



Daten – der Treibstoff für künstliche Intelligenz (KI)

Daten als Grundlage für den Erfolg von KI- und datenbasierten Projekten



Gefördert durch:



Vernetzt durch:



1 Management Summary

Der vorliegende Leitfaden widmet sich detailliert den **Anforderungen an Daten** für ein erfolgreiches Gelingen von KI- und datenbasierten Projekten. Ziel ist es, ein **grundlegendes Verständnis** zu schaffen, das es auch Fachfremden ermöglicht, die Bedeutung und die Anforderungen von Daten an KI-Systeme zu erkennen und zu verstehen.

In unserer heutigen, schnelllebigen digitalen Welt sind Daten das neue Gold. Die Auseinandersetzung mit den eigenen Daten bringt viele **Mehrwerte**. So ermöglicht sie Unternehmen eine präzisere Entscheidungsfindung und optimierte Geschäftsprozesse zu entwickeln. Sie hilft bei der Identifikation von Markttrends und Kundenbedürfnissen und treibt die Entwicklung innovativer Produkte und Dienstleistungen voran. **Insbesondere für KI-Projekte** legt ein solides Datenmanagement die Grundlage, indem es qualitativ hochwertige und gut strukturierte Daten bereitstellt, die für das Training und die Implementierung erfolgreicher KI-Modelle unerlässlich sind und fortschrittliche Analyse überhaupt erst möglich machen.

Trotz ihrer zentralen Bedeutung werden Daten in vielen Unternehmen nicht mit der notwendigen Sorgfalt behandelt. Der Leitfaden soll Aufschluss dazu geben, welche Voraussetzungen an Daten es für die erfolgreiche Durchführung KI- und datenbasierter Projekte gibt. Diese **vier wesentlichen Anforderungen an Daten** werden im Leitfaden aufgezeigt und Handlungsempfehlungen gegeben:

- **Die Bedeutung von Datenqualität:** Sie ist oft unzureichend, was zu ineffizienten und fehleranfälligen KI-Systemen führen kann. Eine schlechte Datenqualität kann verschiedene Ursachen haben, darunter unvollständige Datensätze, Inkonsistenzen, veraltete Informationen und fehlerhafte Eingaben.
- **Die Notwendigkeit klar definierter Verfügbarkeiten und Verantwortlichkeiten:** In vielen Fällen sind die Verantwortlichkeiten für die Datenpflege und -sicherheit innerhalb eines Unternehmens nicht klar definiert. Dies führt zu Unsicherheiten darüber, wer Zugang zu den Daten hat, wer sie aktualisieren darf und wer für ihre Sicherheit verantwortlich ist. Außerdem liegen die Daten in vielen Unternehmen häufig isoliert in sogenannten Datensilos und werden nicht zentral verwaltet.
- **Die unterschiedlichen Anforderungen an Datenmenge:** Eine ausreichende Datenmenge ist Voraussetzung für eine Umsetzung von KI. Je nach Anwendungsfall und KI-Algorithmus kann dies jedoch stark variieren.
- **Die Berücksichtigung der Privatsphäre und rechtlicher Rahmenbedingungen:** Gesetzliche Vorgaben zur Datennutzung und Privatsphäre müssen stets eingehalten werden und sollten bei KI-Projekten frühestmöglich geprüft werden.

2 Anforderungen an Daten für KI

Praxisbeispiel

Ein mittelständisches Unternehmen im Einzelhandel möchte seine Lagerbestände optimieren und die Nachfrageprognose verbessern, um Überbestände und Fehlbestände zu minimieren. Hierzu plant das Unternehmen, ein KI-gestütztes System einzuführen, das auf historischen Verkaufsdaten, saisonalen Trends und externen Faktoren wie Wetterdaten basiert.

Das Ziel der Datennutzung in diesem KI-Projekt ist es, präzise Vorhersagen über zukünftige Verkaufszahlen zu treffen, um die Lagerbestände effizient zu verwalten. Durch die Analyse und Integration verschiedener Datenquellen sollen Muster und Korrelationen erkannt werden, die menschliche Analysen übersteigen. Dies führt zu einer verbesserten Bestandsplanung, reduzierten Lagerkosten und einer höheren Kundenzufriedenheit durch die Verfügbarkeit der richtigen Produkte zur richtigen Zeit.

Welche Datenanforderung muss das Unternehmen für ein die Umsetzung eines erfolgreichen KI-Projekts beachten?

