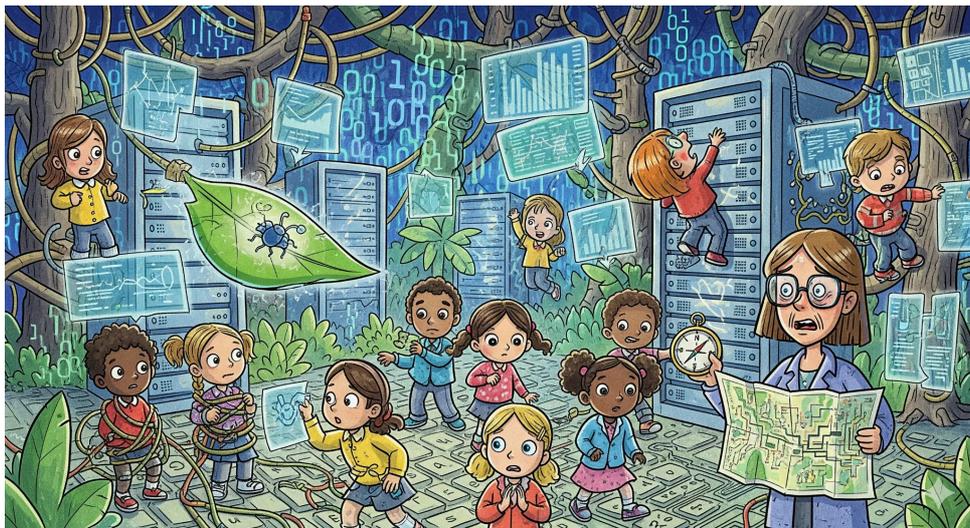


# Datenqualität für KI- Anwendungen

Qualität vor Quantität: Der Schlüssel zum KI-Erfolg



## Teilnehmerhandbuch

Daten KI-ready machen. Arno Schimmelpfennig  
Arno Schimmelpfennig | Qualität Ihrer KI-Ergebnisse deutlich zu

KI-Experte & Digitalisierungsberater



📍 Amalienstraße 11, 96047 Bamberg

☎ 0172 271 26 00

✉ info@arno-schimmelpfennig.de

🌐 as-digitalmarketing.de

🔗 linkedin.com/in/arno-schimmelpfennig

---

## Inhaltsverzeichnis

---

<b>1. Einführung: Der Daten-Realitätsschock</b> .....	3
1.1 Warum ChatGPT bei schlechten Daten versagt .....	3
1.2 Die 80/20-Regel der KI .....	3
<b>2. Die 5 kritischen Datentypen für KMU</b> .....	4
2.1 Kundendaten: Das digitale Gedächtnis Ihres Unternehmens .....	4
2.2 Produktdaten: Die Basis für intelligente Empfehlungen .....	5
2.3 Kommunikationsdaten: Verborgene Schätze in Gesprächen .....	6
2.4 Finanzdaten: Wenn Zahlen Geschichten erzählen .....	6
2.5 Dokumenten-Archive: Wissen bewahren und nutzen .....	7
<b>3. Von Chaos zu KI-Input: Der Datenaufbereitungsprozess</b> .....	8
3.1 Die Pyramide der Datenqualität .....	8
3.2 Der Weg zur Datenexzellenz .....	9
<b>4. Kostenlose Tools für die Datenaufbereitung</b> .....	10
4.1 OpenRefine: Der Datenreiniger .....	10
4.2 Excel/Google Sheets: Die Allrounder .....	11
4.3 LibreOffice Base: Die Datenbank-Alternative .....	12
<b>5. Rechtssichere Umsetzung</b> .....	13
5.1 DSGVO-konforme Datenverarbeitung .....	13

5.2 AI Act: Die neue Spielregeln für KI .....	14
<b>6. Praxisbeispiele und Anwendungsfälle .....</b>	<b>15</b>
6.1 Branchenspezifische Erfolgsgeschichten .....	15
© 2025 AS Digital Marketing   KI.Akademie - Modul 3: Datenbasis aufbauen - Qualität vor Quantität	
6.2 Vom Datenchaos zum KI-Erfolg: Zwei Wege .....	16
<b>7. Aktionsplan und nächste Schritte .....</b>	<b>17</b>
7.1 Daten-Audit: Wo stehen Sie heute? .....	17
7.2 30-Tage-Plan zur Datenqualitätsverbesserung .....	18

### **Wie Sie dieses Handbuch nutzen sollten:**

Dieses Handbuch ist kein trockenes Nachschlagewerk, sondern ein lebendiger Begleiter auf Ihrer Reise zu besseren Daten und erfolgreicherem KI-Anwendungen. Lesen Sie es wie einen Roman – von vorne nach hinten – oder springen Sie direkt zu den Kapiteln, die Sie am meisten interessieren. Unterwegs finden Sie Geschichten, praktische Tipps und Aha-Momente, die Ihnen helfen werden, das Thema Datenqualität mit neuen Augen zu sehen.

Nehmen Sie sich Zeit zum Schmökern, machen Sie sich Notizen und lassen Sie sich inspirieren. Die Reise zu exzellenten Daten beginnt mit dem ersten Schritt – und dieses Handbuch ist Ihre Landkarte.

---

## 1. Einführung: Der Daten-Realitätsschock

---

Es war ein Montagmorgen wie jeder andere, als Michael, der Geschäftsführer eines mittelständischen Handwerksbetriebs, voller Enthusiasmus sein neues ChatGPT-Konto einrichtete. Die Fachzeitschriften hatten es versprochen, die Berater hatten es bestätigt: Künstliche Intelligenz würde sein Unternehmen revolutionieren. Personalisierte Angebote erstellen, Kundenanfragen automatisch beantworten, Ressourcenplanung optimieren – die Möglichkeiten schienen grenzenlos.

Drei Stunden später saß Michael mit gerunzelter Stirn vor seinem Bildschirm. Die von ChatGPT erstellten Angebote enthielten falsche Kundennamen, nicht existierende Produkte und unrealistische Preise. Die automatischen Antworten auf Kundenanfragen waren bestenfalls generisch, schlimmstenfalls komplett irrelevant. Was war schiefgelaufen?

