



Datenmarktplätze für Künstliche Intelligenz im Gesundheitswesen: Potenziale, Herausforderungen und Strategien zur Bewältigung

Richard Guse · Scott Thiebes · Phil Hennel ·
Christoph Rosenkranz · Ali Sunyaev

Eingegangen: 15. Juli 2022 / Angenommen: 11. Oktober 2022 / Online publiziert: 7. November 2022
© Der/die Autor(en) 2022

Zusammenfassung Das Training von Künstliche Intelligenz (KI)-Modellen, die auf maschinellem Lernen (ML) beruhen, erfordert eine große Menge qualitativ hochwertiger Daten. Besonders im Gesundheitswesen mit seinen hochsensiblen Daten und hohen Anforderungen an den Datenschutz besitzen einzelne Akteur:innen oft jedoch nicht ausreichend hochwertige Daten. Datenmarktplätze für KI zielen darauf ab, dieses Problem zu lösen, indem sie Datenanbieter und Datenkonsumenten miteinander verbinden und den Handel von Daten ermöglichen. Allerdings haben sich Datenmarktplätze im Gesundheitswesen, trotz erster technischer Konzepte und einiger Pilotprojekte, bisher noch nicht erfolgreich durchsetzen können. Im Rahmen der vorliegenden Studie wurden daher Interviews mit einer Reihe von relevanten Expert:innen und Akteur:innen durchgeführt, um Potenziale, Herausforderungen und mögliche Strategien zur Bewältigung der Herausforderungen zu identifizieren. Die Ergebnisse der Studie verdeutlichen anhand der drei Dimensionen des Technology-

✉ Richard Guse · Scott Thiebes · Ali Sunyaev

Institut für Angewandte Informatik und Formale Beschreibungsverfahren (AIFB), Fakultät für
Wirtschaftswissenschaften, Karlsruher Institut für Technologie, Karlsruhe, Deutschland
E-Mail: richard.guse@kit.edu

Scott Thiebes

E-Mail: scott.thiebes@kit.edu

Ali Sunyaev

E-Mail: sunyaev@kit.edu

Phil Hennel · Christoph Rosenkranz

Kölner Institut für Wirtschaftsinformatik (KIWI), Wirtschafts- und Sozialwissenschaftliche Fakultät,
Universität zu Köln, Köln, Deutschland

Phil Hennel

E-Mail: hennel@wiso.uni-koeln.de

Christoph Rosenkranz

E-Mail: rosenkranz@wiso.uni-koeln.de

Organization-Environment-Frameworks spezifische Potenziale von Datenmarktplätzen für KI im Gesundheitswesen, aber gleichzeitig auch eine Reihe von Herausforderungen, die es zu adressieren gilt. Die erarbeiteten Bewältigungsstrategien liefern hierbei erste Ansätze zur Beseitigung der identifizierten Herausforderungen, zeigen jedoch auch die Notwendigkeit der weiteren Forschung auf diesem Gebiet auf.

Schlüsselwörter Digitalisierung · Gesundheitswesen · Künstliche Intelligenz · Datenmarktplätze · Herausforderungen

Data Marketplaces for Artificial Intelligence in Healthcare: Opportunities, Challenges and Mitigation Strategies

Abstract Training artificial intelligence (AI) models requires a large amount of high-quality data. However, especially in healthcare with its highly sensitive data and high privacy requirements, individual stakeholders often do not own sufficient high-quality data. Data marketplaces for AI aim to solve this problem by connecting data providers and data consumers and enabling data trading. However, despite initial technical concepts and some pilot projects, data marketplaces have not yet been successful in the healthcare sector. Within this study, expert interviews were therefore conducted with a number of relevant experts and stakeholders to identify potentials, challenges and possible strategies for overcoming the challenges. Based on the three dimensions of the technology, organization and environment framework, the results of the study highlight specific potentials of data marketplaces for AI in healthcare, but at the same time also a number of challenges that need to be addressed. The mitigation strategies developed here provide initial approaches for eliminating the challenges identified, but also highlight the need for further research in this area.

Keywords Digitalization · Healthcare · Artificial intelligence · Data markets · Challenges

1 Einleitung

Die Digitalisierung erfasst immer mehr Bereiche des alltäglichen Lebens, wie auch das Gesundheitswesen. Dabei weist die Digitalisierung im Gesundheitswesen viele unterschiedliche Facetten auf. Insbesondere die Entwicklung und der Einsatz von Künstlicher Intelligenz (KI) ist ein aktueller Trend, der eine bessere, effizientere und kostensparende Versorgung der Bevölkerung verspricht (Chew und Achananuparp 2022). Bspw. zeigen Studien, dass KI-Modelle Krankheiten auf Röntgenaufnahmen genauso gut oder sogar mit einer höheren Genauigkeit als medizinische Fachkräfte erkennen können (Joy Mathew et al. 2020).

Die meisten KI-Modelle, die heutzutage entwickelt und eingesetzt werden, basieren auf den KI-Methoden des maschinellen Lernens (ML) oder des Deep Learnings (DL). Sie beruhen auf einem abstrakten dreiphasigen Prozess, der die Schritte Modellierung, Training und Inferenz umfasst (Thiebes et al. 2021). Innerhalb die-

ses Prozesses kommt den Eingabedaten eine besonders wichtige Rolle zu, da die Verfügbarkeit einer großen Menge qualitativ hochwertiger Eingabedaten eine fundamentale Voraussetzung für die Entwicklung von qualitativ hochwertigen KI-Modellen ist. Besonders innerhalb des Gesundheitswesens stellt dies die Entwickler:innen von KI-Modellen vor große Herausforderungen. Denn obwohl heutzutage große Mengen von (Gesundheits-)Daten generiert werden, bilden Gesundheitssysteme üblicherweise komplexe Netzwerke aus einer Vielzahl unterschiedlicher Akteur:innen (bspw. Patient:innen, Gesundheitsdienstleister:innen, Versicherungen, Gesetzgebende), bei denen jeweils oft nur ein Bruchteil der generierten Daten anfällt. Gleichzeitig können Gesundheitsdaten nicht beliebig zwischen diesen Akteur:innen ausgetauscht werden, da sie zu den hochsensiblen und besonders schützenswerten Daten zählen und ihr Austausch und ihre Verwendung stark reglementiert sind.

