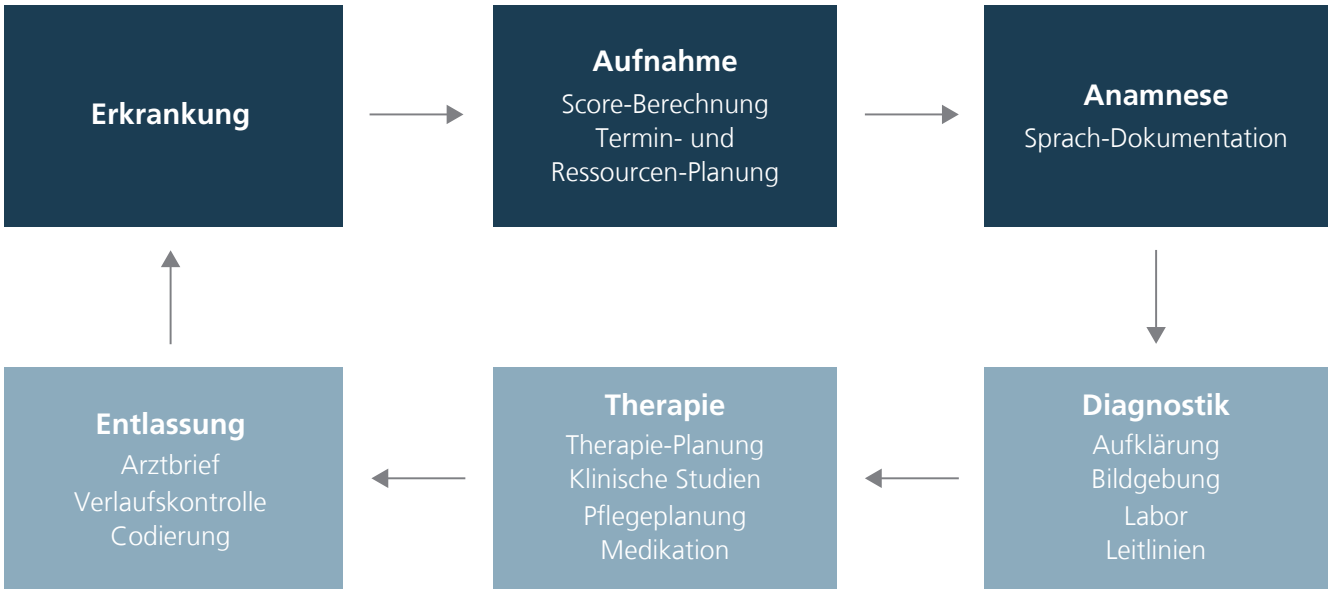

Abbildung 3: Ablaufschema für die Versorgungkette innerhalb eines Krankenhausaufenthaltes mit KI-gestützten Prozessen zur Effizenzsteigerung

und Vertrauenswürdigkeit sowie Zuverlässigkeit zu verbessern. Ein konkreter Lösungsansatz zur Verbesserung der Zuverlässigkeit von KI-Systemen ist die *Retrieval Augmented Generation* (RAG), mit deren Hilfe Wissen aus Suchmaschinen, Datenbanken oder aktuellen klinischen Leitlinien abgerufen werden kann. Solche externen Wissensgrundlagen können ebenfalls für eine interne Prüfung genutzt werden, sodass eine Gegenkontrolle stattfindet.

Herausforderungen bei der Verwendung externer Werkzeuge wiederum sind die Verfügbarkeit von interoperablen Schnittstellen, eine einheitliche Nomenklatur sowie Datenschutz und Informationssicherheit. Eine große Anzahl beteiligter Systeme erhöht die Gefahr von Ausfällen oder unerkannten Veränderungen und macht die Prüfung der Datenqualität und damit Vertrauenswürdigkeit der verwendeten Informationen schwieriger. Ebenso sinkt die Erklärbarkeit der Entscheidungen, da die innere Komplexität der Agentensysteme ungleich höher ist.