Michael erlebte, was ich den "Daten-Realitätsschock" nenne – jenen Moment, in dem die glänzenden Versprechen der KI-Revolution auf die raue Wirklichkeit unternehmerischer Datenwelten treffen. Ein Moment, den fast jedes Unternehmen auf seinem Weg zur KI-Integration durchlebt.

*"Garbage In, Garbage Out" – Dieses alte Informatiker-Prinzip gilt für KI mehr denn je.*

### 1.1 Warum ChatGPT bei schlechten Daten versagt

ChatGPT und andere KI-Systeme sind keine Zauberer, die aus dem Nichts brillante Ergebnisse erschaffen. Sie sind eher wie hochtalentierte Köche: Mit erstklassigen Zutaten können sie Meisterwerke kreieren, mit minderwertigen Zutaten bleibt selbst dem besten Koch nur Mittelmaß.

In Michaels Fall waren die "Zutaten" – seine Unternehmensdaten – alles andere als erstklassig. Seine Kundendatenbank enthielt Duplikate mit leicht abweichenden Schreibweisen (Familie Müller, Familie Mueller, Fam. Muller), veraltete Kontaktinformationen und unstrukturierte Notizen zu früheren Aufträgen. Seine Produktdaten waren über drei verschiedene Systeme verteilt, mit inkonsistenten Bezeichnungen und Preisen. Kein Wunder, dass ChatGPT damit überfordert war.

Was Michael erlebte, ist kein Einzelfall. In meiner Beratungspraxis sehe ich täglich Unternehmen, die mit großen Erwartungen in KI-Projekte starten, nur um dann festzustellen, dass ihre Daten nicht "KI-ready" sind. Die gute Nachricht: Dieses Problem lässt sich lösen – mit dem richtigen Ansatz und etwas Geduld.

*Ein Kunde von mir, Inhaber einer Schreinerei mit 12 Mitarbeitern, wollte ChatGPT nutzen, um automatisch Angebote für Standardmöbel zu erstellen. Sein erster Versuch endete in einem Desaster: Die KI erfand Holzarten, kalkulierte unrealistische Preise und ignorierte wichtige Kundenspezifikationen.*

*Wir nahmen uns drei Wochen Zeit, um seine Produktdaten zu strukturieren, Materialkosten zu standardisieren und Kundenpräferenzen in ein klares Format zu bringen. Das Ergebnis? ChatGPT erstellt heute 80% seiner Standardangebote – mit einer Genauigkeit, die selbst erfahrene Mitarbeiter beeindruckt.*

## 1.2 Die 80/20-Regel der KI

Wenn es um KI-Projekte geht, gilt eine überraschende Faustregel: 80% der Zeit und Ressourcen fließen in die Datenaufbereitung, nur 20% in die eigentliche KI-Anwendung. Diese Verteilung wird von vielen Unternehmen drastisch unterschätzt – mit fatalen Folgen für ihre KI-Initiativen.



Stellen Sie sich vor, Sie planen eine Weltreise. Würden Sie 80% Ihrer Vorbereitungszeit in die Auswahl des perfekten Koffers investieren und nur 20% in die Planung der Reiseroute, die Buchung von Unterkünften und das Festlegen der Reiseziele? Wahrscheinlich nicht. Doch genau diesen Fehler begehen viele Unternehmen bei KI-Projekten: Sie konzentrieren sich auf die Auswahl der KI-Tools (den "Koffer") und vernachlässigen die Datengrundlage (die "Reiseroute").

## 2. Die 5 kritischen Datentypen für KMU

Stellen Sie sich Ihr Unternehmen als ein lebendiges Ökosystem vor, in dem verschiedene Datenarten wie unterschiedliche Pflanzenarten gedeihen. Manche sind auffällig und offensichtlich wertvoll, andere unscheinbar, aber nicht weniger wichtig. In diesem Ökosystem gibt es fünf besonders kritische Datentypen, die für KI-Anwendungen von zentraler Bedeutung sind. Lassen Sie uns diese näher betrachten – nicht als trockene Kategorien, sondern als lebendige Bestandteile Ihres Unternehmens.

### 2.1 Kundendaten: Das digitale Gedächtnis Ihres Unternehmens

Kundendaten sind wie das kollektive Gedächtnis Ihres Unternehmens. Sie speichern nicht nur Namen und Adressen, sondern Geschichten, Vorlieben, Interaktionen und Vertrauen, das über Jahre aufgebaut wurde. In einer idealen Welt würden diese Daten ein perfektes, lückenloses Abbild jeder Kundenbeziehung darstellen. Die Realität sieht oft anders aus.

Ich erinnere mich an einen Besuch bei einem Möbelhaus, das stolz sein "hochmodernes CRM-System" präsentierte. Als wir einen Blick hineinwarfen, fanden wir ein faszinierendes Sammelsurium an Daten: Kunden, die gleichzeitig unter drei verschiedenen Schreibweisen existierten, Adressen ohne Postleitzahlen, Telefonnummern in fünf verschiedenen Formaten und Notizen, die von detaillierten Kaufhistorien bis hin zu kryptischen Abkürzungen reichten, die nur ein längst pensionierter Mitarbeiter hätte entschlüsseln können.

Für KI-Systeme sind solche Daten ein Albtraum. Stellen Sie sich vor, Sie würden einem neuen Mitarbeiter Kundeninformationen in einer fremden Sprache, mit unleserlicher Handschrift und auf zufällig zusammengewürfelten Notizzetteln präsentieren. Genau so fühlt sich eine KI mit chaotischen Kundendaten.

Der Weg zu KI-tauglichen Kundendaten beginnt mit Konsistenz und Struktur:

### Typische Probleme:

Kundendaten leiden häufig unter Duplikaten mit leicht abweichenden Informationen, veralteten Kontaktdaten und inkonsistenten Formatierungen. Besonders problematisch ist die fehlende Kategorisierung und Segmentierung, die für personalisierte KI-Anwendungen unerlässlich ist.

Ein weiteres häufiges Problem ist die Fragmentierung: Kundendaten existieren in verschiedenen Systemen – im CRM, in Excel-Listen, in E-Mail-Verteilern und sogar in den Köpfen langjähriger Mitarbeiter – ohne dass diese Quellen miteinander verbunden sind.