Innerhalb der Literatur werden verschiedene Ansätze zur Begegnung des Mangels an Trainingsdaten für KI-Modelle im Gesundheitswesen diskutiert. Hierzu zählen insbesondere Datenmarktplätze und Ansätze der sicheren Mehrparteienberechnung (secure multi party computation (SMPC)) (Zhao et al. 2019). Da SMPC jedoch hohen Abstimmungsaufwand zwischen den einzelnen Parteien erfordert, keinen Überblick über weitere verfügbare Datenquellen liefert und keine Anreize für Individuen bietet, ihre Daten zu teilen, stellen Datenmarktplätze weiterhin ein wichtiges Konzept dar, um diese Probleme zu beheben (Ghaffaripour und Miri 2020). Im Gesundheitswesen haben sich, trotz erster technischer Konzepte und einiger Pilotprojekte (z. B. Nebula Genomics¹ und EncrypGen² für genetische Daten) Datenmarktplätze bisher noch nicht in der Breite etablieren können. Bisherige Forschung hierzu ist primär technisch, konzeptionell und verbunden mit Blockchain-Technologie (Maher und Khan 2016). Sie ist somit nicht geeignet zur Beantwortung der Frage, weshalb sich Datenmarktplätze für KI im Gesundheitswesen bisher nicht etablieren konnten. Mit dieser Studie möchten wir einen Beitrag dazu leisten Potenziale und aktuelle Herausforderungen bei der Realisierung von Datenmarktplätzen für KI im Gesundheitswesen besser zu verstehen. Das Hauptziel der Arbeit besteht demnach in der Beantwortung der Frage: *Welche Potenziale und Herausforderungen bietet die Realisierung von Datenmarktplätzen für KI im Gesundheitswesen?*

2 Künstliche Intelligenz im Gesundheitswesen

Innerhalb der Literatur existieren vielfältige Verständnisse des Begriffs KI. Im Allgemeinen kann unterschieden werden zwischen schwacher und starker KI (Searle 1980). Schwache KI kann nur einen Teil der menschlichen Intelligenz abbilden und wird für die Problemlösung von spezifischen Aufgaben verwendet. Bisher entspricht jede verfügbare KI dieser Art von KI. Starke KI kann im Gegensatz zur schwachen KI jedes Problem lösen. In dieser Studie verstehen wir KI im Sinne einer schwachen KI als Computer, die komplexe Entscheidungsprobleme lösen, indem Ergebnisse auf Basis von Erfahrungswerten erlernt werden (Goodfellow et al. 2016). Die Verwen-

¹ <https://nebula.org/whole-genome-sequencing-dna-test/>.

² <https://encrypgen.com/>.

derung von KI ist somit ein Prozess, der aktuelle Technologien (z. B. ML) nutzt, um Entscheidungsprobleme zu lösen. ML umfasst dabei definierte Algorithmen, die eine Mustererkennung in Daten ermöglichen. Ein Beispiel für ein solches Problem im Gesundheitswesen ist die Entscheidung, ob auf einem medizinischen Bild (bspw. generiert durch bildgebende Verfahren wie eine Röntgenaufnahme) ein Tumor vorliegt. DL ist ein Teil von ML, nutzt aber mehrere aufeinander aufbauende Datenschichten für die Mustererkennung. Die Entscheidungsfindung bei KI findet im Allgemeinen in drei Schritten statt: (1) Der KI werden Daten übergeben, (2) die KI verarbeitet und interpretiert die Daten mithilfe eines vortrainierten Modells und (3) gibt eine Entscheidung aus (Russell und Norvig 2010). Um ein vortrainiertes Modell zu erstellen, ist es dabei erforderlich, dass die Zieldaten in ausreichender Menge vorhanden, möglichst korrekt und repräsentativ sind.

Aktuelle Anwendungen von KI im Gesundheitssektor umfassen unter anderem KI zur Diagnose von Erkrankungen anhand von Daten aus bildgebenden Verfahren, zur Genanalyse oder die Steuerung von Operationsrobotern (Joy Mathew et al. 2020). Dabei übertrifft KI bei der Genauigkeit der Diagnose zunehmend medizinische Expert:innen. Der Einsatz von KI im Gesundheitswesen hat insgesamt viele potenzielle Vorteile. Die Unterstützung, die KI in vielen Bereichen Mitarbeiter:innen im Gesundheitssektor bietet, entlastet Mitarbeiter:innen, führt zu einer höheren Behandlungsqualität und spart durch eine bessere Planung und Priorisierung von Patient:innen Kosten ein (Masmoudi et al. 2021). Dennoch bringt der Einsatz von KI auch Risiken mit sich. KI kann z. B. bei unbekanntem Verzerrungen in Datensätzen gegenüber bestimmten Gruppen diskriminieren oder bei Fehlern zu Problemen bei Haftungsfragen führen (Russell und Norvig 2010).

3 Datenmarktplätze für Künstliche Intelligenz

Moderne Ansätze zur Entwicklung von KI-Modellen, wie ML oder DL, erfordern in der Regel große Mengen hochwertiger Eingabedaten für das Training der Modelle. Diese sind aber häufig aus Datenschutzgründen und aufgrund der starken Verteilung der Daten auf viele Akteur:innen im Gesundheitswesen für viele Organisationen nicht einfach zu beschaffen. Datenmarktplätze zielen darauf ab, dieses Problem zu lösen, indem sie Datenanbieter und Datenkonsumenten miteinander verbinden und den Akteur:innen den sicheren Handel, die Speicherung und den Zugriff auf hochwertige Datenbestände ermöglichen (Bergman et al. 2022). Dabei erfüllen Datenmarktplätze drei wesentliche Funktionen: (1) Zusammenbringen von Angebot und Nachfrage; (2) Bereitstellen einer Infrastruktur für die Ausgestaltung des Datenaustauschs; (3) Erleichtern von Transaktionen für die Vergütung und Übertragung von Datenbeständen (Bergman et al. 2022; vgl. Abb. 1).

Obschon Datenmarktplätze in diesen drei Funktionen traditionellen elektronischen Marktplätzen für Produkte und Dienstleistung ähneln, unterscheiden sie sich von traditionellen Marktplätzen dadurch, dass Daten ein Erfahrungsgut sind (d. h. ein Gut dessen Qualität und Wert nur schwer zu beurteilen sind), welches beliebig vervielfältigt werden kann (Koutroumpis et al. 2020). Bisherige Lösungen am Markt

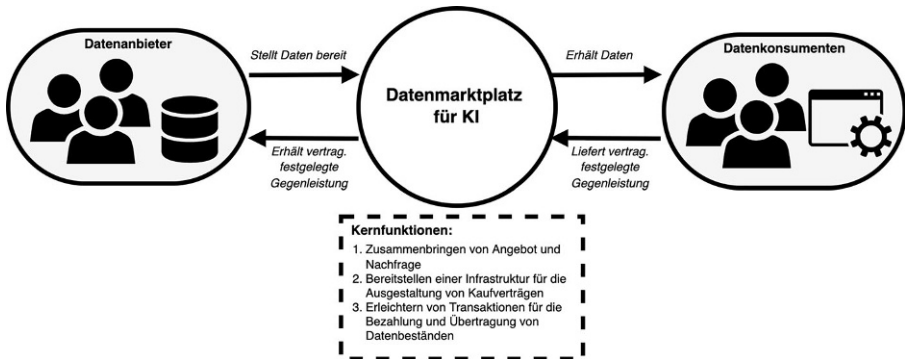


Abb. 1 Schematische Darstellung eines Datenmarktplatzes

werden meist durch nationale (z. B. USA³) oder supranationale (z. B. EU⁴) Initiativen oder einzelne private Anbieter (z. B. kaggle⁵) angeboten. Diese beschränken sich häufig auf ausgewählte aggregierte Datensätze und liefern keine Möglichkeit individuelle Datensätze zu teilen.

