

Ergänzende Zitate zu Herausforderungen und Bewältigungsstrategien

ID	Herausforderung	Beschreibung	Beispiele für unterstützende, stichhaltige Zitate für Herausforderungen aus den Interviews	Strategien zur Bewältigung der Herausforderung	Beispiele für unterstützende, stichhaltige Zitate für Bewältigungsstrategien aus den Interviews
Technologische Herausforderungen					
H01	Technischer Datenschutz	Schutz gesundheitsbezogener Daten vor unbefugten Zugriffen im Einklang mit dem erhöhten Schutzbedürfnis gesundheitsbezogener Daten.	<p>“Da haben die Ärzte überhaupt keine Sorgen. Der Datenschutz ist für die Ärzte, also bei den Gesunden ist das ein Problem.” (i07)</p> <p>“Es darf kein Rückschluss auf die Person möglich sein. Wie man die Anonymisierung in der Praxis umsetzt, ist eine andere Frage. Das ist ein Problem der Verschlüsselung. Da wären die Informatiker gefragt. Diese Verschlüsselungsverfahren müssen regelmäßig überprüft werden damit auch in 5 Jahren kein Rückschluss auf die Person möglich ist” (i04)</p> <p>“Und wie können sehr große Datenmengen genutzt werden und dennoch eine Sicherheit gewährleistet werden?” (i05)</p> <p>“wichtig die Qualität der Daten sicherzustellen. Denn wenn</p>	<ul style="list-style-type: none"> • Anwendung höchster, gängiger Datenschutz- und Sicherheitsstandards • Lokales Training ermöglichen • Metriken zur Bewertung von Datenqualität • Nachvollziehbarkeit der Daten gewährleisten • Umfangreiche Metadaten bereitstellen • Breite Beteiligung der Bevölkerung 	<p>“Wenn es notwendig ist sie zu anonymisieren also alle personenbezogenen Daten entfernen sowohl aus Bilddaten als auch die ganzen Metadaten dieser Bilder” (i02)</p> <p>“Beteiligung möglichst großer Teile der Bevölkerung.” (i05)</p> <p>“Könnte ich dort dann sozusagen eine spezifischere, sensiblere Abfrage machen kann und die nicht mit Daten arbeitet, die auf dem Datenmarktplatz verfügbar gemacht werden müssen” (i02)</p> <p>“Und da wird aktuell sehr viel mehr investiert von Unternehmen in Richtung verteiltem Lernen, wo man dann eben die Daten gar nicht austauschen muss, sondern nur Modelle oder Schwarm Lernen, wo es nicht mal eine zentrale Instanz gibt,</p>

			Datenmarktplätze mit „Müll“, also unwichtigen Daten, überflutet werden, dann werden diese schnell uninteressant für Käufer, also die Leute, die KI trainieren wollen.” (i03)		sondern ein Peer-to-Peer, die Modelle ausgetauscht werden.” (i05)
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.	<p>“Was gilt als anonym und was nicht?” (i01)</p> <p>“Also eine vollständige Anonymisierung ist ganz oft nicht möglich, weil es bestimmte weitere Themen gibt. Also wenn ich an eine Röntgen Thorax Aufnahme, Aufnahme vom Brustkorb eines Menschen denke dann seh ich grad gegebenenfalls 2 gebrochene Rippen und das können so eindeutige Erkennungsmuster sein ist natürlich ich da kein Anonymes Bild habe” (i02)</p>		<p>“Dann eine aufmerksame Datenerfassung und am besten so was wie ein Fact-Sheet für jeden Datensatz erstellen.” (i05)</p> <p>“Also interaktiv und versioniert, dass man zu jederzeit auch zurückgehen kann in der Versionshistorie und schaut, was war denn meine ursprüngliche Annotation? Wie haben Sie sich über die Zeit verändert und warum?” (i06)</p> <p>“Also ich brauche einfach zusätzlichen Daten um die relevanten Bilddaten, Datensätze fürs Training zu identifizieren, und die muss ich mir quasi vorher raussuchen” (i02)</p>