Die Anforderungen an Daten für ein erfolgreiches Gelingen von KI- und datenbasierten Projekten sind vielfältig und können je nach Anwendungsfall und Datentyp variieren. Dennoch gibt es grundlegende Prinzipien, die stets beachtet werden sollten. Diese Anforderungen bilden ein solides Grundgerüst, das unabhängig vom spezifischen Kontext gilt. Zu diesen wesentlichen Anforderungen an Daten für KI-Projekte gehören, wie oben genannt, **Datenqualität, Verfügbarkeit und Verantwortlichkeiten, die Datenmenge, sowie Privatsphäre und rechtliche Aspekte.**

Diese Anforderungen werden im Folgenden detailliert erläutert, um ein umfassendes Verständnis dafür zu schaffen, wie Daten effektiv für KI- und datenbasierte Projekte genutzt werden können.

Grundsätzlich sind alle dieser Anforderungen für ein erfolgreiches Gelingen von datenbasierten und KI-Projekten wichtig. Die nachfolgende Grafik (*Abbildung 1*) zeigt jedoch, an welchen Anforderungen es in der Praxis hauptsächlich scheitert. Die Grafik basiert auf den Erkenntnissen aus einer Befragung der KI-Labs (Regionale Labore für Künstliche Intelligenz in Baden-Württemberg), die durch ihre Rolle als regionale Ansprechpartner für Unternehmen einen vielseitigen Einblick in diese und deren Herausforderungen haben. An der Befragung nahmen 12 Personen von unterschiedlichen KI-Labs teil.

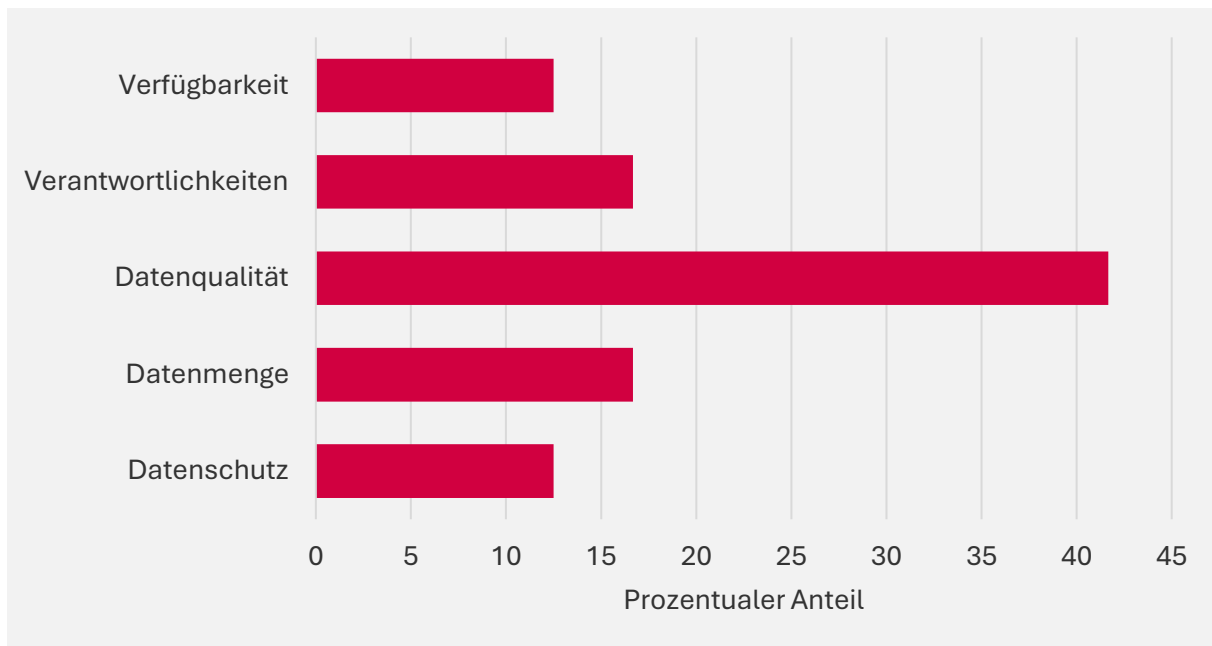


Abbildung 1: Häufigste Hindernisse bei der Umsetzung von datenbasierten und KI-Projekten als Prozentuale Anteil der Nennung der aufgeführten Anforderungen. Bis zu zwei Antwortmöglichkeiten waren zulässig.