Eine der ersten Studien, die das Phänomen Datenmarktplätze konzeptualisiert hat, unterscheidet Datenmarktplätze anhand der beiden Dimensionen Ausrichtung (Engl. *Orientation*) und Eigentümerschaft (Engl. *Ownership*) (Stahl et al. 2016). Die Ausrichtung eines Datenmarktplatzes kann hierbei von hierarchisch (d. h., die Eigentümer:innen eines Datenmarktplatzes legen Datenpreise und die Regeln für Käufer:innen und Verkäufer:innen fest) bis hin zu marktbezogen (d. h., die Preise werden von Käufer:innen und Verkäufer:innen in Abhängigkeit vom Wettbewerb festgelegt) reichen (Bergman et al. 2022). Bezogen auf die Eigentümerschaft können Datenmarktplätze entweder im Privatbesitz einer Organisation sein, sich im Besitz eines Konsortiums (d. h., mehrerer Organisationen) befinden oder im Besitz einer unabhängigen dritten Partei sein. Koutroumpis et al. (2020) hingegen unterscheiden vier Arten von Datenmarktplätzen anhand unterschiedlicher Matching-Algorithmen. In einem One-to-One Datenmarktplatz kann ein Datenanbieter gleichzeitig mit einem oder mehreren Datenkonsumenten handeln. In einem One-to-Many Datenmarktplatz handelt ein Datenanbieter mit vielen Datenkonsumenten über dieselben Daten. Im Unterschied zu einem One-to-One Datenmarktplatz gelten jedoch in der Regel standardisierte Transaktionsbedingungen, da es unwirtschaftlich sein kann, jede Transaktion einzeln auszuhandeln. An Many-to-One-Datenmarktplätzen sind viele Datenanbieter, aber nur ein einziger großer Datenkonsument beteiligt. Many-to-Many-Datenmarktplätze schließlich sind Handelsplattformen, auf denen eine große Zahl von Datenanbietern für eine große Anzahl von Datenkonsumenten bereitstellt. Der Zugang zu den Daten und ihre Nutzung sind hierbei durch unterschiedliche, entweder standardisierte oder ausgehandelte Lizenzierungsmodelle geregelt.

³ <https://healthdata.gov/>.

⁴ <https://data.europa.eu/en>.

⁵ <https://www.kaggle.com/>.

4 Forschungsmethodisches Vorgehen

Obwohl Datenmarktplätze Gegenstand aktueller Forschung sind, ist der theoretische Erkenntnisstand zu Datenmarktplätzen im Gesundheitswesen nicht gesichert. Daher folgt unsere Forschungsarbeit einem qualitativen Vorgehen, mit welchem explorativ ein Überblick über die Wahrnehmung von Datenmarktplätzen im Gesundheitswesen erstellt wird. Hierfür werden Expert:inneninterviews mit verschiedenen Akteur:innen durchgeführt. Das methodische Vorgehen zur Datenerhebung ist dabei die Durchführung von semi-strukturierten Interviews, die sich an Myers (2019) orientieren.

4.1 Akquise von Interviewpartner:innen

Als Interviewpartner:innen wurden verschiedene Expert:innen akquiriert, die an Datenmarktplätzen teilhaben können. Die Expert:innen gliedern sich in drei Zielgruppen. Die erste Gruppe umfasst wissenschaftliche Mitarbeitende und Professor:innen, welche sich mit KI und Datenmarktplätzen im Gesundheitswesen in ihrer Forschung beschäftigen. Die zweite Gruppe umfasst Start-Ups und etablierte Organisationen, die sich mit der Entwicklung von KI und Datenmarktplätzen im Gesundheitswesen beschäftigen. Die dritte Gruppe umfasst Gesundheitsämter und Kliniken, die sich direkt mit der Bereitstellung von Daten beschäftigen. Die Interviewpartner:innen wurden durch Kaltakquise und direkte Kontakte rekrutiert. Für die Studie konnten so sieben Expert:innen gewonnen werden. Tab. 1 gibt einen anonymisierten Überblick über die Expert:innen.

4.2 Durchführung der Interviews

Die Interviews dauerten durchschnittlich ca. 45 min und hatten das Ziel, die Potenziale und Herausforderungen von Datenmarktplätzen im Gesundheitswesen zusammenzutragen. Dazu wurde erfragt, welche Strategien es gibt, um die Herausforderungen zu adressieren und Datenmarktplätze im Gesundheitswesen zu realisieren. Für die Durchführung der Interviews wurde ein Interviewleitfaden mit vorwiegend offenen Fragen entwickelt. Um die unterschiedlichen Perspektiven und Kontexte der Personen einzuordnen, umfasste der Interviewleitfaden ebenfalls Fragen zur Person und zu deren Organisation sowie zu grundlegenden Begriffen wie KI und Datenmarktplätzen. Für die weitere Analyse wurden die Interviews mit Zustimmung der Teilnehmer:innen aufgezeichnet und transkribiert.

4.3 Analyse der Interviewdaten

Zur Analyse der Interviewdaten wurde ein mehrstufiges, an Corbin und Strauss (2015) angelehntes Kodierungsverfahren eingesetzt. Die Ergebnisse aus den Interviews mit einer längeren Dauer und Interviewpartner:innen mit mehr Erfahrung mit Datenmarktplätzen und Gesundheitsdaten wurden dabei stärker berücksichtigt. Um die Subjektivität des Kodierungsprozesses zu verringern, wurde das Kodieren von zwei Forscher:innen unabhängig voneinander durchgeführt. Nach der Kodierung der ersten beiden Interviews (i1, i2) wurden die Codes verglichen und gemeinsam

Tab. 1 Zusammenfassung der Beschreibung der Interviewpartner:innen

ID	Organisation	Position	Jahre in der Org.	Begründung für die Auswahl	Dauer (in Minuten)
I1	Start-Up für die Entwicklung von KI für die Medizin	Geschäftsführung	1	Entwickeln KI für die Medizin Regel Austausch mit vielen Akteur:innen des Gesundheitswesens	60
I2	Mittelständisches Unternehmen für Langzeitarchivierungslösungen und einem Datenmarktplatz für KI	Business Developer	3	Unterstützen Entwicklung von KI auf Basis der archivierten Gesundheitsdaten Regel Austausch mit vielen Akteur:innen des Gesundheitswesens	85
I3	Gesundheitsamt	Arzt/ Medizininformatiker	8	Haben mehrere wissenschaftliche Artikel über Blockchain und Datenmarktplätze geschrieben	25
I4	Universität/ Forschungseinrichtung	Professor für Rechtswissenschaften	20	Analysieren KI und Datenmarktplätze aus rechtswissenschaftlicher Perspektive	40
I5	Forschungsinstitut für angewandte Wissenschaften (KI)	Projektleiter	3	Entwicklung von KI für verschiedene Branchen (u. a. Gesundheitswesen) Analyse von Möglichkeiten der Datenakquise für KI	55
I6	Universität	Wissenschaftlicher2 Mitarbeiter		Entwickeln KI-basierte Operationsrobotik Zugang zu Gesundheitsdaten bereits erworben	48
I7	Universitätsklinikum/ Universität	Professor für Radiologie	20	Entwickeln KI zur Bestimmung des Knochenalters auf Basis von Gesundheitsdaten in Benutzung	32

diskutiert, um ein gemeinsames Verständnis der zugrunde liegenden Konzepte zu erzeugen. Die verbliebenen fünf Interviews wurden separat kodiert und anschließend mit dem gesamten Autorenteam diskutiert. Für die Analyse wurde zunächst offenes Kodieren mithilfe des Softwarepakets „f4analyse 3.3“, dann axiales Kodieren und schließlich selektives Kodieren angewendet.