### KI-Potenzial:

Gut strukturierte Kundendaten sind Gold wert für KI-Anwendungen. Sie ermöglichen personalisierte Marketingkampagnen, die genau die Produkte vorschlagen, die ein Kunde wahrscheinlich kaufen wird. Sie befähigen Chatbots, auf frühere Interaktionen Bezug zu nehmen und so ein natürlicheres Gesprächserlebnis zu bieten.

Mit qualitativ hochwertigen Kundendaten kann KI sogar vorhersagen, welche Kunden abwanderungsgefährdet sind (Churn-Prognose), und proaktive Maßnahmen zur Kundenbindung vorschlagen.

### 2.2 Produktdaten: Die Basis für intelligente Empfehlungen

Wenn Kundendaten das Gedächtnis Ihres Unternehmens sind, dann sind Produktdaten sein Wissensschatz. Sie beschreiben, was Sie anbieten, wie es funktioniert, was es kostet und wie es sich von Alternativen unterscheidet. Für Menschen mit Branchenerfahrung sind viele dieser Informationen selbstverständlich – eine KI hingegen ist auf explizite, strukturierte Daten angewiesen.

Ich erinnere mich an ein Projekt mit einem Elektronikfachhändler, der seine Produktdaten für KI-gestützte Empfehlungen optimieren wollte. Als wir seinen Produktkatalog analysierten, stießen wir auf ein faszinierendes Phänomen: Das gleiche Produkt existierte unter verschiedenen Bezeichnungen, mit unterschiedlichen Attributen und sogar zu verschiedenen Preisen – je nachdem, über welchen Kanal es verkauft wurde.

Für den Händler war das "normal" – er hatte die mentale Fähigkeit, diese Inkonsistenzen zu überbrücken. Für die KI war es ein unüberwindbares Hindernis. Sie konnte nicht verstehen, dass "Premium-HDMI-Kabel 2m" und "HDMI-Anschlusskabel High-Speed 2,0m" das gleiche Produkt waren, oder dass "4K-fähig" und "Ultra-HD-kompatibel" die gleiche Eigenschaft beschrieben.

*Ein Elektronikhändler wollte KI nutzen, um Kunden passende Zubehörprodukte zu empfehlen. Die ersten Ergebnisse waren katastrophal: Die KI empfahl iPhone-Hüllen zu Android-Smartphones und Druckerpatronen zu Modellen, mit denen sie nicht kompatibel waren.*

*Das Problem? In den Produktdaten fehlten strukturierte Kompatibilitätsinformationen. Nach einer gründlichen Überarbeitung der Produktdatenbank mit klaren Kategorien, standardisierten Attributen und*

*expliziten Kompatibilitätsangaben verbesserte sich die Relevanz der Empfehlungen dramatisch – und der Durchschnittswarenkorb stieg um 23%.*

Der Weg zu KI-tauglichen Produktdaten führt über Standardisierung und Anreicherung:

Produktdaten müssen eine einheitliche Struktur aufweisen, mit konsistenten Kategorien, Attributen und Beschreibungen. Jedes Produkt sollte eindeutig identifizierbar sein und alle relevanten Informationen in einem maschinenlesbaren Format enthalten. Besonders wichtig sind Beziehungen zwischen Produkten – wie Kompatibilität, Ergänzungsprodukte oder Alternativen.

Die Herausforderung liegt oft darin, dass Produktdaten aus verschiedenen Quellen stammen – von Herstellern, aus ERP-Systemen, aus Webshops – und in jedem System anders strukturiert sind. Eine zentrale Produktdatenbank (PIM – Product Information Management) kann hier Abhilfe schaffen, erfordert aber zunächst eine gründliche Bereinigung und Harmonisierung der vorhandenen Daten.

Mit qualitativ hochwertigen Produktdaten kann KI nicht nur relevante Empfehlungen geben, sondern auch bei der Erstellung von Produktbeschreibungen helfen, Bestandsoptimierungen vorschlagen und sogar dynamische Preisanpassungen basierend auf Nachfrage, Wettbewerb und Lagerbestand vornehmen.

Ein häufiges Problem bei Produktdaten ist das "Silodenken": Jede Abteilung pflegt ihre eigene Sicht auf die Produkte. Der Einkauf interessiert sich für Einkaufspreise und Lieferanten, der Vertrieb für Verkaufspreise und Features, die Logistik für Maße und Gewichte. Für KI-Anwendungen müssen diese verschiedenen Perspektiven in einer konsistenten, ganzheitlichen Produktansicht zusammengeführt werden.

---

### 2.3 Kommunikationsdaten: Verborgene Schätze in Gesprächen

Stellen Sie sich vor, Sie könnten jedes Gespräch, jede E-Mail und jede Chat-Nachricht zwischen Ihrem Unternehmen und Ihren Kunden analysieren – welche Erkenntnisse würden Sie gewinnen? Welche Muster würden Sie erkennen? Genau hier liegt das Potenzial von Kommunikationsdaten für KI-Anwendungen.

Kommunikationsdaten sind vielleicht der am meisten unterschätzte Datenschatz in Unternehmen. In E-Mails, Chat-Protokollen, Kundenservice-Tickets und Social-Media-Interaktionen verbergen sich wertvolle Informationen über Kundenbedürfnisse, häufige Probleme, erfolgreiche Lösungsansätze und Stimmungen.

Das Problem: Diese Daten liegen meist unstrukturiert vor – als Fließtext, ohne klare Kategorisierung oder Verknüpfung mit anderen Daten. Für Menschen mit Kontext sind sie verständlich, für KI-Systeme zunächst ein undurchdringlicher Dschungel aus Worten.

Der Schlüssel zur Nutzbarmachung von Kommunikationsdaten liegt in der Strukturierung und Kategorisierung. E-Mails und Tickets sollten nach Themen, Dringlichkeit und Status kategorisiert werden. Chat-Protokolle sollten mit Kundendaten verknüpft sein, um eine ganzheitliche Sicht auf die Kundenbeziehung zu ermöglichen.