2.1 Verfügbarkeit und Verantwortlichkeiten

Datensilos

Eine der größten Herausforderungen, denen Unternehmen heute gegenüberstehen, ist die Verfügbarkeit und Verwaltung von Daten. Daten sind oft in Silos gespeichert, das heißt, sie werden in verschiedenen Abteilungen und von unterschiedlichen Personen verwaltet und verantwortet und häufig in verschiedenen Systemen abgelegt. In vielen Unternehmen existieren Datensilos in Form von separaten Datenbanken, Excel-Tabellen oder speziellen Softwarelösungen, die nicht miteinander kommunizieren. Diese Isolation führt dazu, dass wertvolle Informationen nicht vollständig genutzt werden können. Dies erschwert einen schnellen, zuverlässigen Zugriff auf die Daten, vor allem wenn von verschiedenen Silos Daten benötigt werden und die Zuständigkeiten nicht klar definiert sind. Andererseits erschwert dies ein umfassendes Bild aller vorliegender Daten, sowie ihrer Menge und Qualität zu erhalten.

KI-Wissen von Mitarbeitenden

Zusätzlich ist es für Mitarbeitende ohne KI-Wissen meist unklar, welche Daten für KI-Projekte genutzt werden könnten und welche Anforderungen daran bestehen. Dies führt in der Praxis häufig dazu, dass entweder irrelevante Daten gesammelt und verarbeitet werden, was unnötige Kosten verursacht, oder dass Daten nicht mit der nötigen Sorgfalt und Qualität eingepflegt werden, was die Qualität und Leistungsfähigkeit der KI-Lösungen beeinträchtigt.

Klare Verantwortlichkeiten und Prozesse

Um diese Herausforderungen zu bewältigen, ist es unerlässlich, klare Verantwortlichkeiten und Prozesse zu etablieren. Dabei gibt es einige Maßnahmen, die dabei helfen diesen Herausforderungen entgegenzutreten. Etablierte Maßnahmen sind unter anderem die Zentralisierung der Datenverwaltung, eine Erstellung eines Datenkatalogs, eine klare Definition von Verantwortlichkeiten, Ansprechpersonen und klare Kommunikationswege, sowie die

Sensibilisierung von Mitarbeitenden durch Schulungen und Weiterbildungen. Unter dem Begriff **Data Governance** werden all diese Punkte zusammengefasst.

Durch die Implementierung einer Data Governance können Unternehmen die Verfügbarkeit und Verwaltung ihrer Daten erheblich verbessern. Dies schafft die Grundlage für erfolgreiche KI- und datenbasierte Projekte und ermöglicht es, das volle Potenzial der Daten auszuschöpfen.

Praxisbeispiel „Effizientes Lagerbestandsmanagement durch KI-Prognosen“

Um Datensilos zu vermeiden, sollte das Unternehmen dafür sorgen, dass alle relevanten Daten zentralisiert und leicht zugänglich sind. Das bedeutet, dass Verkaufsdaten, Lagerbestände und externe Faktoren wie Wetterdaten in einem zentralen Datenspeicher, z.B. einer Datenbank, gespeichert werden. Es sollte geprüft werden, dass die Mitarbeitenden das notwendige Wissen im Bereich KI haben, um die Daten korrekt zu nutzen und zu interpretieren – falls nicht müssten z. B. Schulungen in Betracht gezogen werden. Zudem sollten klare Verantwortlichkeiten und Prozesse definiert werden. Beispielsweise sollte ein Teammitglied verantwortlich dafür sein, dass aktuelle Verkaufsdaten regelmäßig in das System eingepflegt werden, während ein anderes Teammitglied die Datenqualität überwacht. Auch die Zugriffsrechte und die Entscheidungshoheit über die Daten müssen klar und transparent sein. Durch klare Prozesse und Zuständigkeiten kann sichergestellt werden, dass die Daten immer aktuell und korrekt sind und ein Zugriff auf die Daten einfach möglich ist.

2.2 Datenqualität

Neben Verfügbarkeit und Verantwortlichkeiten ist die Qualität der Daten eines der zentralen Kriterien und entscheidend über die erfolgreiche Umsetzung von datenbasierten Projekten und die Effektivität von KI-Systemen. Ein KI-System kann schon mit relativ kleinen Datenmengen von hoher Qualität gute bis sehr gute Ergebnisse liefern.

2.2.1 Ursachen für unzureichende Datenqualität

Für eine schlechte Datenqualität gibt es viele verschiedene Gründe. Dabei ist häufig eine fehlende Sensibilisierung von Mitarbeitenden und die niedrige Priorität von Daten im Unternehmen eine zentrale Herausforderung. Dies führt einerseits zu fehlerhaften oder unvollständigen Daten vor allem bei Dateneingabe durch den Menschen, da den Personen, die die Daten eingeben die Bedeutung und Auswirkungen nicht klar sind. Eine weitere häufige Ursache für schlechte Datenqualität ist, dass die Grundsätze der Datenspeicherung nicht beachtet werden. Zum Beispiel werden Teils Vor- und Nachnamen von Personen als ein Feld in einer Datenbank gespeichert.