<p>H03</p>	<p>Sicherstellung von Datenqualität</p>	<p>Datensätze liegen oft nicht in ausreichender Qualität vor oder müssen erst aufwändig auf ein gewisses Qualitätsniveau gebracht werden.</p>	<p>“Bei einer Vorhersage, die sehr gut für 40-jährige Männer passt, aber eine schlechtere Qualität bei jungen Frauen aufweist, ist die Frage, welches Vorgehen gerecht ist. Sollte dem Arzt das Ergebnis entweder gar nicht angezeigt werden, weil es ungerecht wäre, oder sollte das Ergebnis ausgegeben werden mit einer Warnung, dass Ergebnis von schlechter Qualität sein könnte?” (i01)</p> <p>“wichtig für gute Ergebnisse, dass man extrem ja, gute Krankheitsbilder auswählt.” (i02)</p> <p>“wichtig die Qualität der Daten sicherzustellen. Denn wenn Datenmarktplätze mit „Müll“, also unwichtigen Daten, überflutet werden, dann werden diese schnell uninteressant für Käufer, also die Leute, die KI trainieren wollen.” (i03)</p> <p>“Dann eine Riesenherausforderung ist die Datenqualität, die erhebliche Schwankungen haben wird, denn in einem Krankenhaus werden die Radiologie Bilder so angefertigt, im nächsten Krankenhaus anders und man kann die Leute nicht zwingen, sich an irgendeinem Thema zu halten,</p>		<p>“Dann eine aufmerksame Datenerfassung und am besten so was wie ein Fact-Sheet für jeden Datensatz erstellen.” (i05)</p>
------------	---	---	--	--	--

			<p>sondern mit denen Unterschieden muss man irgendwie leben.” (i05)</p> <p>“ebenso wie jetzt ein Dorfarzt höchstwahrscheinlich häufiger zu Fehldiagnosen kommt als ein Chefarzt in der Charité. Müsste auch irgendwie sichergestellt werden, dass die Daten, die hochgeladen werden, dementsprechend auf Qualität kontrolliert werden.” (i06)</p> <p>“Der Unterschied ist wahrscheinlich, wenn ich jetzt Daten für die Werbung verkaufe und ich kaufe den Werbung und stelle fest die Teddybären die ich verkaufe da verkaufe ich nicht mehr wenn ich diesen Algorithmus gekauft habe, mit dem Trainingsdatensatz, dann ist das jetzt nicht gut, aber sagen wir mal der Schaden ist begrenzt, weil andere Leute auch Teddybären anbieten. Ja, wenn ich aber sozusagen Diagnosen damit mache und die Leute sterben wie die Fliegen, weil der Datensatz nicht gut ist, dann ist das halt nicht akzeptabel.” (i07)</p>		
H04	Auffindbarkeit von Daten	Die Auffindbarkeit benötigter Datensätze ist innerhalb einer großen Menge heterogener Datensätze erschwert.	“gute Selektion über die Daten habe. Also ich muss einfach genau sagen können wir sozusagen was für Daten ich benötige” (i02)		
Organisationale Herausforderungen					

<p>H05</p>	<p>Vorbehalte gegenüber Datenmarktplätzen</p>	<p>Vorbehalte relevanter Stakeholder bezogen auf den Datenschutz und die Nützlichkeit von Datenmarktplätzen mindern deren Akzeptanz.</p>	<p>“Es ist sehr schädlich für so einen Prozess, wenn der Entwickler nicht verrät, wie das System funktioniert. Und genau das muss auch bei einem Daten Marktplatz beachtet werden. Es sollte nicht einfach nur ein abstraktes Konstrukt sein, welches irgendwo in einer vermeintlich sicheren Cloud liegt, sondern es muss genau erklärt sein, wo kommen die Daten rein und was passiert mit diesen.” (i01)</p> <p>“Und auch so was wie eine globale Fairness. Also eigentlich muss man dafür sorgen, dass eine internationale Beteiligung aller Nationen möglich ist. Ansonsten haben wir noch eine Verstärkung der Unterschiede” (i05)</p> <p>“Da muss man vielleicht nur an sich selbst denken. Also das ist eine Frage, die man nur aufgrund von empirischen Studien beantworten kann. Würde man selbst so eine Einwilligung unterzeichnen seine Daten zu Forschungszwecken wegzugeben. Da kommt es auf die Vertrauenswürdigkeit der Organisation drauf an zu der die Daten kommen und ob ich mich informiert fühle.” (i04)</p>	<ul style="list-style-type: none"> • Relevante Stakeholder aufklären und frühzeitig in Entwicklung einbinden • Widerruf der Datennutzung ermöglichen • Nachvollziehbarkeit der Datennutzung gewährleisten • Freigabe durch die Ethikkommission • Lokales Training ermöglichen • Aufträge zur Datensammlung ermöglichen 	<p>“Strategien sind Einbindung der Anwenderinnen und Anwender in die Entwicklung von KI-Systemen.” (i05)</p> <p>“Eine transparente Möglichkeit seine Zustimmung zu geben” (i05)</p> <p>“Würden Sie so Ihren optimalen Datenmarktplatz definieren? Also, dass die Interaktion nicht anonym ist, die Daten an sich anonym sind, dass es gut standardisiert ist und qualitativ sehr hochwertig ist!” (i06)</p> <p>“Ethikkommission, wie sie jedes Krankenhaus hat, auch als Datenmarktplatz betreiben, unabhängig von meinem kommerziellen Interesse und sagen eine Forschung muss sozusagen von mir erstmal als relevante Forschung validiert werden, die einen positiven Zweck hat und erst dann darfst du sozusagen Daten aus meinem Datenmarktplatz erwerben, wenn eine Prüfung erfolgt ist.” (i02)</p> <p>“in der Einrichtung und dort kann ich dann trainieren und das was dann sozusagen das Ergebnis für die Forscher ist oder was herausziehen ist nicht ein erworbener Datensatz bei</p>
------------	---	--	---	--	---