Mit gut aufbereiteten Kommunikationsdaten kann KI automatische Antworten auf häufige Fragen generieren, Stimmungen analysieren (Sentiment-Analyse), Anfragen automatisch kategorisieren und priorisieren oder sogar proaktiv Lösungen für wiederkehrende Probleme vorschlagen.

## 2.4 Finanzdaten: Wenn Zahlen Geschichten erzählen

Finanzdaten werden oft als trocken und nüchtern wahrgenommen – doch für KI-Systeme sind sie faszinierende Geschichtenerzähler. In Rechnungen, Buchungen und Zahlungsverläufen verbergen sich Muster und Trends, die wertvolle Einblicke in die Gesundheit und Entwicklung eines Unternehmens geben können.

Ich erinnere mich an ein Projekt mit einem Großhändler, der KI nutzen wollte, um Cashflow-Prognosen zu erstellen. Als wir seine Finanzdaten analysierten, stießen wir auf ein grundlegendes Problem: Die Buchungstexte waren so inkonsistent und kryptisch, dass selbst erfahrene Buchhalter Mühe hatten, sie zu entschlüsseln. "RE 123/23", "Zahlung Kunde M.", "Ausgleich Jan" – für Menschen mit Kontext verständlich, für KI-Systeme bedeutungslos.

Für KI-Anwendungen müssen Finanzdaten strukturiert, konsistent und mit ausreichend Kontext versehen sein. Jede Buchung sollte eindeutig kategorisiert sein, mit standardisierten Beschreibungen und klaren Verknüpfungen zu Kunden, Lieferanten oder Projekten.

Mit qualitativ hochwertigen Finanzdaten kann KI nicht nur Cashflow-Prognosen erstellen, sondern auch Anomalien erkennen (Betrugsschutz), Kostenoptimierungen vorschlagen oder automatisch Rechnungen verarbeiten und kategorisieren.

*Ein mittelständisches Handelsunternehmen wollte KI nutzen, um Zahlungsausfälle vorherzusagen. Die ersten Modelle lieferten enttäuschende Ergebnisse mit einer Trefferquote kaum besser als der Zufall.*

*Die Ursache? Die Zahlungshistorien waren unvollständig und nicht mit den Kundendaten verknüpft. Nach einer gründlichen Datenbereinigung und -integration konnte das KI-Modell Zahlungsausfälle mit einer Genauigkeit von 78% vorhersagen – früh genug, um proaktive Maßnahmen zu ergreifen.*

### 2.5 Dokumenten-Archive: Wissen bewahren und nutzen

In jedem Unternehmen schlummert ein Schatz an Wissen – in Verträgen, Handbüchern, technischen Dokumentationen, Protokollen und Berichten. Dieses Wissen ist oft nur schwer zugänglich, verteilt auf verschiedene Ablagesysteme, Formate und Medien. Für KI-Anwendungen liegt hier enormes Potenzial – wenn die Dokumente entsprechend aufbereitet sind.

Das typische Dokumenten-Archiv eines mittelständischen Unternehmens gleicht oft einem verwunschenen Labyrinth: PDF-Dokumente ohne Texterkennung, gescannte Verträge ohne Metadaten, Word-Dokumente mit inkonsistenten Formatierungen und Benennungen, verteilt auf Netzlaufwerke, lokale Rechner und Cloud-Speicher. Für Menschen mit Insiderwissen navigierbar, für KI-Systeme ein undurchdringliches Chaos.

Der Weg zu KI-tauglichen Dokumenten-Archiven führt über Digitalisierung, Texterkennung und Metadaten-Anreicherung. Dokumente müssen in durchsuchbaren Formaten vorliegen, mit konsistenten Metadaten versehen sein und einer klaren Struktur folgen.

Mit gut aufbereiteten Dokumenten-Archiven kann KI Verträge analysieren und vergleichen, technische Dokumentationen durchsuchen und relevante Informationen extrahieren, Wissensmanagement-Systeme mit automatischen Antworten auf häufige Fragen füttern oder Compliance-Risiken in Dokumenten identifizieren.

#### **Datenqualitäts-Check: Wo beginnen?**

Wenn Sie Ihre Daten KI-ready machen wollen, stellt sich die Frage: Wo anfangen? Meine Empfehlung: Identifizieren Sie den Datentyp, der in Ihrem Unternehmen am chaotischsten ist *und* den größten Geschäftswert verspricht.

Für ein Handelsunternehmen sind das oft die Produktdaten, für einen Dienstleister die Kundendaten, für ein Produktionsunternehmen vielleicht die technischen Dokumentationen. Beginnen Sie Ihre Datenbereinigung mit diesem Bereich, um schnell sichtbare Verbesserungen zu erzielen.

© 2025 AS Digital Marketing | KI.Akademie - Modul 3: Datenbasis aufbauen - Qualität vor Quantität

Die fünf kritischen Datentypen existieren nicht isoliert – sie sind miteinander verwoben und beeinflussen sich gegenseitig. Ein Kunde (Kundendaten) kauft Produkte (Produktdaten), kommuniziert mit dem Kundenservice (Kommunikationsdaten), bezahlt Rechnungen (Finanzdaten) und ist vielleicht Gegenstand von Verträgen oder Dokumentationen (Dokumenten-Archive).

Für wirklich leistungsfähige KI-Anwendungen ist es entscheidend, diese Verbindungen herzustellen und eine ganzheitliche Sicht auf Daten zu entwickeln. Dies erfordert nicht nur technische Lösungen, sondern auch ein Umdenken in der Organisation – weg vom Silodenken, hin zu einer integrierten Datenstrategie.

*"Daten sind das neue Öl – aber wie Öl müssen sie raffiniert werden, um wertvoll zu sein."*

Im nächsten Kapitel werden wir uns anschauen, wie dieser Raffinerieprozess aussieht – wie Sie Ihre Daten Schritt für Schritt von chaotischen Rohdaten zu KI-ready Informationen transformieren können.

### 3. Von Chaos zu KI-Input: Der Datenaufbereitungsprozess

Stellen Sie sich vor, Sie betreten eine Küche, in der alle Zutaten durcheinandergeworfen sind – Mehl neben Waschmittel, Salz vermischt mit Zucker, Gewürze ohne Beschriftung. Könnten Sie dort ein Gourmet-Menü zubereiten? Wahrscheinlich nicht. Genau so fühlt sich eine KI, wenn sie mit chaotischen, unstrukturierten Daten arbeiten soll.