2.4. Large Language Models (LLMs) oder Agenten? Einsatzgebiete und Entscheidungsfaktoren

Large Language Models (LLMs) sind besonders gut geeignet für Aufgaben, die eine Verarbeitung und Generierung von Text basierend auf bestehendem Wissen erfordern. Sie können bei der Analyse von Patientendaten, der Erstellung von Dokumentationen oder der Beantwortung einfacher Anfragen unterstützen. Durch ihre Fähigkeit, große Mengen an Informationen zu verarbeiten und Muster zu erkennen, finden sie Anwendung in der Diagnostik und Patientenkommunikation. Allerdings stoßen LLMs an ihre Grenzen, wenn es darum geht, Echtzeit-Entscheidungen zu treffen oder auf dynamische Ereignisse in komplexen Umgebungen zu reagieren. Wenn jedoch Kontinuität kein Fokus ist und qualitativ hochwertige Prompts bereitgestellt werden können, bieten LLMs eine robuste und schnell integrierbare Lösung, die in vielen Szenarien effizient eingesetzt werden kann.

Im Gegensatz dazu sind KI-Agenten für komplexere Aufgaben besser geeignet. Sie erfordern proaktive Entscheidungsfindung und die Fähigkeit, in Echtzeit auf Ereignisse zu reagieren. Diese Agenten nutzen die Integration verschiedener KI-Technologien, um Prozesse zu optimieren, und benötigen externe Module wie Wahrnehmung, Gedächtnis und Aktionsmodule. Sie können vergangene Interaktionen abrufen und personalisierte Antworten basierend auf der Krankengeschichte, den Medikamenten und den Vorlieben der Patienten geben. Dies ermöglicht eine verbesserte Kontinuität der Versorgung und die Fähigkeit, sich an neue Informationen anzupassen und durch Interaktionen zu lernen.

Wissenschaftliche Studien zeigen, dass die starke Leistung von LLMs bei medizinischen Zulassungsprüfungen das Interesse an ihrem Einsatz in klinischen Entscheidungsszenarien mit realen Patienten geweckt hat [7]. Allerdings testen diese Prüfungen nicht die Fähigkeiten, die für die klinische Entscheidungsfindung in der realen Welt erforderlich sind. Zudem zeigen andere Untersuchungen, dass LLMs keine persistente Erinnerung haben, was ihre Relevanz in dynamischen, sich schnell ändernden Umgebungen einschränkt [4]. Weiterhin zeigen LLMs eine geringe Performanz bei spezifischen Anwendungsfällen wie medizinischer Codierung, da sie wenig hochqualitative Trainingsdaten gesehen hatten [8]. In solchen Fällen hilft besonders die Ergänzung durch externes Hintergrundwissen [4], [6].

Zusammenfassend sind LLMs vor allem für einfachere, textbasierte bzw. unimodale Aufgaben geeignet, während KI-Agenten in komplexeren, interaktiven und dynamischen Szenarien Vorteile bieten. Diese Agenten sind entscheidend, um die Herausforderungen und Anforderungen in der modernen Patientenversorgung zu bewältigen, insbesondere wenn es darum geht, proaktive Entscheidungen zu treffen und die Qualität der Versorgung zu verbessern.

3. Anwendungsszenarien

Aufgrund der signifikanten Potenziale, insbesondere in Bezug auf die Steigerung der Effizienz, ist in den kommenden Jahren ein verstärkter Einsatz von LLMs und KI-Agenten im klinischen Alltag zu erwarten. Diese Technologien können als Serviceschicht fungieren, die verschiedene Datensilos innerhalb einer Klinik aufbrechen und Informationen effizient kombinieren.

Es ist jedoch zu berücksichtigen, dass die Einbindung von KI in den Behandlungsprozess, beispielsweise durch die Generierung von Empfehlungen oder die Unterstützung bei Entscheidungen, die Klassifizierung als Medizinprodukt erfordert. Dies wiederum hat Konsequenzen für die Zulassung solcher Produkte.

3.1. Medizinische Prozesse

Entscheidungsunterstützung in Diagnose und Therapie

Die Fähigkeit der KI, große Informationsmengen zu aggregieren und zu bewerten, eröffnet neue Möglichkeiten der Entscheidungsunterstützung im klinischen Alltag. KI-Agenten können beispielsweise sowohl die gesamte Patientenhistorie zusammenfassen als auch Echtzeitinformationen aus verschiedenen Datensilos (z. B. Krankenhaus-Informationssystem KIS, Picture Archiving and Communication System PACS) verarbeiten, um proaktiv Überlegungen zu möglichen Behandlungspfaden und -mustern anzustellen. Diese können dem medizinischen Personal als Impulse dienen, um die Qualität der Patientenversorgung zu verbessern.