H06	Komplexe Datenmarktplatz-Strukturen	Aufwendige Regelungen von Besitz- und Teilnahmestrukturen in Datenmarktplätzen erhöhen die Komplexität für die Teilnehmer:innen.	<p>“Wer legt die Regeln für den Marktplatz fest? Wer darf daran teilnehmen und wer nicht?” (i01)</p> <p>“Das dritte Problem bezieht sich auf die Frage: “Wem gehört der Datenmarktplatz?” (i01)</p> <p>“Die Frage ist etwas allgemein. Die Frage nach der Regulierung ist die 100 Millionen Dollar Fragen. Für die Regulierung von KI gibt es da schon Vorschläge. Man könnte es bereichsspezifisch bei unterschiedlichen Behörden implementieren. Aber ich kann diese Frage pauschal nicht beantworten. Da müsste man wirklich ins Detail gehen. Gerade im medizinischen Bereich. Es betrifft nicht nur die medizinische Forschung, sondern einen großen Bereich.” (i04)</p> <p>Gerade bei medizinischen Forschungsdaten sollte nur von wenigen und vor allem von Ärzten ein Zugriff erfolgen.” (i04)</p>		<p>uns. Da kriegen sie nicht 10.000 Bild Aufnahmen, die sich dann kopieren und irgendwie bei sich verfügbar haben, sondern sie bekommen quasi Zugang zu diesen 10.000 Bild aufnahmen und können die auf dem System trainieren und entnehmen sozusagen am Ende die mathematischen Gewichte” (i02)</p>
H07	Unklarheit über Kosten-Nutzen-Relation	Die Teilnahme an Datenmarktplätzen ist für Organisationen mit gewissen Aufwänden verbunden, während der Nutzen unklar	<p>“Wenn man dann so ein Standard hat, bedeutet das erstmal Zusatz-Arbeit in Kliniken, also zum Beispiel Personal wird benötigt” (i01)</p> <p>“Also es scheint ja doch irgendwie ein Anreiz, oder die technischen und</p>		

		und vorab nur schwer abschätzbar ist.	rechtlichen Möglichkeiten, nicht da zu sein.” (i06)		
H08	Einhaltung ethischer Standards	Die Einhaltung bestimmter ethischer Standards in Bezug auf die Teilnahme an Datenmarktplätzen führt zu erhöhtem Aufwand für die Teilnehmer:innen.	<p>“Im medizinischen Bereich hat man nur sensible Daten. Ich kann das konkret nicht beantworten man muss sich das im Einzelfall anschauen wann das ethisch vertretbar ist:” (i04)</p> <p>“Dann bei Output Data ist eine Herausforderung/. KI-Systeme können unabsichtlich für nicht gedachte Einsatzzwecke verwendet werden. Wenn ich ein KI-System habe/. Ein Negativbeispiel, dass irgendjemand sagt, ob jemand rückfällig wird und ich setze das einen ganz anderen Ort ein, dann macht das überhaupt keinen Sinn, dieses System zu verwenden.” (i05)</p> <p>“ich finde, dass die Interaktion mit den Daten doch einem Experten zugeordnet werden müssen, weil sonst das klassische YouTube Kommentare Phänomen einschlägt, wo aus dieser Anonymität im Internet ergibt sich ja doch sehr häufig ein, ich blas mal einen Schwachsinn ins Internet, kann ja nicht auf mich zurückfallen. Aber ich hab das Gefühl, sobald man einen Kommentar oder einen Beitrag im Internet liefert, der dir persönlich auf eine Person zurückfallen kann, wird</p>		