Der Weg von Datenchaos zu KI-tauglichen Daten ist eine Reise der Transformation – von rohen, unbearbeiteten Informationsbrocken zu fein säuberlich strukturierten, konsistenten und kontextreichen Datensätzen. Diese Reise folgt einem bestimmten Pfad, den ich als "Pyramide der Datenqualität" bezeichne.

#### 3.1 Die Pyramide der Datenqualität

Die Datenqualität entwickelt sich in Stufen, wobei jede Stufe auf der vorherigen aufbaut und einen höheren Grad an Ordnung, Konsistenz und Nutzbarkeit darstellt:



**Stufe 1: Rohdaten – Das Chaos der Realität**

Auf dieser Stufe befinden sich Daten in ihrem Urzustand – unbearbeitet, ungeordnet, mit Inkonsistenzen, Duplikaten, Fehlern und fehlenden Werten. Sie existieren in verschiedenen Formaten, Systemen und Strukturen. Denken Sie an eine Excel-Tabelle, in der jeder Mitarbeiter nach eigenem Gutdünken Daten eingetragen hat, oder an einen Ordner voller PDF-Dokumente ohne erkennbares Benennungsschema.

Rohdaten sind wie ungeschliffene Diamanten – wertvoll, aber in ihrer aktuellen Form kaum nutzbar. Für KI-Systeme sind sie ein Minenfeld voller Missverständnisse und Fehlinterpretationen.

### **Stufe 2: Bereinigt – Der erste Schritt zur Ordnung**

Auf dieser Stufe wurden die größten Probleme beseitigt. Duplikate wurden entfernt, offensichtliche Fehler korrigiert, fehlende Werte identifiziert und behandelt. Die Daten sind konsistenter, aber noch nicht vollständig strukturiert.

Bereinigte Daten sind wie sortierte Zutaten in einer Küche – alle Gewürze stehen zusammen, Mehl und Zucker sind getrennt, verdorbene Lebensmittel wurden aussortiert. Man kann damit arbeiten, aber für ein Gourmet-Menü fehlt noch die feine Organisation.

### **Stufe 3: Strukturiert – Klare Ordnung und Beziehungen**

Auf dieser Stufe haben Daten eine klare Struktur mit definierten Spalten, Datentypen und Beziehungen. Kategorien und Hierarchien sind festgelegt, Formate sind einheitlich. Die Daten folgen einem konsistenten Schema und sind in sich schlüssig.

Strukturierte Daten sind wie eine perfekt organisierte Küche – jede Zutat hat ihren festen Platz, ist korrekt beschriftet und nach logischen Kriterien angeordnet. Ein guter Koch kann damit effizient arbeiten.

## Stufe 4: KI-Ready – Bereit für intelligente Verarbeitung

Auf der höchsten Stufe sind Daten nicht nur strukturiert, sondern auch validiert, angereichert und in standardisierten Formaten verfügbar, die direkt von KI-Systemen verarbeitet werden können. Sie enthalten alle notwendigen Kontextinformationen und Metadaten, sind vollständig, konsistent und aktuell.

KI-ready Daten sind wie vorbereitete Zutaten für einen Sternekoch – nicht nur perfekt organisiert, sondern auch in der idealen Form und Qualität, um sofort verarbeitet zu werden. Der Koch kann sich voll auf die kreative Zubereitung konzentrieren, ohne sich um die Grundlagen sorgen zu müssen.

*In einem Workshop fragte ich die Teilnehmer, wo ihre Unternehmensdaten in dieser Pyramide stehen. Die meisten antworteten "zwischen Bereinigt und Strukturiert" – ein typischer Zwischenzustand in vielen Unternehmen. Die Daten sind nicht mehr völlig chaotisch, aber noch weit entfernt von dem Qualitätsniveau, das für anspruchsvolle KI-Anwendungen erforderlich ist.*

*Ein Teilnehmer aus dem Maschinenbau beschrieb seine Situation treffend: "Unsere Daten sind wie ein aufgeräumtes Chaos – man findet zwar alles, aber es dauert."*

## 3.2 Der Weg zur Datenexzellenz

Wie gelangt man von chaotischen Rohdaten zu KI-ready Informationen? Der Prozess folgt einem strukturierten Pfad, den ich als "Datenveredelung" bezeichne. Lassen Sie mich Sie auf diese Reise mitnehmen – nicht als trockene Anleitung, sondern als Geschichte der Transformation.

### Schritt 1: Bestandsaufnahme und Analyse – Den Dschungel kartieren

Jede erfolgreiche Expedition beginnt mit einer guten Karte. Im Fall der Datenaufbereitung bedeutet das, alle relevanten Datenquellen zu identifizieren und deren Qualität zu analysieren. Welche Systeme enthalten welche Daten? Wo liegen die größten Probleme? Welche Daten sind für Ihre KI-Anwendungen besonders wichtig?

© 2025 AS Digital Marketing | KI.Akademie - Modul 3: Datenbasis aufbauen - Qualität vor Quantität

Ich erinnere mich an ein Projekt mit einem Großhändler, der seine Produktdaten KI-ready machen wollte. Als wir mit der Bestandsaufnahme begannen, entdeckten wir nicht weniger als sieben verschiedene Systeme, die Produktinformationen enthielten – vom ERP über Excel-Listen bis hin zu einem veraltetem Produktkatalog-System, das nur noch ein Mitarbeiter bedienen konnte, der kurz vor der Rente stand. Jedes System hatte seine eigene Struktur, seine eigenen Bezeichnungen und seine eigene "Wahrheit" über die Produkte.

Eine gründliche Bestandsaufnahme bringt solche versteckten Datenquellen ans Licht und schafft Klarheit über den tatsächlichen Zustand Ihrer Datenlandschaft. Sie ist der erste, entscheidende Schritt auf dem Weg zur Datenexzellenz.