Medikationsmanagement

Der Einsatz von KI-Agenten zur Erstellung individueller Medikationspläne basiert auf Patientendaten wie Alter, Gewicht, Diagnosen und Unverträglichkeiten. Diese Pläne werden kontinuierlich anhand von Echtzeitinformationen weiterentwickelt und modifiziert.

Der KI-Agent nutzt dazu Echtzeitinformationen über die Medikamentengabe, die in digitaler Form dokumentiert sind. Er überwacht die Einhaltung des erstellten Plans und gibt bei Abweichungen, beispielsweise bei vergessenen Vergaben, entsprechende Hinweise an das medizinische Personal.

Zusätzlich werden das Klinikpersonal und die Logistik entlastet, indem Medikamente automatisch nachbestellt werden.

3.2. Administrative Prozesse

Informationsextraktion für die Befüllung von Qualitätsregistern

Qualitätsregister sind systematische Datensammlungen zur Bewertung und Verbesserung der medizinischen Versorgung. Im Prozess der Erstellung dieser Register ist es essenziell, dass die entsprechenden medizinischen Dokumente erneut durch das klinische oder administrative Personal geprüft werden, um die erforderlichen Merkmale für die Füllung der Qualitätsregister zu ermitteln.

KI-Agenten und LLMs können diese Arbeit erleichtern, indem automatisch die entsprechenden Dokumente gescreent und die erforderlichen Parameter herausgefiltert werden. Dies entlastet das Personal und erleichtert der Klinik den Erhalt von Zertifizierungen, sofern das Füllen von Qualitätsregistern dafür notwendig ist.

Verfassen von Arztbriefen

Beim Verfassen des Arztbriefes werden zahlreiche medizinische Informationen abgebildet, welche unter anderem für die nachgelagerte Versorgung des Patienten von Relevanz sind. Für die Zusammenfassung von Informationen aus unterschiedlichen Quellen können LLMs verwendet werden, um auf Basis von strukturierten Daten zusammenhängende und qualitätsgesicherte Texte zu formulieren.

Diese Vorgehensweise kann als Grundlage für KI-Agenten dienen, die beispielsweise die Koordination von nachgelagerten Leistungserbringern proaktiv unterstützen, um die medizinische Versorgung der Patienten zu gewährleisten.

Allerdings werden bereits beim Einsatz von LLMs für das Verfassen des Arztbriefes erhebliche Vorteile für das medizinische Personal realisiert, indem die Zeit für den administrativen Aufwand reduziert wird und mehr Zeit für die Versorgung des medizinischen Personals eingesetzt werden kann.

4. Herausforderungen

4.1. Datenschutz und Informationssicherheit

Beim Einsatz von KI im Gesundheitswesen und auch bei der Verwendung von KI-Agenten müssen hohe Anforderungen an Datenschutz und IT-Sicherheit erfüllt werden. KI-Agenten müssen als Softwareprodukt den allgemeinen IT-Sicherheitsanforderungen entsprechen, aber auch besondere Regularien für den Umgang mit personenbezogenen und hochsensiblen Daten bedienen.

Ein KI-Agent kann aus zwei Perspektiven betrachtet werden:

1. **Softwareprodukt**: Klassische IT-Sicherheitsmaßnahmen stehen im Vordergrund, um die Integrität und Verfügbarkeit der Software sicherzustellen. Außerdem sorgen die neuen Anforderungen des EU AI Acts für neue Standards zum Qualitätsmanagement und Sicherheitseinstufungen der Produkte. KI-Systeme werden entsprechend ihrem Risikopotenzial klassifiziert und insbesondere für Anwendungen im medizinischen Bereich gelten strenge Vorgaben. [9]

2. **Umgang mit Gesundheitsdaten**: Insbesondere KI-Agenten kommen auch nach der Integration der Anwendung mit besonders schützenswerten und hochsensiblen Daten in Berührung. Somit muss nicht nur bei der Entwicklung und dem Training des Systems ein Standard gehalten werden, sondern auch während der Laufzeit und der Benutzung, da Daten gespeichert werden und auch ad hoc aus ihnen gelernt werden kann (siehe auch Kapitel 2.2.3). Die Grundlage für den Schutz personenbezogener Daten in der Europäischen Union bildet auch in diesem Fall die Datenschutz-Grundverordnung (DSGVO), welche konkrete Regelungen zu der Erhebung, Verarbeitung und Speicherung von personenbezogenen Daten festlegt.