			auch die Qualität dessen, was man dort beiträgt besser.” (i06)		
H09	Fehlende Zustimmung für Datennutzung	Bei vorhandenen Datensätzen mangelt es oft an der Zustimmung der Patient:innen zur Nutzung der Daten.	“Es geht darum, dass man im Sinne des Datenschutzrechts darauf achten muss, dass man die Einwilligung freiwillig gibt.” (i04)		
Umgebungsbezogene Herausforderungen					
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.	<p>“Also die brauchen vor allem erstmal Rechtssicherheit.” (i03)</p> <p>“Diese Frage wird gerade noch erforscht. Da ist man gerade in der Diskussion was von DSGVO aufgefangen wird.” (i04)</p> <p>“Dann ist eine Herausforderung, dass unterschiedliche Nationen unterschiedliche Normen und Werte haben, aber unsere Daten und unsere Modelle heute schon in dem Zug und noch sehr viel mehr durch die Globalisierung international existieren und agieren und letztendlich für trustworthiness. Wenn man da auf Normen und Werte basieren möchte und die unterschiedlich sind, dann wird es schwer, ein gemeinsames Bild dafür zu definieren, was die Anforderungen sind” (i05)</p> <p>“Sind die Leute nicht gewillt das aufzuziehen oder sind die Leute nicht</p>	<ul style="list-style-type: none"> • Entwicklung (und ggf. Beschaffung) von Datenstandards • Verpflichtende Veröffentlichung von Daten aus öffentlich finanzierten Projekten • Vergütung von Datenbeiträgen • Vereinheitlichung und Vereinfachung der Gesetzgebung • Breite Einwilligung ermöglichen 	<p>“Ja, also diese ganzen Daten Marktplatz Themen da geht es glaube ich am Ende einfach nur darum, dass man gute Wege finden muss, um mit solchen Daten DSGVO konform zu arbeiten und arbeiten zu können, und das heißt, wir müssen Standards definieren.” (i02)</p> <p>“Gleichzeitig muss man, glaube ich, auch ein bisschen an den Standards arbeiten, wie Daten Gelabelt werden also das medizinische Selbstverständnis” (i02)</p> <p>“Wenn mit öffentlichen Mitteln gezahlt wird und es sich um geförderte Projekte handelt, könnte verlangt werden, dass das Unternehmen bestimmte Standards einhalten muss. Also wenn öffentliche Ausschreibungen gemacht werden, muss sich auch an bestimmte Daten-Bereitstellung und Datenschutz-Regelungen gehalten werden.” (i01)</p>

			gewillt, die technischen und rechtlichen Hürden anzugehen?" (i06)		"Eine Vergütung von Datenbeiträgen von Personen" (i05)
H11	Fehlende Datenstandards	Ein Mangel an Datenstandards erschwert die Nutzung von Daten über verschiedene Akteur:innen hinweg.	"Gleichzeitig muss man, glaube ich, auch ein bisschen an den Standards arbeiten, wie Daten Gelabelt werden also das medizinische Selbstverständnis" (i02)		
H12	Fehlende Anreizmechanismen	Ein Mangel an Anreizmechanismen, die eine breite Beteiligung von mögl. Datenanbietern an Datenmarktplätzen sicherstellen.	<p>"Auf der anderen Seite muss beantwortet werden, wie groß der ‚Gain‘ ist – also was habe ich davon? Ich denke, es ist legitim zu sagen, dass derjenige der die Daten bereitstellt, vielleicht eine Vergünstigung in Richtung Datennutzung bekommt. Also dass die Person von den Ergebnissen profitiert, wenn Modelle entstanden sind. Dass ein Incentive da ist, die Daten zur Verfügung zu stellen." (i01)</p> <p>"Also eine Herausforderung ist die gleiche, dass Unternehmen momentan keinen Anreiz haben, ihre Daten zu teilen" (i05)</p>		