## **Schritt 2: Datenbereinigung – Aufräumen im Chaos**

Mit einem klaren Bild der Ausgangssituation beginnt die eigentliche Arbeit: das Aufräumen. In diesem Schritt werden Duplikate entfernt, offensichtliche Fehler korrigiert und fehlende Werte behandelt. Formate werden standardisiert, Schreibweisen vereinheitlicht.

Die Datenbereinigung ist oft mühsam und zeitaufwändig, aber sie legt den Grundstein für alle weiteren Schritte. Sie ist wie das Aussortieren und Waschen von Zutaten, bevor man mit dem Kochen beginnt – nicht glamourös, aber unerlässlich für ein gutes Ergebnis.

## **Schritt 3: Datenstrukturierung – Ordnung schaffen**

Nach der grundlegenden Bereinigung folgt die Strukturierung. In diesem Schritt werden klare Spaltenbezeichnungen und Datentypen definiert, konsistente Kategorien und Hierarchien erstellt und Beziehungen zwischen Datensätzen hergestellt.

Die Strukturierung verwandelt bereinigte Daten in ein logisches, zusammenhängendes Ganzes. Sie ist wie das Organisieren einer Küche nach einem durchdachten System – jede Zutat hat ihren festen Platz, der einem logischen Prinzip folgt.

Ein entscheidender Aspekt der Strukturierung ist die Normalisierung – das Aufteilen von Daten in logische Einheiten, um Redundanzen zu vermeiden. Statt in jeder Kundenbestellung die vollständigen Kundendaten zu wiederholen, wird eine separate Kundentabelle erstellt und über eine eindeutige ID verknüpft. Dies spart nicht nur Speicherplatz, sondern verhindert auch Inkonsistenzen bei Änderungen.

## **Schritt 4: Datenvalidierung – Qualitätssicherung**

Mit einer klaren Struktur im Rücken ist es Zeit für die Qualitätssicherung. In diesem Schritt werden die Daten anhand definierter Kriterien überprüft. Stimmen die Werte mit der Realität überein? Sind alle Beziehungen korrekt? Gibt es versteckte Inkonsistenzen?

Die Validierung kann durch Stichproben, automatisierte Tests oder den Abgleich mit bekannten Referenzwerten erfolgen. Sie ist wie die Qualitätskontrolle in einer Küche – der Koch probiert die Zutaten, um sicherzustellen, dass sie frisch und hochwertig sind.

## **Schritt 5: Export in KI-ready Formate – Die letzte Meile**

Der letzte Schritt ist die Bereitstellung der Daten in Formaten, die von KI-Systemen optimal verarbeitet werden können. Dies können CSV- oder JSON-Dateien sein, API-

Endpunkte oder spezielle Datenbanken für maschinelles Lernen.

Der Export ist mehr als nur eine Formatkonvertierung – er umfasst auch die Anreicherung mit Metadaten, die Dokumentation der Datenstruktur und die Sicherstellung der Kompatibilität mit den gewählten KI-Tools.

© 2025 AS Digital Marketing | KI.Akademie - Modul 3: Datenbasis aufbauen - Qualität vor Quantität

### **Praxis-Tipp: Dokumentieren Sie jeden Schritt**

Ein oft übersehener, aber entscheidender Aspekt der Datenaufbereitung ist die Dokumentation. Halten Sie fest, welche Entscheidungen Sie getroffen haben, welche Transformationen Sie durchgeführt haben und warum. Dies ermöglicht es Ihnen, den Prozess zu wiederholen, bei Bedarf anzupassen und Entscheidungen nachvollziehbar zu machen.

Eine gute Dokumentation ist wie ein Rezept – sie ermöglicht es anderen (oder Ihnen selbst in der Zukunft), die gleichen Ergebnisse zu erzielen, ohne das Rad neu erfinden zu müssen.

Der Weg von Rohdaten zu KI-ready Informationen ist keine einmalige Reise, sondern ein kontinuierlicher Prozess. Daten ändern sich, neue Quellen kommen hinzu, Anforderungen entwickeln sich weiter. Eine nachhaltige Datenstrategie umfasst daher nicht nur die initiale Aufbereitung, sondern auch Prozesse zur kontinuierlichen Qualitätssicherung und Aktualisierung.

Im nächsten Kapitel werden wir uns anschauen, welche kostenlosen Tools Ihnen auf dieser Reise helfen können – praktische Helfer, die den Weg von Datenchaos zu Datenexzellenz erleichtern.

## 4. Kostenlose Tools für die Datenaufbereitung

Die gute Nachricht vorweg: Sie benötigen keine teuren Spezialsoftware oder Enterprise-Lösungen, um Ihre Daten KI-ready zu machen. Es gibt eine Reihe leistungsstarker, kostenloser Tools, die Ihnen bei der Datenbereinigung und -strukturierung helfen können. Lassen Sie uns drei besonders nützliche Werkzeuge näher betrachten – nicht als trockene Funktionslisten, sondern als praktische Helfer auf Ihrer Datenreise.

Bevor wir in die Details eintauchen, hier ein Überblick über die drei Tools, die wir vorstellen werden:

Tool	Haupteinsatzgebiet	Schwierigkeitsgrad	Besonders geeignet für
OpenRefine	Bereinigung und Transformation großer Datensätze	Mittel	Unstrukturierte und "messy" Daten
Excel/Google Sheets	Allgemeine Datenbearbeitung und -analyse	Niedrig bis Mittel	Tabellarische Daten, einfache Bereinigungsaufgaben
LibreOffice Base	Relationale Datenbanken	Mittel bis Hoch	Komplexe Datenstrukturen mit Beziehungen

### 4.1 OpenRefine: Der Datenreiniger

Stellen Sie sich vor, Sie haben eine riesige Kiste voller Legosteine in allen möglichen Farben, Formen und Größen, wild durcheinandergewürfelt. Ihre Aufgabe ist es, sie

zu sortieren, beschädigte Teile auszusortieren und ähnliche Teile zu gruppieren. Genau das macht OpenRefine mit Ihren Daten.

OpenRefine (früher Google Refine) ist ein leistungsstarkes Tool zur Bereinigung und Transformation von "messy" Daten. Es eignet sich besonders für die Arbeit mit großen Datensätzen und erkennt automatisch Muster und Inkonsistenzen.