Agenten, welche eigenständig beliebige Infrastrukturen kompromittieren, waren laut Aussage des Bundesamts für Sicherheit in der Informationstechnik BSI im April 2024 nicht verfügbar [10]. Dass LLM-basierte Agenten Teile eines Angriffs in Zukunft automatisieren, ist allerdings derzeit nicht auszuschließen [4], [10].

Umso wichtiger ist es, dass entsprechende Sicherheitskonzepte eingehalten werden. Diese gelten jedoch global für die gesamte Einrichtung und Infrastruktur und beziehen sich nicht ausschließlich auf eine einzelne Anwendung wie die KI-Agenten. Daher ist es elementar, für IT- und Cybersicherheit höchste Standards und Prioritäten einzuräumen und entsprechende Abwehrmaßnahmen zu ergreifen. Dies umfasst die Verbesserung des Patchmanagements, Aufbau einer resilienten IT-Infrastruktur, gezielte Angriffserkennung sowie ausgereifte Benutzergruppen- und Sicherheitskonzepte. Auch die Schulung zur korrekten Nutzung von KI-Agenten und die Sensibilisierung für Datenschutz und IT-Sicherheit bei Nutzenden sowie Patienten ist maßgeblich für ein gutes Sicherheitskonzept und für das Vertrauen in die Technologie (siehe auch Kapitel 4.3).

4.2. Vertrauenswürdigkeit

Wie auch bei LLMs muss die Vertrauenswürdigkeit von KI-Agenten in mehreren Dimensionen verankert sein [11]. Besonders hervorzuheben sind hier die Herausforderungen, die die erhöhte Autonomie im Vergleich zu LLMs mit sich bringt. Die Komplexität der Abläufe erschwert die Kontrolle und Überwachung der Agenten, sodass eine sorgfältige Rechtfertigung ihrer Entscheidungen erforderlich ist. Trotz der mehrstufigen Lösungsansätze müssen die Agenten transparent in ihren Entscheidungen sein, da eine mangelnde Nachvollziehbarkeit zu Misstrauen führen kann. Risiken wie Halluzinationen, falsche Informationen und Manipulation durch unerwartete Eingaben müssen aktiv angegangen werden, um das Vertrauen der Nutzenden zu gewinnen.

Fairness spielt ebenfalls eine wesentliche Rolle, denn sie äußert sich darin, dass alle betroffenen Personen gerecht behandelt werden und keine Diskriminierung stattfindet. Dies erfordert eine kontinuierliche Überprüfung der Algorithmen und zugrundeliegenden Trainingsdaten. Die Herausforderung liegt darin, Fairness in allen Komponenten und Tools zu gewährleisten und die Nachverfolgbarkeit der endgültigen Entscheidungen zu sichern.

Insgesamt ist eine ausgewogene Betrachtung dieser Dimensionen unerlässlich, um die Herausforderungen der modernen Patientenversorgung erfolgreich zu meistern und die Integrität sowie Akzeptanz von KI-Agenten zu sichern.

4.3. Digitale Bildung des Personals und der Patienten

Die erfolgreiche Integration eines KI-Agenten in den klinischen Alltag, sei es bei administrativen Prozessen oder auch auf den Stationen am »Point of Care«, hängt maßgeblich von der Akzeptanz der Fachkräfte und der Patienten ab. Vereinzelt steht medizinisches Personal automatisierten Systemen skeptisch gegenüber, da eine Einschränkung in der eigenen Entscheidungsgewalt befürchtet oder gar eine zusätzliche Belastung anstatt einer Entlastung [4] wahrgenommen wird.