In einigen Fällen müssen Daten für Anwendung von KI-Algorithmen ein Label besitzen, um das Training von KI-Modellen zu ermöglichen. Darunter versteht man, die Zielausgabe bzw. die Antwort, die das KI-System vorherzusagen versucht. Probleme können auftreten, wenn Label ungenau oder falsch sind oder wenn es Schwierigkeiten gibt, Experten für die Annotation der Daten zu finden.

Weitere häufige Gründe sind ungenaue Messungen durch unzuverlässige oder minderwertige Sensorik, fehlende Metadaten oder veraltete Daten.

2.2.2 Dimensionen der Datenqualität

Datenqualität wird häufig in die sechs Dimensionen Vollständigkeit, Validität, Genauigkeit, Aktualität, Einzigartigkeit und Konsistenz eingeteilt, die bereits 2013 definiert wurden¹.

Zusätzlich zu diesen sechs Dimensionen ist noch die Repräsentativität der Daten als siebte Dimension als wichtiges Kriterium für KI-Systeme zu nennen. Welche der unterschiedlichen Dimensionen der Datenqualität in der Praxis besonders häufig zu einem Scheitern von KI-Projekten führt zeigt *Abbildung 2*. Diese Grafik wurde durch die bereits erwähnte Befragung der KI-Labs durchgeführt. Zu sehen ist, dass insbesondere die schlechte Repräsentativität der Daten, sowie die Unvollständigkeit und teils auch die fehlende Aktualität der Daten die häufigsten Herausforderungen darstellen.