Was OpenRefine besonders wertvoll macht, ist seine Fähigkeit, ähnliche Werte zu erkennen und zu gruppieren. Stellen Sie sich vor, Sie haben eine Kundendatenbank mit verschiedenen Schreibweisen für die gleiche Stadt: "München", "Muenchen", "Munich", "Mnchen" (Tippfehler). OpenRefine erkennt diese Ähnlichkeiten und ermöglicht es Ihnen, sie mit wenigen Klicks zu vereinheitlichen.

Ein weiteres Highlight ist die GREL (Google Refine Expression Language) – eine mächtige Sprache zur Datentransformation. Mit GREL können Sie komplexe Operationen durchführen, wie das Extrahieren von Informationen aus unstrukturiertem Text, das Umformatieren von Daten oder das Kombinieren von Werten aus verschiedenen Spalten.

*Ein Kunde von mir, ein kleiner Onlinehändler, hatte eine Produktdatenbank mit über 5.000 Artikeln, die über Jahre hinweg von verschiedenen Mitarbeitern gepflegt wurde. Das Ergebnis war ein Chaos aus inkonsistenten Kategorien, Schreibweisen und Formaten.*

*Mit OpenRefine konnten wir innerhalb eines Tages die Produktkategorien von über 200 verschiedenen Schreibweisen auf 15 standardisierte Kategorien reduzieren. Die Clustering-Funktion erkannte automatisch ähnliche Bezeichnungen wie "Küchenzubehör", "Küchenhelfer" und "Küchen-Accessoires" und schlug vor, sie zu vereinheitlichen.*

*Was manuell Wochen gedauert hätte, war mit dem richtigen Tool in Stunden erledigt.*

OpenRefine ist besonders nützlich für die ersten Phasen der Datenaufbereitung – die Bereinigung und grundlegende Strukturierung. Es hilft Ihnen, aus chaotischen Rohdaten saubere, konsistente Datensätze zu machen, die dann weiter verarbeitet werden können.

Ein typischer Workflow mit OpenRefine könnte so aussehen:

- 1 Importieren Sie Ihre Daten (z.B. eine CSV-Datei mit Kundendaten)
- 2 Nutzen Sie "Facets" und "Filters", um einen Überblick über die Daten zu bekommen und Probleme zu identifizieren
- 3 Verwenden Sie "Cluster and Edit", um ähnliche Werte zu finden und zu vereinheitlichen
- 4 Transformieren Sie Daten mit GREL (z.B. Formatierung von Telefonnummern)
- 5 Exportieren Sie die bereinigten Daten für die weitere Verarbeitung

**Download und Installation:** OpenRefine ist kostenlos verfügbar unter [openrefine.org](https://openrefine.org). Es läuft auf Windows, macOS und Linux und benötigt Java.

## 4.2 Excel/Google Sheets: Die Allrounder

Manchmal sind die besten Werkzeuge diejenigen, die wir bereits kennen und täglich nutzen. Microsoft Excel und Google Sheets sind weit verbreitete Tabellenkalkulationsprogramme, die überraschend leistungsstarke Funktionen zur Datenbereinigung und -analyse bieten.

Was Excel und Google Sheets besonders wertvoll macht, ist ihre Zugänglichkeit. Fast jeder hat schon einmal mit einer Tabellenkalkulation gearbeitet, die Lernkurve

ist flach, und die Ergebnisse sind sofort sichtbar. Sie eignen sich besonders für kleinere bis mittlere Datensätze und alltägliche Datenaufgaben.

Für die Datenbereinigung bieten Excel und Google Sheets eine Reihe nützlicher Funktionen:

© 2025 AS Digital Marketing | KI.Akademie - Modul 3: Datenbasis aufbauen - Qualität vor Quantität

**Duplikate finden und entfernen:** Mit wenigen Klicks können Sie doppelte Einträge identifizieren und beseitigen. In Excel finden Sie diese Funktion unter "Daten" > "Duplikate entfernen", in Google Sheets unter "Daten" > "Daten bereinigen" > "Duplikate entfernen".

**Datenvalidierung:** Sie können Regeln festlegen, welche Werte in bestimmten Zellen erlaubt sind – z.B. nur Zahlen in einem bestimmten Bereich oder nur Werte aus einer vordefinierten Liste. Dies hilft, neue Fehler zu vermeiden.

**Bedingte Formatierung:** Heben Sie Zellen hervor, die bestimmten Kriterien entsprechen – z.B. alle leeren Zellen oder Werte außerhalb eines normalen Bereichs. Dies macht Probleme in den Daten sofort sichtbar.

**Formeln und Funktionen:** Eine Vielzahl von Funktionen hilft bei der Datenbereinigung, darunter:

- **TRIM()** - Entfernt Leerzeichen am Anfang und Ende eines Textes
- **PROPER()** - Wandelt den ersten Buchstaben jedes Wortes in einen Großbuchstaben um
- **CONCATENATE()** oder **&** - Verbindet Texte aus verschiedenen Zellen
- **LEFT()**, **MID()**, **RIGHT()** - Extrahiert Teile eines Textes
- **SUBSTITUTE()** - Ersetzt bestimmte Zeichen oder Texte innerhalb eines Strings

Ein besonderes Highlight in Excel ist Power Query (unter "Daten" > "Daten abrufen und transformieren"). Dieses leistungsstarke Tool ermöglicht es Ihnen, Daten aus verschiedenen Quellen zu kombinieren, zu transformieren und zu bereinigen – mit einer benutzerfreundlichen Oberfläche und ohne komplexe Formeln.

Mit Power Query können Sie:

- Daten aus verschiedenen Quellen kombinieren (Excel-Tabellen, CSV-Dateien, Datenbanken, Webseiten)
- Spalten umbenennen, löschen oder neu anordnen
- Datentypen ändern und validieren
- Zeilen filtern und sortieren
- Berechnete Spalten erstellen
- Duplikate entfernen und Fehler behandeln

Das Beste an Power Query: Es zeichnet jeden Schritt Ihrer Transformation auf, sodass Sie den Prozess jederzeit wiederholen können – ideal für regelmäßige Datenaktualisierungen.