Auch eine ethisch motivierte skeptische Haltung ist häufig zu beobachten. Um einen echten Mehrwert für alle Beteiligten zu schaffen, müssen KI-Agenten in die bestehenden Prozesse integriert und spezifisch darauf abgestimmt werden. Sie dürfen nicht verkomplizieren, sondern sollen effizient unterstützen und Arbeitsabläufe rationalisieren. Die benutzerfreundliche Gestaltung der Schnittstellen ist essenziell, um die Akzeptanz zu fördern und eine intuitive Nutzung zu ermöglichen [5].

Beim Einsatz von KI ist zudem durchdachtes Change-Management inklusive einer umfassenden Schulung sowohl von Fachpersonal als auch Patienten erforderlich. Die Nutzenden müssen ein grundlegendes Verständnis für die Funktionsweise und vor allem auch für die Grenzen der Anwendung entwickeln, um die präsentierten Unterstützungsleistungen auch kritisch bewerten zu können.

Eine kontinuierliche Weiterbildung im Umgang mit dieser Technologie ist unabdingbar, um deren Nutzung vollumfänglich auszuschöpfen und die Fehlnutzung und potenzielles Versagen der Technologie bewerten zu können. Transparenz über den Einsatz von KI schafft dabei zusätzliches Vertrauen.

5. Fazit und Ausblick

Die Herausforderungen im Gesundheitswesen erfordern eine dringende Integration innovativer Technologien wie KI, um die Effizienz und Qualität der Patientenversorgung zu verbessern. Während LLMs wertvolle Werkzeuge zur Verarbeitung und Generierung von Text bieten, sind KI-Agenten notwendig, um komplexe Aufgaben zu bewältigen, die dynamische Entscheidungen und Echtzeit-Reaktionen erfordern.

Zukünftige Entwicklungen sollten sich auf die Schaffung umfassender Agentensysteme konzentrieren, die verschiedene KI-Technologien integrieren, um eine proaktive Patientenversorgung zu ermöglichen. Die Implementierung solcher Systeme wird jedoch auch von der Berücksichtigung ethischer Fragestellungen, Datenschutz und der Akzeptanz der Nutzenden abhängen.

Der Fokus sollte aus diesem Grund darauf liegen, benutzerfreundliche Schnittstellen zu schaffen und die digitale Bildung von Personal und Patienten zu fördern, um eine erfolgreiche Implementierung zu gewährleisten. Die Zusammenarbeit zwischen Technologieanbietern, Gesundheitsdienstleistern und anderen Stakeholdern wird entscheidend sein, um die Potenziale von KI im Gesundheitswesen vollständig auszuschöpfen und die Herausforderungen der Zukunft zu meistern.

Um die Chancen der digitalen Transformation im Gesundheitswesen voll auszuschöpfen, müssen sich Einrichtungen an gemeinsamen, interdisziplinären Projekten beteiligen. Die Implementierung von KI-Systemen ist keine ferne Zukunftsvision mehr, sondern jetzt dringend nötig, um die Versorgungseffizienz, das Patientenwohl aber auch die Wirtschaftlichkeit von Krankenhäusern und Kliniken zu verbessern. Es bedarf jedoch mehr als nur technologischer Innovation – es erfordert einen iterativen und interdisziplinären Ansatz, der durch Experimentierfreude und Mut geprägt ist.

Ein KI-Implementierungsprojekt im Krankenhaus folgt typischerweise einem strategischen strukturierten Ablauf:

1. **Vision entwickeln**
   Gemeinsam mit Stakeholdern aus Medizin, IT, Verwaltung und Datenschutz wird eine Vision entwickelt und eine Zielsetzung definiert.
2. **Anforderungen und Voraussetzungen klären**
   Gemeinsam werden regulatorische Rahmenbedingungen geklärt und technische Anforderungen analysiert.
3. **Daten sichten und evaluieren**
   Verfügbare Datenquellen werden geprüft, hinsichtlich Qualität, Vollständigkeit und Datenschutzanforderungen bewertet und für das Training von KI-Modellen aufbereitet.
4. **KI-Werkzeuge auswählen und anpassen**
   Basierend auf den Anforderungen werden geeignete KI-Lösungen identifiziert, angepasst und mit den relevanten Daten validiert.
5. **Prototypen entwickeln und evaluieren**
   In einem iterativen Prozess werden Prototypen erstellt, in klinischen Testumgebungen evaluiert und mit Nutzerfeedback optimiert.
6. **Schulungs- und Akzeptanzstrategien entwickeln**
   Fortbildungsprogramme und Aufklärungskampagnen werden entwickelt und eine nutzerfreundliche Schnittstelle sichergestellt.
7. **Langfristigen Betrieb klären**
   Für den produktiven Einsatz werden Aspekte wie Integration in bestehende Systeme, Wartung, Skalierbarkeit und kontinuierliche Validierung geregelt.