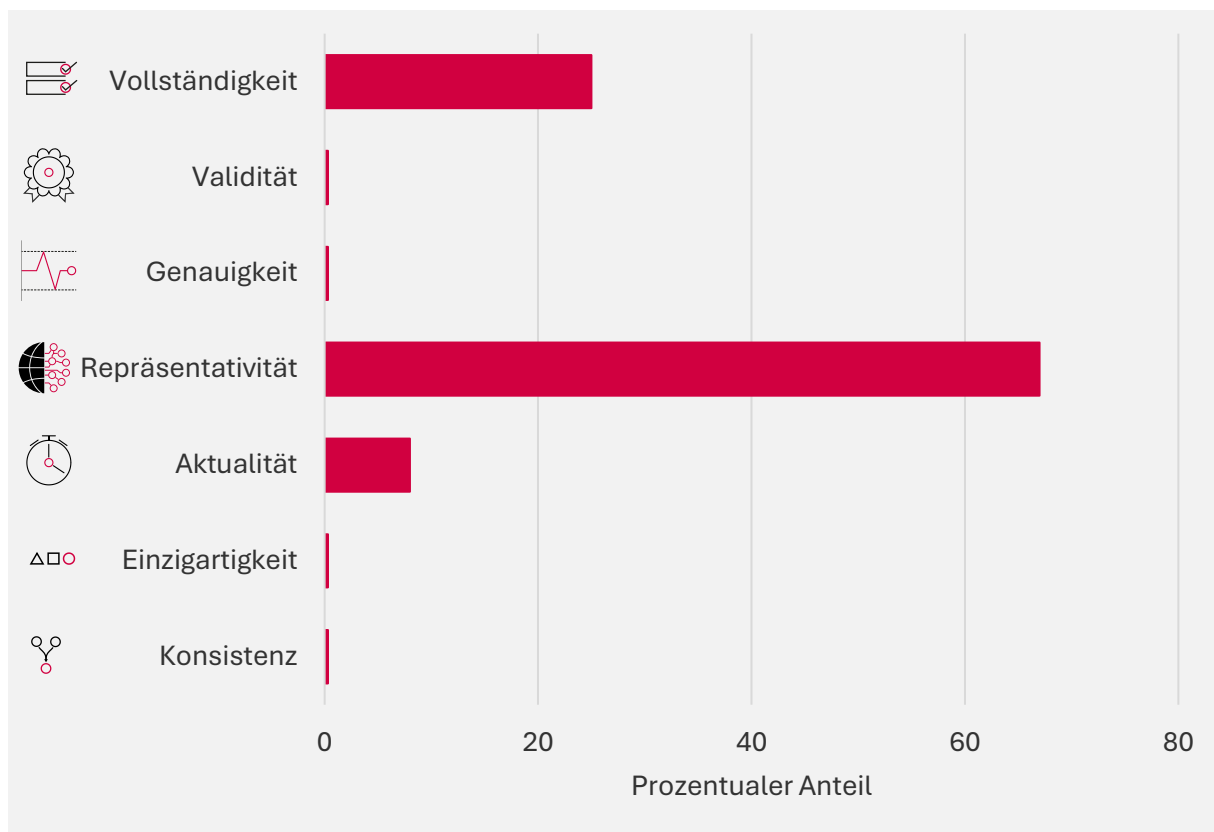


Abbildung 2: Dimensionen der Datenqualität, die am häufigste zu Problemen bei der Umsetzung von datenbasierten und KI-Projekten führen als prozentualer Anteil der Nennung der aufgeführten Dimensionen. Bis zu zwei Antwortmöglichkeiten waren zulässig.

Vollständigkeit

Daten werden als vollständig bezeichnet, wenn alle für einen Anwendungsfall erforderlichen Daten ohne Lücken oder fehlende Daten vorhanden sind. Es wird empfohlen, für jeden Anwendungsfall zu definieren, was Vollständigkeit bedeutet. In der Praxis wird eine Vollständigkeit von 100 % selten erreicht. Einzelne fehlende Datenpunkte können entweder aus dem Datensatz entfernt werden, oder durch geeignete Verfahren ersetzt werden, was in der Fachliteratur unter dem Begriff Imputation zusammengefasst wird.

¹ DAMA UK Working Group on "Data Quality Dimensions". „THE SIX PRIMARY DIMENSIONS FOR DATA QUALITY ASSESSMENT - Defining Data Quality Dimensions“. In: (2013). %\cite{DAMA_2013



Validität

Die Validität beschreibt, inwieweit die Daten in einem definierten Format vorliegen oder einem gültigen Wert entsprechen. Beispielsweise besteht eine Postleitzahl in Deutschland aus fünf Ziffern. Weicht ein Eintrag davon ab, ist er nicht valide. Außerdem werden die Postleitzahlen in Deutschland von der Deutschen Post AG vergeben. Sie können mit einer Liste aller existierenden Postleitzahlen verglichen werden. Beispielsweise ist die fünfstellige Nummer 01010 keine gültige Postleitzahl und damit nicht valide. Valide Daten ermöglichen den Austausch und die Nutzung über Systemgrenzen hinweg.



Genauigkeit

Die Genauigkeit beschreibt, wie genau oder korrekt die Daten die Realität widerspiegeln. Zum Beispiel, wie genau ein Thermometer die Temperatur misst. Ein Messfehler von 0,2°C kann je nach Anwendungsfall vernachlässigbar oder von entscheidender Bedeutung sein. Daher ist es wichtig, die Genauigkeit im Hinblick auf den Anwendungsfall zu prüfen und geeignete Maßnahmen zur Überwachung der Genauigkeit zu implementieren. Bei Daten, die von Menschen erzeugt und eingegeben werden, ist die Überprüfung der Genauigkeit besonders relevant. Da sich Daten und Konzepte im Laufe der Zeit ändern können, ist es wichtig, die Genauigkeit in regelmäßigen Abständen zu überprüfen.



Repräsentativität

Die Repräsentativität eines Datensatzes gibt an, inwieweit die Daten die Vielfalt und Variation sowie deren Verteilung im realen Anwendungsbereich widerspiegeln. Probleme können auftreten, wenn die Daten nicht alle relevanten Szenarien abdecken oder bestimmte Gruppen oder Merkmale unterrepräsentiert sind. In der Praxis führt dies zu einer Verzerrung des KI-Systems, einem sogenannten Bias.

Ein Bias kann durch technische Methoden ausgeglichen werden, setzt aber Kenntnisse über die Repräsentativität der Daten voraus.

Beispiel eines Bias behafteten KI-Systems

Ein KI-System zur Vorhersage von Gesundheitsrisiken von US-Bürger, zeigte einen Bias in Form von rassistischen Vorurteilen. Dies geschah, da es auf einem nicht repräsentativen Maß zur Bestimmung des Bedarfs basierte. Der Algorithmus bevorzugte daher weiße gegenüber schwarzen Personen^a.

^aObermeyer, Ziad, et al. "Dissecting racial bias in an algorithm used to manage the health of populations." *SScience* 366.6464 (2019): 447-453



Aktualität

In vielen Anwendungsfällen ist die Aktualität der Daten von entscheidender Bedeutung. Die erforderliche Aktualität der Daten hängt vom jeweiligen Anwendungsfall ab und sollte für jeden Anwendungsfall definiert werden. Generell gilt jedoch, dass ältere Daten an Repräsentativität verlieren und zum Teil auch die Richtigkeit der Daten aufgrund sich ändernder Rahmenbedingungen und Konzepte nicht mehr gewährleistet werden kann. Zudem wird der Zusammenhang zu aktuellen Fragestellungen oftmals geringer.