Beim *offenen Kodieren* werden Textstellen ohne bestimmte Vorgaben kodiert. Textstellen können dabei ebenfalls mit bereits erstellten Codes kodiert werden, wenn das zugrunde liegende Konzept ähnlich ist. Neue Codes werden laufend mit bereits erstellten Codes verglichen und bei ähnlichen zugrunde liegenden Konzepten zu einem neuen Code zusammengefasst. Die Wiederverwendung und Aggregation ermöglichten dabei eine robustere Identifikation von grundlegenden Konzepten. Auf diese Weise wurden insgesamt 13 Codes für Potenziale, 46 Codes für Herausforderungen und 34 Codes für Bewältigungsstrategien generiert.

Auf Basis der kodierten Textstellen wurden anschließend durch *axiales Kodieren* Zusammenhänge zu weiteren Textstellen hergestellt. Damit wurden die 34 Strategien zur Bewältigung der Herausforderungen den 46 Herausforderungen zugeordnet.

Schließlich wurden durch *selektive Kodierung* die bereits identifizierten Konzepte (Potenziale, Herausforderungen und Strategien) zu Oberkategorien aggregiert und in einen hierarchischen Zusammenhang gebracht. Dadurch ließen sich insgesamt 7 Oberkategorien für Potenziale, 12 Oberkategorien für Herausforderungen und 18 Oberkategorien für Strategien zur Bewältigung von Herausforderungen von Datenmarktplätzen identifizieren.

Im letzten Schritt der Analyse wurden die Oberkategorien für Potenziale und Herausforderungen anhand der drei Dimensionen Technologie, Organisation und Umgebung des Technology-Organization-Environment-Frameworks (TOE) gemeinsam diskutiert und in das Framework eingeordnet (Tornatzky und Fleischer 1990). Das Framework beschreibt, wie diese Dimensionen den Prozess beeinflussen, mit dem die Adoption und Implementierung von neuen Technologien in Organisationen stattfindet. Zusammengefasst stellen diese Dimensionen innerhalb der Theorie den Organisationskontext dar, in welchem die Organisation arbeitet. Der *technologische Kontext* umfasst dabei alle Aspekte der IT, die innerhalb und außerhalb der Organisation verfügbar sind. Der *organisatorische Kontext* beinhaltet dabei Aspekte der Struktur, der Mitarbeiter:innen, der internen Kommunikation, der verfügbaren Ressourcen und der Größe. Der *umgebungsbezogene Kontext* umfasst Aspekte der Industrie, wie Wettbewerber oder Arbeitsmarkt, und die Gesetzgebung.

5 Potenziale von Datenmarktplätzen im Gesundheitswesen

Insgesamt konnten im Rahmen der Studie sieben Potenziale von Datenmarktplätzen für KI im Gesundheitswesen identifiziert werden (vgl. Tab. 2). Daneben wurde darauf aufbauend und ergänzend das umfassendere Meta-Potenzial *Höhere Datenverfügbarkeit* identifiziert. Dieses Meta-Potenzial lässt sich nicht eindeutig einer Dimension des TOE-Frameworks zuordnen, sondern ist ein grundlegendes Potenzial von Datenmarktplätzen, welches als eine Voraussetzung für diese drei Dimensionen zu betrachten ist. Im Folgenden werden die sieben aggregierten Potenziale von Datenmarktplätzen für KI im Gesundheitswesen vorgestellt.

Datenmarktplätze bieten nicht nur mehr Daten, sondern können auch hochwertigere Daten liefern, die mehr unterschiedliche Szenarien und Umstände (z. B. bei der Behandlung von Krankheiten) abdecken als aktuell verfügbare Datenbestände. Dadurch bieten Datenmarktplätze das technologische Potenzial, bessere KI-Modelle und Ergebnisse zu erzeugen (P01) und eine bessere Abdeckung von Minoritäten (P02) zu erreichen. Damit verbunden sind organisationale Potenziale wie die Erschließung neuer Anwendungsgebiete für KI (P05), da der Datenmarktplatz es ermöglicht, unerforschte Krankheiten besser zu untersuchen. Für Akteur:innen im Gesundheitswesen bietet dies ebenso die Möglichkeit, ihre Ressourcen effizienter einzusetzen (P03), da ein Datensatz auf einem Datenmarktplatz genutzt und rekombiniert werden kann, anstatt die Daten selbst zu erheben. Damit verbunden ist ebenfalls eine Absenkung der Einstiegshürden für kleine und mittelständische Unter-

Tab. 2 Potenziale von Datenmarktplätzen für KI im Gesundheitswesen

ID	Potenzial	Beschreibung
P00	Höhere Datenverfügbarkeit	Datenmarktplätze führen zu einer höheren Verfügbarkeit von Daten für das Training von KI-Modellen im Gesundheitswesen
<i>Technologische Potenziale</i>		
P01	Bessere KI-Modelle und Ergebnisse	KI-Modelle und Ergebnisse werden robuster, können verfügbare Daten ergänzen und damit Verzerrungen überwinden
P02	Bessere Abdeckung von Minoritäten	Weniger Verzerrungen und Diskriminierung und größere Datenvielfalt
<i>Organisationale Potenziale</i>		
P03	Effizientere Ressourcenverwendung	Akteur:innen im Gesundheitswesen können verfügbare Ressourcen schneller und besser einsetzen
P04	Demokratisierung von Datenzugängen	Kleine und mittelständische Unternehmen erhalten Zugriff auf große Datenmengen
P05	Erschließung neuer Anwendungsgebiete für KI	Unerforschte Krankheiten und neue Anwendungsgebiete können untersucht werden
<i>Umgebungsbezogene Potenziale</i>		
P06	Etablierung von Standards und Schnittstellen	Standards werden etabliert und Schnittstellen werden vereinheitlicht
P07	Höheres Vertrauen in Digitalisierung	Durch positive Anwendungsbeispiele wird das Vertrauen in Digitalisierung gestärkt

nehmen (P04), die erst durch den Datenmarktplatz Zugriff auf große Datenmengen erhalten. Letztlich gibt es umgebungsbezogene Potenziale wie bspw. die Etablierung von Standards und einheitlichen Schnittstellen (P06) durch den Datenmarktplatzbetreiber selbst oder einen großen Datenanbieter. Dadurch wird der Zugriff auf die Daten weiter vereinfacht und durch die Vereinheitlichung lassen sich die Daten leichter miteinander kombinieren. Letztlich bieten Datenmarktplätze ebenso eine Möglichkeit, ein höheres Vertrauen in Digitalisierung (P07) zu schaffen. Dies kann durch Anwendungsszenarien (z.B. eine verbesserte Krebserkennung) geschehen, die erst durch Datenmarktplätze in der Breite ermöglicht werden. Eine ausführliche Zuordnung mit Beispielen aus den Interviews findet sich in Online-Material 1.