Durch enge Zusammenarbeit, kontinuierliches Lernen und die Bereitschaft, neue Wege zu gehen, können die Herausforderungen der Zukunft gemeistert und nachhaltige Verbesserungen im Gesundheitswesen erzielt werden.

6. Referenzen

[1] »Mitten im demografischen Wandel«, Statistisches Bundesamt. Zugegriffen am 24. Feb. 2025. [Online]. Verfügbar über: https://www.destatis.de/DE/Themen/Querschnitt/Demografischer-Wandel/demografie-mitten-im-wandel.html

[2] »Statistics | Eurostat.« Zugegriffen am 05. März 2025. [Online]. Verfügbar über: https://ec.europa.eu/eurostat/databrowser/view/HLTH_SHA11_HC__custom_15646460/default/table?lang=de

[3] »Bis 2049 werden voraussichtlich mindestens 280 000 zusätzliche Pflegekräfte benötigt«, Statistisches Bundesamt. Zugegriffen am 05. März 2025. [Online]. Verfügbar über: https://www.destatis.de/DE/Presse/Pressemitteilungen/2024/01/PD24_033_23_12.html

[4] J. Qiu et al., »LLM-based agentic systems in medicine and healthcare«, Nat Mach Intell, vol. 6, no. 12, pp. 1418–1420, Dec. 2024, doi: 10.1038/s42256-024-00944-1.

[5] M. Klumpp et al., »Artificial Intelligence for Hospital Health Care: Application Cases and Answers to Challenges in European Hospitals«, Healthcare, vol. 9, no. 8, Art. no. 8, Aug. 2021, doi: 10.3390/healthcare9080961.

[6] J. Li et al., »Agent Hospital: A Simulacrum of Hospital with Evolvable Medical Agents«, Jan. 17, 2025, arXiv: arXiv:2405.02957. doi: 10.48550/arXiv.2405.02957.

[7] P. Hager et al., »Evaluation and mitigation of the limitations of large language models in clinical decision-making«, Nat Med, vol. 30, no. 9, pp. 2613–2622, Sep. 2024, doi: 10.1038/s41591-024-03097-1.

[8] A. Soroush et al., »Large Language Models Are Poor Medical Coders – Benchmarking of Medical Code Querying«, NEJM AI, vol. 1, no. 5, p. AIdbp2300040, Apr. 2024, doi: 10.1056/AIdbp2300040.

[9] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance). 2024. Zugegriffen am 06. März 2025. [Online]. Verfügbar über: http://data.europa.eu/eli/reg/2024/1689/oj/eng

[10] »Einfluss von KI auf die Cyberbedrohungslandschaft.« [Online]. Verfügbar über: https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/KI/Einfluss_KI_auf_Cyberbedrohungslage.pdf?__blob=publicationFile&v=2

[11] M. Mock et al., »Vertrauenswürdige KI-Anwendungen mit Foundation-Modellen entwickeln«, Fraunhofer IAIS Whitepaper. [Online]. Verfügbar über: https://doi.org/10.24406/publica-2475

Impressum

Herausgeber
Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS
Schloss Birlinghoven 1
53757 Sankt Augustin

Redaktion
Silke Loh
Eléna Zay-Vanvoorden

Grafik und Layout
Achim Kapusta
Lea Sophie Scharmach Silva

Bildnachweise
Cover: © ipopba - stock.adobe.com
Abbildung 1-3: Fraunhofer IAIS

Stand
März 2025

Kontakt

Dario Antweiler
Abteilung Knowledge Discovery
Telefon +49 2241 14-2516
dario.antweiler@iais.fraunhofer.de

Fraunhofer-Institut für Intelligente
Analyse- und Informationssysteme IAIS
Schloss Birlinghoven 1
53757 Sankt Augustin

www.iais.fraunhofer.de