Konsistenz







Daten gelten als konsistent, wenn zwei oder mehr Werte in verschiedenen Datenbanken oder Systemen identisch oder widerspruchsfrei sind.

Inkonsistente Daten können sowohl bei der Verknüpfung mehrerer Datensätze als auch beim Training von KI-Algorithmen zu erheblichen Problemen führen. Inkonsistente Daten werden oft als solche erkannt, aber es ist oft schwierig, die Inkonsistenzen aufzulösen.

△□○ **Einzigartigkeit**

Die einzelnen Datenpunkte in einem Datensatz müssen einzigartig sein. Doppelte Datenhaltung führt häufig zu Problemen bei der Aktualität der einzelnen Einträge, da meist nur ein System gepflegt wird. Darüber hinaus führt das mehrfache Auftreten desselben Datenpunktes zu Problemen der Repräsentativität und kann beim Training von KI-Systemen zu so genannten Leakage führen. Leakage bezeichnet die Verwendung von Informationen beim Training eines KI-Systems, die zum Zeitpunkt der Vorhersage nicht verfügbar sind. Dies wiederum führt zu einem Überschätzen der Leistung des KI-Systems.

Dimensionen der Datenqualität als Fragen formuliert

-  **Vollständigkeit:** Sind alle erforderlichen Datenpunkte vorhanden und vollständig erfasst?
-  **Validität:** Entsprechen die Daten den definierten Formaten und Regeln?
-  **Genauigkeit:** Sind die Daten korrekt und frei von Fehlern?
-  **Repräsentativität:** Spiegeln die Daten die Vielfalt und Breite des realen Kontextes und Szenarien wieder?
-  **Aktualität:** Sind die Daten auf dem neuesten Stand und regelmäßig aktualisiert?
-  **Konsistenz:** Stimmen die Daten über verschiedene Systeme und Quellen hinweg überein?
- △□○ **Einzigartigkeit:** Sind die Daten frei von Duplikaten und eindeutig?

Durch die Berücksichtigung dieser sieben Dimensionen können Unternehmen sicherstellen, dass ihre Daten die notwendige Qualität aufweisen, um KI- und datenbasierte Projekte erfolgreich umzusetzen. Ein kontinuierliches Qualitätsmanagement ist dabei unerlässlich, um die Datenqualität dauerhaft auf einem hohen Niveau zu halten und den damit verbundenen Mehrwert für das Unternehmen zu maximieren. Maßnahmen wie Mitarbeiterschulungen und die Implementierung automatisierter Validierungstools sowie Korrekturprozesse zur Behebung von Qualitätsproblemen können dabei eine entscheidende Rolle spielen.

Insgesamt ist die Sicherung der Datenqualität ein kontinuierlicher Prozess, der ständige Aufmerksamkeit und Pflege erfordert. Nur so kann sichergestellt werden, dass die Daten verlässlich und nutzbar bleiben und somit bestmögliche Ergebnisse für KI- und datenbasierte Projekte erzielt werden können.

Praxisbeispiel „Effizientes Lagerbestandsmanagement durch KI-Prognosen“

In unserem Praxisbeispiel sollte das Unternehmen sicherstellen, dass die Verkaufsdaten vollständig, valide, genau, aktuell, konsistent und einzigartig sind. Das bedeutet, dass alle relevanten Datenpunkte vorhanden und korrekt sein müssen, um zuverlässige Prognosen zu ermöglichen. Beispielsweise müssen Verkaufszahlen, Lagerbestände und saisonale Schwankungen genau erfasst werden. Regelmäßige Aktualisierungen der Daten sind notwendig, um aktuelle Trends und Änderungen im Kaufverhalten zu berücksichtigen. Zudem sollten doppelte Einträge vermieden werden, um die Datenintegrität zu wahren.

2.3 Datenmenge

Die Datenmenge ist ein weiterer entscheidender Faktor für den Erfolg von KI-Projekten. Der tatsächliche Bedarf variiert jedoch stark je nach Komplexität des Anwendungsfalls und des verwendeten KI-Algorithmus.

Einfache vs. komplexe Anwendungen

- Einfache Anwendungsfälle: Bereits einige wenige, qualitativ hochwertige Datenpunkte können ausreichen, um ein brauchbares Modell zu erstellen.
- Komplexe Anwendungen: Bei Anwendungen wie Bild- oder Spracherkennung sind oft Tausende oder Millionen von Datenpunkten notwendig, um die geforderte Genauigkeit zu erreichen.