6 Herausforderungen bei der Realisierung von Datenmarktplätzen

Insgesamt konnten 12 Herausforderungen identifiziert werden (siehe Tab. 3). Entsprechend unserer Datenanalyse auf Basis des TOE-Frameworks erfolgt im Folgenden die Darstellung der Herausforderungen anhand der drei Dimensionen des Frameworks. Eine ausführliche Zuordnung mit Beispielen aus den Interviews findet sich in Online-Material 2.

Tab. 3 Herausforderungen und mögliche Bewältigungsstrategien

ID	Herausforderung	Beschreibung	Strategie zur Bewältigung
<i>Technologische Herausforderungen</i>			
H01	Technischer Datenschutz	Schutz gesundheitsbezogener Daten vor unbefugten Zugriffen im Einklang mit dem erhöhten Schutzbedürfnis gesundheitsbezogener Daten	Anwendung höchster, gängiger Datenschutz- und Sicherheitsstandards
H02	Zielkonflikt zwischen Anonymisierung und Datennutzbarkeit	Ein hohes Maß an Anonymisierung von Daten zum Zwecke des Datenschutzes kann insbes. im Gesundheitswesen die Nutzbarkeit dieser Daten für das Training von KI-Modellen beeinträchtigen	Lokales Training ermöglichen Metriken zur Bewertung von Datenqualität Nachvollziehbarkeit der Daten gewährleisten
H03	Sicherstellung von Datenqualität	Datensätze liegen oft nicht in ausreichender Qualität vor oder müssen erst aufwändig auf ein gewisses Qualitätsniveau gebracht werden	Umfangreiche Metadaten bereitstellen Breite Beteiligung der Bevölkerung
H04	Auffindbarkeit von Daten	Die Auffindbarkeit benötigter Datensätze ist innerhalb einer großen Menge heterogener Datensätze erschwert	
<i>Organisationale Herausforderungen</i>			
H05	Vorbehalte gegenüber Datenmarktplätzen	Vorbehalte relevanter Stakeholder bezogen auf den Datenschutz und die Nützlichkeit von Datenmarktplätzen mindern deren Akzeptanz	Relevante Stakeholder aufklären und frühzeitig in Entwicklung einbinden
H06	Komplexe Datenmarktplatz-Strukturen	Aufwendige Regelungen von Besitz- und Teilnahmestrukturen in Datenmarktplätzen erhöhen die Komplexität für die Teilnehmer:innen	Widerruf der Datennutzung ermöglichen Nachvollziehbarkeit der Datennutzung gewährleisten
H07	Unklarheit über Kosten-Nutzen-Relation	Die Teilnahme an Datenmarktplätzen ist für Organisationen mit gewissen Aufwänden verbunden, während der Nutzen unklar und vorab nur schwer abschätzbar ist	Freigabe durch die Ethikkommission
H08	Einhaltung ethischer Standards	Die Einhaltung bestimmter ethischer Standards in Bezug auf die Teilnahme an Datenmarktplätzen führt zu erhöhtem Aufwand für die Teilnehmer:innen	Lokales Training ermöglichen Aufträge zur Datensammlung ermöglichen
H09	Fehlende Zustimmung für Datennutzung	Bei vorhandenen Datensätzen mangelt es oft an der Zustimmung der Patient:innen zur Nutzung der Daten	
<i>Umgebungsbezogene Herausforderungen</i>			
H10	Unklare und inkonsistente Gesetzeslage	Unklarheit darüber, welche Gesetze wie im Kontext der Datenmarktplätze anzuwenden sind, insbesondere in Bezug auf den Datenschutz und inkonsistente Gesetzgebung	Entwicklung (und ggf. Beschaffung) von Datenstandards Verpflichtende Veröffentlichung von Daten aus öffentlich finanzierten Projekten
H11	Fehlende Datenstandards	Ein Mangel an Datenstandards erschwert die Nutzung von Daten über verschiedene Akteur:innen hinweg	Vergütung von Datenbeiträgen
H12	Fehlende Anreizmechanismen	Ein Mangel an Anreizmechanismen, die eine breite Beteiligung von mögl. Datenanbietern an Datenmarktplätzen sicherstellen	Vereinheitlichung und Vereinfachung der Gesetzgebung Breite Einwilligung ermöglichen

6.1 Technologische Herausforderungen

6.1.1 Technischer Datenschutz

Wie alle elektronischen Marktplätze bergen auch Datenmarktplätze die Gefahr, dass Unbefugte Zugriff auf geschützte Daten erhalten (z. B. durch unbeabsichtigte Datenlecks oder Hacker). Dies ist im Gesundheitswesen von hoher Brisanz, da Gesundheitsdaten in der Regel als besonders sensibel und schützenswert gelten. Eine durch die Expert:innen genannte Herausforderung besteht darin, einerseits einen aus technischer Perspektive möglichst hohen Datenschutz zu erreichen und andererseits die Verfügbarkeit der Daten für Akteur:innen mit berechtigten Interessen (d. h., die Datenkonsumenten) zu gewährleisten. Des Weiteren müssen eingesetzte Sicherheitsmechanismen regelmäßig auf ihr Schutzniveau überprüft werden. Ein:e Interviewpartner:in merkte diesbezüglich an: *„Diese Verschlüsselungsverfahren müssen regelmäßig überprüft werden, damit auch in fünf Jahren kein Rückschluss auf die Person möglich ist.“* (i4).

Eine mögliche Strategie zur Begegnung dieser Herausforderung liegt in der Anwendung und regelmäßigen Überprüfung höchster, gängiger Datenschutz- und Sicherheitsstandards.

6.1.2 Zielkonflikt zwischen Anonymisierung und Datennutzbarkeit

Eine weitere technologische Herausforderung liegt in dem oft vorherrschenden Zielkonflikt zwischen Anonymisierung und Datennutzbarkeit. Einerseits handelt es sich bei Gesundheitsdaten um spezifische, personenbezogene Daten, die anonymisiert werden müssen, bevor sie über Datenmarktplätze ausgetauscht werden können. Andererseits stellen sich Fragen nach einem ausreichenden Grad der Anonymisierung und der Nutzbarkeit der Daten. So merkt bspw. ein:e Interviewpartner:in an, dass etwa Röntgenaufnahmen so spezifische Krankheitsbilder wiedergeben können, dass sie nur schwer anonymisierbar sind: *„Also wenn ich an eine Röntgen-Thoraxaufnahme [...] denke, dann seh ich gerade gegebenenfalls zwei gebrochene Rippen und das können so eindeutige Erkennungsmuster sein, [dass] ich da kein anonymes Bild habe“* (i2).

Zwar lässt sich in der Regel auch heute schon eine sehr hohe Anonymisierung gesundheitsbezogener Datensätzen erreichen, bspw. durch das Löschen derartiger identifizierender Strukturen, jedoch steht dies in direktem Konflikt zur Nutzbarkeit der Daten, wenn etwa wichtige Strukturen gelöscht werden müssen, die für das Training eines bestimmten KI-Modells relevant sind.

Eine Strategie zur Bewältigung dieser Herausforderung besteht darin, anstatt direkt Datensätze auf Datenmarktplätzen zu handeln, mit dem Zugriff auf Datensätze zu handeln und KI-Modelle lokal zu trainieren oder nur trainierte Modelle auszutauschen.