Wahl des KI-Algorithmus

- Traditionelle Algorithmen: Algorithmen aus der Statistik oder des maschinellen Lernens, wie Entscheidungsbäume oder Support-Vector-Machines, benötigen in der Regel weniger Daten als moderne neuronale Netze, können aber teils sehr gute Ergebnisse liefern.
- Moderne KI-Modelle: Diese Modelle, meist tiefen neuronale Netze, sind oft auf große und umfangreiche Datensätze angewiesen, um ihre komplexen Strukturen zu trainieren.

Techniken zur Datengenerierung

Wenn sehr große Datenmengen benötigt werden, aber nur eine begrenzte Datenmenge zur Verfügung steht, kann Datengenerierung, auch Datenaugmentation genannt, eine nützliche Technik sein. Dabei werden aus vorhandenen Daten zusätzliche Datenpunkte erzeugt (z. B. durch Transformation oder Synthese).

Datenspeicherung und -zugriff

Bei speicherintensiven Datentypen (z.B. Bilder) oder sehr großen Datenmengen spielt die Datenspeicherung und der Datenzugriff eine wichtige Rolle. Unternehmen müssen sicherstellen, dass sie über die notwendige Infrastruktur verfügen, um große Datenmengen effizient speichern und verarbeiten zu können.

Wirtschaftliche Überlegungen

Unternehmen sollten prüfen, ob die Investitionen in die Datenerhebung und -speicherung durch den Mehrwert der Daten gerechtfertigt sind, da dies zu zusätzlichen, nicht zu vernachlässigenden Mehrkosten führen kann.

Qualität vor Quantität

Zwar bereits in Abschnitt 2.2.2 *Dimensionen der Datenqualität* erwähnt, soll hier jedoch nochmals betont werden, dass es besser ist, eine kleine Menge qualitativ hochwertiger Daten zu haben, als eine große Menge fehlerhafter oder irrelevanter Daten. Die Datenqualität sollte niemals zugunsten der Datenmenge vernachlässigt werden.

Dabei spielt auch insbesondere die Vielfalt und Repräsentativität der Daten eine große Rolle, da das Modell in der Lage sein muss, aus einer Vielzahl von Szenarien zu lernen.

Praxisbeispiel „Effizientes Lagerbestandsmanagement durch KI-Prognosen“

Die Menge der benötigten Daten in unserem Praxisbeispiel hängt von der Komplexität der KI-Anwendung ab. Für komplexe Prognosemodelle werden umfangreiche historische Verkaufsdaten benötigt. Das Unternehmen sollte Techniken zur Datengenerierung und -augmentation in Betracht ziehen, um die Datenmenge zu erhöhen. Beispielsweise könnten historische Wetterdaten integriert werden, um saisonale Trends besser zu verstehen. Zudem muss die Infrastruktur zur Datenspeicherung und zum Datenzugriff skalierbar und zuverlässig sein. Dies bedeutet, die IT-Abteilung muss prüfen, ob ein zusätzlicher Server oder externer Dienstleister zur Datenspeicherung benötigt wird und der Zugriff auf die Daten schnell und effizient erfolgen kann. Wirtschaftliche Überlegungen müssen ebenfalls berücksichtigt werden. Es muss geprüft werden, ob die Nutzung komplexer KI-Modelle oder einfacher Modelle Gewinnbringender ist. Grundsätzlich sollte dabei stets die Qualität der Daten im Vordergrund stehen. Hochwertige Daten führen zu besseren Modellen und genaueren Vorhersagen.

2.4 Privatsphäre und rechtliche Aspekte

Rechtliche Aspekte spielen eine entscheidende Rolle bei der Nutzung von Daten für KI-Projekte. Unternehmen müssen sicherstellen, dass sie alle relevanten Gesetze und Vorschriften einhalten, die den Umgang mit Daten regeln, da sonst hohe Strafen drohen können.

Wichtige Datenschutzgesetze

- Die Datenschutz-Grundverordnung (DSGVO) auf europäischer Ebene gilt als das umfassendste Datenschutzgesetz, das den Umgang mit personenbezogenen Daten regelt und ist im Geltungsbereich der EU stets zu beachten.
- Zusätzlich existieren branchenspezifische Regelungen, die eingehalten werden müssen, beispielsweise für den medizinischen Bereich oder für Banken.
- Für KI-Systeme, die nach dem EU AI Act als Hochrisiko-KI-Systeme eingestuft werden, gibt es konkrete Auflagen an die Daten und deren Verarbeitung.