6.1.3 Sicherstellung von Datenqualität

Die dritte Herausforderung auf technologischer Ebene liegt in der Sicherstellung der Qualität von auf Datenmarktplätzen gehandelten Daten. Gesundheitsdaten werden primär im Rahmen der Behandlung von Patient:innen erhoben und nicht zum Training von KI-Modellen. Folglich liegen sie oft in unterschiedlicher Qualität und in unterschiedlichen Formaten vor, sodass sie nicht unmittelbar für das Training von KI-Modellen zu verwenden sind. Bspw. müssen Daten aus bildgebenden Verfahren erst durch Expert:innen annotiert werden, um für das Training von KI-Modellen geeignet zu sein. Dieser Prozess ist aufwändig und fehleranfällig. Zudem ist es ein Problem, dass für Datenkonsumenten, ohne Wissen über die Rohdaten und deren Entstehungsprozess, nur schwer beurteilbar ist, von welcher Qualität die vorliegenden Daten sind: „*Ob ich dem Ganzen jetzt traue, ja, da müsste ich den ganzen Prozess jetzt praktisch nachvollziehen können*“ (i7).

Mögliche Strategien zur Bewältigung dieser Herausforderung liegen in der Entwicklung und Bereitstellung von Metriken zur Bewertung der Datenqualität sowie in der Gewährleistung der Nachvollziehbarkeit bereitgestellter Daten (z. B. durch lückenlose Versionierung).

6.1.4 Auffindbarkeit von Daten

Die letzte genannte Herausforderungen auf technologischer Ebene liegt in der Auffindbarkeit der Daten. Für das Training von KI-Modellen im Gesundheitswesen werden zumeist sehr spezifische Daten benötigt. Das Auffinden der richtigen Daten kann sich in Datenmarktplätzen als schwierig herausstellen, da Datensätze, auch zu einem spezifischen Krankheitsbild, in unterschiedlicher Form und Güte vorliegen. Gleichzeitig muss sichergestellt werden, dass die Auffindbarkeit von Daten nicht dazu genutzt werden kann, Datensätze beabsichtigt oder unbeabsichtigt zu de-anonymisieren: „*Optimalerweise ist ein Datenmarktplatz so groß, dass er trotz hoch spezifischer Informationen über diese Abfrage nicht einzelne Fälle auffindbar macht*“ (i1).

Um diese Herausforderung zu adressieren, ist es wichtig, umfangreiche und aussagekräftige Metadaten mit einzelnen Datensätzen bereitzustellen. Außerdem kann eine breite Beteiligung der Bevölkerung und die damit erhoffte Steigerung der verfügbaren Daten es erschweren, Rückschlüsse auf einzelne spezifische Fälle zu ziehen.

6.2 Organisatorische Herausforderungen

6.2.1 Vorbehalte gegenüber Datenmarktplätzen

Damit Datenmarktplätze genutzt werden, muss ein gewisses Grundvertrauen vorliegen. Dennoch gibt es aufgrund intransparenter Datenmarktplatzstrukturen und unklaren Akteur:inneninteressen Vorbehalte: „*Datenmarktplätze hören sich für mich momentan ein bisschen zu ökonomisch an. Das hört sich nicht sehr vertrauenswürdig an*“ (i4). Zudem kann es sein, dass sowohl bei der Nutzung als auch beim

Angebot von Daten ein Akteur:innenungleichgewicht entsteht. So haben vor allem Organisationen mit größeren finanziellen Mitteln mehr Möglichkeiten Daten zu nutzen, während Organisationen ohne ausreichende finanzielle Mittel unter Umständen nicht von Datenmarktplätzen profitieren können. Beim Angebot von Daten entsteht ein Ungleichgewicht dadurch, dass gerade im internationalen Vergleich sich besonders Staaten beteiligen können, die die Infrastruktur und Gelder für das Publizieren von Daten auf einem Datenmarktplatz bereitstellen können. Letztlich kann auch das Konzept und die Funktionsweise von Datenmarktplätzen für Akteur:innen des Gesundheitswesens zu abstrakt oder unbekannt sein.

Um die Vorbehalte gegenüber Datenmarktplätzen zu reduzieren, können zum einen bei der Freigabe von Daten alle relevanten Stakeholder über die Verwendungszwecke der Daten und die Funktionsweise des Datenmarktplatzes aufgeklärt werden. Zum anderen sollte den Akteur:innen ein Widerruf der Datennutzung (z. B. in Form einer dynamischen Einwilligung) ermöglicht werden.

6.2.2 Komplexe Datenmarktplatz-Strukturen

Datenmarktplätze für Gesundheitsdaten benötigen klare und einheitliche Strukturen. Bisher sind vor allem drei Aspekte der Datenmarktplatz-Struktur aufgrund ihrer Komplexität ungeklärt. Erstens stellt sich die Frage, wer die Regeln für den Marktplatz festlegt und welche Regeln für einen Datenmarktplatz gelten sollten. Beispielsweise ist bisher ungeklärt, wer an einem solchen Datenmarktplatz teilnehmen darf, wer Besitzer:innen des Datenmarktplatzes sind und inwieweit Besitzer:innen Regeln festlegen dürfen. Auf die Frage nach Besitzer:innen oder Treuhänder:innen für Datenmarktplätze merkte ein:e Experte:in an: „Wenn Sie in den aktuellen Entwurf der Europäischen Kommission zur KI-Verordnung schauen, dann kann man sich auch eine europäische Behörde vorstellen. Aber wie das letztendlich praxistauglich gestaltet werden kann, ist eine schwierige Frage“ (i4). Zweitens ist offen, ob die mit den Daten des Datenmarktplatzes erzeugten Ergebnisse denjenigen zur Verfügung gestellt werden sollen, die diese Daten freigegeben haben. Drittens ist die Frage offen, wem die Daten auf einem Datenmarktplatz gehören. Obwohl Patient:innen letztlich die Datenhoheit über ihre Daten haben müssen, bleibt zu klären, wer nach der Freigabe Besitzer:in der Daten ist und welche Rechte Besitzer:innen haben werden. Dies kann sich auch auf die mit den Daten trainierten KI-Modelle beziehen.

Zur Bewältigung dieser Herausforderung konnten keine Strategien identifiziert werden.

6.2.3 Unklarheit über Kosten-Nutzen-Relation

In mehrfacher Hinsicht kann das Bereitstellen von Daten auf einem Datenmarktplatz zu Kosten führen, ohne dabei einen direkten weiteren Nutzen zu erzeugen. Zunächst müssen Organisationen rechtliche Vorgaben und Standards analysieren und umsetzen: „Wenn man dann so ein Standard hat, bedeutet das erstmal Zusatz-Arbeit in Kliniken, also zum Beispiel Personal wird benötigt“ (i1). Des Weiteren müssen Organisationen technische Möglichkeiten schaffen, diese Daten zu aggregieren, die Zustimmung zu den einzelnen Daten nachzuhalten und auf dem Datenmarkt-

platz zur Verfügung zu stellen. Dabei findet weder eine Vergütung des Mehraufwandes statt, noch ergibt sich ein zusätzlicher Nutzen für die Organisation. Aus Patient:innensicht ist ebenfalls unklar, ob die Bereitstellung der eigenen Gesundheitsdaten einen Mehrwert bringt. Dies liegt vor allem an mangelnder Transparenz bezüglich der Verwendungszwecke bei Einwilligungserklärungen oder mangelnden Partizipationsmöglichkeiten an einer späteren Monetarisierung.