Maßnahmen zur Minimierung rechtlicher Risiken & Wahrung der Privatsphäre

- Implementierung von Datenschutzrichtlinien für den Schutz persönlicher und sensibler Daten.
- Ernennung eines Datenschutzbeauftragten als verantwortliche Person für Datenschutzfragen.
- Regelmäßige Schulungen von Mitarbeitenden zu rechtlichen Anforderungen.
- Zustimmung von betroffenen Personen einholen.
- Anonymisierung oder Pseudonymisierung: Entfernen bzw. ersetzen identifizierbarer Informationen, um die Rückverfolgung zu verhindern.

Praxisbeispiel „Effizientes Lagerbestandsmanagement durch KI-Prognosen“

Das Unternehmen in unserem Praxisbeispiel sollte sicherstellen, dass alle Datenschutzgesetze wie die DSGVO eingehalten werden. Maßnahmen zur Minimierung rechtlicher Risiken, wie die Anonymisierung von Kundendaten, sind dabei essenziell. Beispielsweise sollten personenbezogene Daten anonymisiert oder pseudonymisiert werden, um die Privatsphäre der Kunden des Händlers zu schützen. Der Schutz der Privatsphäre der Kunden muss stets gewährleistet sein, um Vertrauen und Compliance sicherzustellen. Dies bedeutet, dass das Unternehmen Datenschutzrichtlinien implementieren und regelmäßig überprüfen muss, um sicherzustellen, dass alle gesetzlichen Anforderungen erfüllt werden. Zudem sollten Schulungen für Mitarbeitende durchgeführt werden, um das Bewusstsein für Datenschutz und rechtliche Aspekte zu schärfen.

3 Fazit

In der heutigen digitalen Welt sind Daten das Herzstück erfolgreicher KI- und datenbasierter Projekte. Um zum Erfolg zu gelangen sind folgende Punkte zu beachten:

✓ **Daten als strategischen Wert begreifen**

Unternehmen, die Daten als strategischen Wert begreifen und entsprechend handeln, sind besser aufgestellt, um die Chancen und Herausforderungen der digitalen Transformation erfolgreich zu meistern.

✓ **Datenmanagement und Verantwortlichkeiten festlegen**

Sorgfältiges Datenmanagement, klare Verantwortlichkeiten und die Auflösung von Datensilos sind entscheidend, um die Verfügbarkeit und Nutzbarkeit von Daten sicherzustellen.

✓ **Anforderungen an die Datenqualität definieren**

Klare Anforderungen an die Datenqualität und regelmäßige Qualitätskontrollen und -prozesse müssen vorhanden sein, um sicherzustellen, dass die Daten den hohen Anforderungen genügen. Es gilt: „Die Qualität der Daten sollte immer Vorrang vor der reinen Quantität haben.“

✓ **Datenschutz und rechtliche Aspekte klären**

Der Schutz der Privatsphäre und relevante Datenschutzbestimmungen müssen bei der Nutzung von Daten stets sichergestellt werden. Auch das sollten ethische Richtlinien befolgt werden. Dies schafft Vertrauen und sorgt für einen verantwortungsvollen und fairen Umgang mit Daten und KI.

✓ **Investition in Datenqualität und Datenmanagement wagen**

Die Investition in Datenqualität und Datenmanagement ist immer mit einem gewissen Aufwand und meist mit Vorleistungen verbunden. Diese Investition zahlt sich in der Regel aus, da sie die eine unverzichtbare Grundlage für erfolgreiche datenbasierte und KI-Projekte ist und das Unternehmen zukunftssicherer macht.

Klären Sie deshalb die nächsten Schritte mit den KI-Labs - den regionalen Laboren für Künstliche Intelligenz in Baden-Württemberg. Diese unterstützen Sie gerne bei Ihren persönlichen Herausforderungen. Darüber hinaus finden sich aktuelle Informationen zu Themen Rund um die Digitalisierung auf dem Portal der Digitalisierungsstrategie des Wirtschaftsministeriums BW www.wirtschaft-digital-bw.de.

Impressum

Fraunhofer-Institut für Arbeitswirtschaft und Organisation IAO

Nobelstraße 12

70569 Stuttgart

www.iao.fraunhofer.de

Kontakt

Niclas Renner

Digital Business Services

Tel. +49 711 970 5172

niclas.renner@iao.fraunhofer.de

Gefördert durch

Ministerium für Wirtschaft, Arbeit und Tourismus Baden-Württemberg

Fraunhofer-Publica

<http://dx.doi.org/10.24406/publica-3757>

Alle Rechte vorbehalten

© Fraunhofer IAO, November 2024