Bezüglich dieser Herausforderung ließen sich keine Strategien zur Bewältigung identifizieren.

6.2.4 *Einhaltung ethischer Standards*

Die Einhaltung ethischer Standards zu gewährleisten und somit Missbrauch der Daten vorzubeugen, ist eine erhebliche Herausforderung für Datenmarktplätze. Es ist bisher nicht ausreichend geklärt, für welche Zwecke die Freigabe von Gesundheitsdaten ethisch vertretbar ist. Zusätzlich lässt sich nur schwer nachvollziehen, wofür die Daten verwendet werden, sobald diese einmal freigegeben sind. Die Einhaltung ethischer Standards zu gewährleisten, erfordert hohen Aufwand, da die entsprechenden administrativen Hürden erst aufgebaut werden müssen.

Um diesen Herausforderungen zu begegnen, kann einerseits die Nachvollziehbarkeit der Datennutzung gewährleistet werden, indem jede Datennutzung auf einem Datenmarktplatz dokumentiert wird. Andererseits kann über die ethische Vertretbarkeit der Datennutzung entschieden werden, wenn eine unabhängige Ethikkommission eingerichtet wird, die die Nutzung der Daten prüft und freigibt.

6.2.5 *Fehlende Zustimmung für Datennutzung*

In vielen Gesundheitseinrichtungen fallen schon heute große Mengen an Gesundheitsdaten an. Um diese Daten über Datenmarktplätze bereitstellen und austauschen zu können, bedarf es der expliziten Zustimmung der jeweiligen Patient:innen. Häufig fehlt eine derartige Zustimmung, da das Einholen einer umfangreichen Zustimmung mit administrativen Aufwänden verbunden ist. Zudem wurden viele der heute vorhandenen Datensätze zu einer Zeit erhoben, in der noch nicht an deren Verwendung für das Training von KI-Modellen oder deren Austausch über Datenmarktplätze gedacht wurde: „*Welcher Verarbeitung hat der Nutzer zugestimmt und welcher nicht? Das nachzuhalten ist schwierig. [...] Viele Daten sind in den letzten Jahrzehnten entstanden, aber da gibt es noch keine Zustimmung*“ (i2). Die letzte Herausforderung auf der organisationalen Ebene besteht demnach in der fehlenden Zustimmung von Patient:innen für die Nutzung von Gesundheitsdaten in Datenmarktplätzen.

Eine Strategie zur Adressierung dieser Herausforderung ist das lokale Training von KI-Modellen, damit Daten nicht die Gesundheitseinrichtung verlassen müssen, in der sie generiert wurden, sodass das Einholen von Zustimmungen vereinfacht werden kann. Des Weiteren wurde auch die Vergabe von Aufträgen für die gezielte Sammlung von Daten für Datenmarktplätze genannt.

6.3 Umgebungsbezogene Herausforderungen

6.3.1 Unklare und inkonsistente Gesetzeslage

Da Gesundheitsdaten einem hohen Datenschutz unterliegen, ist es erforderlich, dass eindeutig geregelt wird, welche Daten auf Datenmarktplätzen gehandelt werden dürfen, was mit diesen Daten erlaubt ist und in welchem Kontext sie genutzt werden dürfen. Dennoch ist momentan unklar, welche dieser Aspekte bspw. von den bestehenden Gesetzen und Regularien wie der Datenschutzgrundverordnung (DSGVO) aufgefangen werden: „*Momentan haben wir nur die DSGVO und die sagt zu Datenmarktplätzen gar nichts*“ (i3). Auch wenn in Zukunft die Einwilligung für die Nutzung von Patient:innenendaten eingeholt werden kann, bleibt weiterhin offen, wie mit bisherigen Patient:innenendaten umgegangen werden darf. Darüber hinaus ist die Regelung dieser Aspekte bspw. in den Landeskrankenhausgesetzen über verschiedene Bundesländer hinweg inkonsistent, was eine breite Anwendung von Datenmarktplätzen erschwert. Diese Herausforderung verschärft sich noch weiter, wenn Datenmarktplätze über Ländergrenzen hinweg etabliert werden sollen.

Strategien, diese Herausforderung zu überwinden, stützen sich auf zwei Handlungsempfehlungen: Erstens die Vereinheitlichung und Vereinfachung der Gesetzgebung in Bezug auf Datenmarktplätze und zweitens das Ermöglichen von breiter Einwilligung, um den Prozess der Einwilligungen zur Datennutzung von Patient:innen zu erleichtern.

6.3.2 Fehlende Datenstandards

Um eine möglichst reibungslose Nutzung der Daten auf einem Datenmarktplatz zu ermöglichen, sind Standards zur Speicherung, Annotation und Integration von Daten unabdinglich (bspw. vereinheitlichte Schemata): „*Die Vergleichbarkeit und Standardisierung sowohl innerhalb der Daten als auch innerhalb der Forschungsdokumente ist für mich die größte Challenge*“ (i1). Ohne Standards lassen sich die Daten nicht oder nur mit erheblichem Aufwand miteinander kombinieren, da keine einheitliche Interpretation durch die KI möglich ist. Dazu zählt bspw. auch die uneinheitliche Annotation von medizinischen Bildern. Während z.B. in einem Datensatz medizinische Bilder genaue Markierungen und Bezeichnungen enthalten, liegen in einem weiteren Datensatz keine Markierungen oder andere Bezeichnungen vor. Diese Markierungen und Bezeichnungen bei sehr großen Datensätzen manuell nachzutragen, würde viel Aufwand erfordern.

Eine mögliche Strategie zur Adressierung dieser Herausforderung ist die Entwicklung und Einführung von Datenstandards, die in Datenmarktplätzen eingehalten werden müssen.

6.3.3 Fehlende Anreizmechanismen

Es muss sichergestellt werden, dass adäquate Anreizmechanismen etabliert werden, die die Akteur:innen im Gesundheitswesen (Patient:innen, Ärzt:innen und Organisationen) dazu motivieren, ihre Daten über einen Datenmarktplatz zur Verfügung

zu stellen und Standards einzuhalten. Da Daten einen sehr hohen Wert für Organisationen haben, die mit diesen Daten Geld verdienen, haben Organisationen ohne entsprechende Anreize kein Interesse daran, ihre Daten offenzulegen oder mit Mehraufwand in einen bestimmten Datenstandard zu konvertieren. Dennoch sollte die reine Monetarisierung gerade auf Patient:innenseite nicht als einziger Anreizmechanismus genutzt werden: „Vergütung finde ich ganz schlecht. Man möchte ja keinen Ausverkauf der Daten forcieren. Es geht auch immer darum, die Freiwilligkeit sicherzustellen.“ (i3). Unklar ist dabei auch, ob neben Geld bspw. eine Vergünstigung der Datennutzung, eine freie Nutzung der Ergebnisse oder andere Anreizmechanismen implementiert werden sollten. Auch bei Daten, die in Projekten erhoben und generiert werden, die mit öffentlichen Mitteln finanziert sind, sind bisher keine Anreize integriert, um die Daten in einem bestimmten Standard der Öffentlichkeit zugänglich zu machen.

Um diese Herausforderung zu überwinden, können zwei Strategien angewendet werden: Einerseits können zunächst Vergütungen von Datenbeiträgen implementiert werden, um einen Anreiz zu schaffen. Andererseits könnten die Veröffentlichung und Standardisierung von Daten aus öffentlich finanzierten Projekten verpflichtend in Ausschreibungen von öffentlichen Projekten übernommen und die Verwendung bestimmter Standards vorgeschrieben werden.

7 Fazit

Mit fortschreitender Digitalisierung werden KI-Anwendungen immer relevanter im Gesundheitswesen. Die Erstellung von KI-Anwendungen erfordert jedoch große Datenmengen, die im Gesundheitswesen nicht immer einfach zu erhalten sind. Als Lösung für dieses Problem werden vermehrt Datenmarktplätze diskutiert, die aber bisher in der Praxis kaum realisiert sind. Vor diesem Hintergrund leistet diese Studie einen Beitrag dazu, das Potenzial und aktuelle Herausforderungen bei der Realisierung von Datenmarktplätzen für KI im Gesundheitswesen besser zu verstehen.

Mit dieser Studie konnten wir sieben Potenziale in drei Dimensionen sowie ein Meta-Potenzial identifizieren. So können Datenmarktplätze durch erhöhte Datenverfügbarkeit zu besseren KI-Modellen und Ergebnissen führen, die Behandlung von seltenen Krankheiten unterstützen und Kosten im Gesundheitswesen reduzieren. Dazu identifizieren wir 12 Herausforderungen und mögliche Strategien zur Bewältigung in drei Dimensionen. Während ein Teil der Herausforderungen, wie die Sicherstellung von Datenqualität und die Auffindbarkeit von Daten, allgemeine Probleme von Datenmarktplätzen abbilden, sind andere Herausforderungen wie die Einhaltung ethischer Standards insbesondere im Gesundheitskontext relevant. Zukünftige Forschung sollte daher vor allem untersuchen, wie die identifizierten Strategien zur Bewältigung der Herausforderungen implementiert werden können. Zudem hat sich gezeigt, dass einige Aspekte von Datenmarktplätzen, gleichzeitig als Potenzial und Herausforderung angesehen werden. So wurde bspw. das Fehlen von Datenstandards als Herausforderungen gesehen, während dies gleichzeitig auch als Potenzial zur Etablierung von (neuen) Standards und Schnittstellen gewertet wurde.

Im Rahmen der verwendeten, explorativen Forschungsmethode ergeben sich für diese Studie einige Limitationen. Es wurde nur eine begrenzte Anzahl von Interviews geführt und somit möglicherweise nur eine beschränkte Sicht auf Datenmarktplätze ohne Data Scientists, Patient:innen-Vertretungen und mit nur wenig medizinischem Fachpersonal abgebildet. Zukünftige Forschung sollte insbesondere diese Perspektiven stärker untersuchen. Auch wurden für die Interviews deutsche Expert:innen befragt, die vor allem Einblicke in das deutsche Gesundheitswesen geben. Weitere Forschung sollte bspw. in Bezug auf grenzüberschreitende Datenflüsse, auch innerhalb der EU und mehr Interviews mit Expert:innen außerhalb Deutschlands beinhalten. In den Expert:inneninterviews wurden außerdem für zwei organisationale Herausforderungen keine Strategien zur Bewältigung identifiziert, da sie mit Fragen verbunden sind, die erst mit Forschung anhand eines implementierten Datenmarktplatz für Gesundheitsdaten beantwortet werden können.

Trotz dieser Einschränkungen sind wir davon überzeugt, dass diese Studie und die darin enthaltenen Ergebnisse einen wichtigen Beitrag zur Realisierung von Datenmarktplätzen und damit zur breiten Anwendung von KI im Gesundheitswesen leisten.

Zusatzmaterial online Zusätzliche Informationen sind in der Online-Version dieses Artikels (<https://doi.org/10.1365/s40702-022-00922-4>) enthalten.

Funding Gefördert von der Deutschen Forschungsgemeinschaft (DFG) – 442171588.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

Literatur

- Bergman R, Abbas AE, Jung S et al (2022) Business model archetypes for data marketplaces in the automotive industry: Contrasting business models of data marketplaces with varying ownership and orientation structures. *Electron Markets*. <https://doi.org/10.1007/s12525-022-00547-x>
- Chew HSJ, Achananuparp P (2022) Perceptions and needs of artificial intelligence in health care to increase adoption: scoping review. *J Med Internet Res* 24:1–19. <https://doi.org/10.2196/32939>
- Corbin JM, Strauss AL (2015) *Basics of qualitative research: techniques and procedures for developing grounded theory*, 4. Aufl. SAGE, Los Angeles
- Ghaffaripour S, Miri A (2020) Enabling medical research through privacy-preserving data markets. In: Moallem A (Hrsg) *HCI for cybersecurity, privacy and trust*. Springer, Cham, S 367–380
- Goodfellow I, Bengio Y, Courville A (2016) *Deep learning*. MIT Press, Cambridge

- Koutroumpis P, Leiponen A, Thomas LDW (2020) Markets for data. *Ind Corp Change* 29:645–660. <https://doi.org/10.1093/icc/dtaa002>
- Maher M, Khan I (2022) From sharing to selling: challenges and opportunities of establishing a digital health data marketplace using blockchain technologies. *Blockchain Healthc Today*. <https://doi.org/10.30953/bhty.v5.184>
- Masmoudi M, Jarboui B, Siarry P (Hrsg) (2021) Artificial intelligence and data mining in healthcare, 1. Aufl. Springer, Cham
- Mathew JC, David AM, Mathew JCM (2020) Artificial intelligence and its future potential in lung cancer screening. *EXCLI J* 19:1552–1562. <https://doi.org/10.17179/EXCLI2020-3095>
- Myers MD (2019) *Qualitative research in business and management*, 3. Aufl. SAGE, Thousand Oaks
- Russell SJ, Norvig P (2010) *Artificial intelligence: a modern approach*, 3. Aufl. Prentice Hall, Upper Saddle River
- Searle JR (1980) Minds, brains, and programs. *Behav Brain Sci* 3:417–424. <https://doi.org/10.1017/S0140525X00005756>
- Stahl F, Schomm F, Vossen G, Vomfell L (2016) A classification framework for data marketplaces. *Vietnam J Comput Sci* 3:137–143. <https://doi.org/10.1007/s40595-016-0064-2>
- Thiebes S, Lins S, Sunyaev A (2021) Trustworthy artificial intelligence. *Electron Markets* 31:447–464. <https://doi.org/10.1007/s12525-020-00441-4>
- Tornatzky LG, Fleischer M (1990) *The processes of technological innovation*. Lexington Books, Lexington
- Zhao C, Zhao S, Zhao M et al (2019) Secure multi-party computation: theory, practice and applications. *Inf Sci* 476:357–372. <https://doi.org/10.1016/j.ins.2018.10.024>