*Eine Kundin von mir, Inhaberin eines kleinen Modegeschäfts, kämpfte mit einem chaotischen Produktkatalog. Jede Saison erhielt sie Excel-Listen von verschiedenen Lieferanten – jede mit einer anderen Struktur, anderen Spaltenbezeichnungen und anderen Konventionen.*

*Mit Power Query in Excel konnten wir einen Transformationsprozess definieren, der die verschiedenen Listen automatisch in ein einheitliches Format brachte. Spalten wurden umbenannt, Preise in ein konsistentes Format gebracht, Produktkategorien standardisiert.*

*Was früher Tage manueller Arbeit bedeutete, war nun mit wenigen Klicks erledigt – und der Prozess konnte für jede neue Lieferantenliste*

wiederverwendet werden.

© 2025 by Spindt Consulting AG. Google Sheets ist kostenlos und ermöglicht einfache Zusammenarbeit, hat aber Einschränkungen bei großen Datensätzen und bietet nicht alle fortgeschrittenen Funktionen von Excel (wie Power Query). Excel bietet mehr Leistung und Funktionen, erfordert aber eine Lizenz.

Für DSGVO-konforme Verarbeitung sensibler Daten ist Excel oft die bessere Wahl, da die Daten lokal verarbeitet werden können, während Google Sheets eine Cloud-Lösung ist.

### 4.3 LibreOffice Base: Die Datenbank-Alternative

Wenn Ihre Daten komplexer werden und viele Beziehungen zwischen verschiedenen Datensätzen bestehen, stoßen Tabellenkalkulationen wie Excel an ihre Grenzen. Hier kommt LibreOffice Base ins Spiel – eine kostenlose Datenbank-Anwendung, die Teil des LibreOffice-Pakets ist.

LibreOffice Base ermöglicht es Ihnen, relationale Datenbanken zu erstellen und zu verwalten – ohne teure Datenbanksoftware oder tiefgreifende technische Kenntnisse. Es bietet eine benutzerfreundliche Oberfläche für die Erstellung von Tabellen, Abfragen, Formularen und Berichten.

Was LibreOffice Base besonders wertvoll macht, ist seine Fähigkeit, Beziehungen zwischen Daten zu modellieren. Statt alle Informationen in einer großen Tabelle zu speichern (was zu Redundanzen und Inkonsistenzen führt), können Sie die Daten in logische Einheiten aufteilen und durch Beziehungen verknüpfen.

Ein klassisches Beispiel ist die Beziehung zwischen Kunden und Bestellungen:

- In einer Tabelle "Kunden" speichern Sie alle Kundeninformationen (Name, Adresse, Kontaktdaten) – jeder Kunde genau einmal, mit einer eindeutigen KundenID.
- In einer Tabelle "Bestellungen" speichern Sie alle Bestellinformationen (Datum, Gesamtbetrag, Status) – mit einem Verweis (Fremdschlüssel) auf die KundenID.
- In einer Tabelle "Bestellpositionen" speichern Sie die einzelnen Positionen jeder Bestellung – mit einem Verweis auf die BestellID und die ProduktID.

Diese Struktur hat mehrere Vorteile:

- Keine Redundanz: Kundeninformationen werden nur einmal gespeichert, nicht für jede Bestellung wiederholt.
- Konsistenz: Wenn sich eine Kundenadresse ändert, muss sie nur an einer Stelle aktualisiert werden.
- Flexibilität: Komplexe Abfragen über mehrere Tabellen hinweg sind möglich.

Für KI-Anwendungen ist diese strukturierte, relationale Datenhaltung oft ideal – sie bietet eine klare Organisation, vermeidet Redundanzen und ermöglicht komplexe Abfragen.

*Ein Handwerksbetrieb, mit dem ich zusammenarbeitete, hatte seine Kundendaten, Aufträge und Materialien in verschiedenen Excel-Tabellen*

gespeichert. Das führte zu Duplikaten, Inkonsistenzen und Schwierigkeiten bei der Analyse von Kundenhistorien.

Mit LibreOffice Base erstellten wir eine einfache relationale Datenbank mit Tabellen für Kunden, Aufträge, Materialien und Arbeitszeiten. Durch die Verknüpfung dieser Tabellen konnte der Betrieb nun leicht Fragen beantworten wie: "Welche Materialien wurden bei diesem Kunden am häufigsten verwendet?" oder "Welche Aufträge haben am meisten Arbeitszeit pro Euro Umsatz erfordert?"

Diese strukturierten Daten bildeten später die Grundlage für KI-Anwendungen zur Angebotskalkulation und Ressourcenplanung.

**Download und Installation:** LibreOffice ist kostenlos verfügbar unter [libreoffice.org](https://libreoffice.org). Es läuft auf Windows, macOS und Linux.

## Welches Tool ist das richtige für Sie?

Die Wahl des richtigen Tools hängt von Ihren spezifischen Anforderungen, dem Zustand Ihrer Daten und Ihrem Komfortniveau mit verschiedenen Technologien ab:

- **OpenRefine** ist ideal, wenn Sie große, unstrukturierte Datensätze mit vielen Inkonsistenzen haben und eine leistungsstarke Lösung für die initiale Bereinigung benötigen.
- **Excel/Google Sheets** sind perfekt für die tägliche Datenarbeit mit überschaubaren Datensätzen, besonders wenn Sie bereits mit diesen Tools vertraut sind.
- **LibreOffice Base** ist die richtige Wahl, wenn Sie komplexe Datenstrukturen mit vielen Beziehungen zwischen Datensätzen haben und eine strukturierte, relationale Datenhaltung benötigen.

Oft ist die Kombination mehrerer Tools am effektivsten. Ein typischer Workflow könnte so aussehen:

1. Initiale Bereinigung großer, chaotischer Datensätze mit OpenRefine
2. Weitere Verfeinerung und Analyse in Excel mit Power Query

3. Strukturierte Speicherung in einer relationalen Datenbank mit LibreOffice Base
4. Export in KI-ready Formate für die Verwendung in KI-Anwendungen



---

## 5. Rechtssichere Umsetzung

---

Während wir uns bisher auf die technischen Aspekte der Datenaufbereitung konzentriert haben, dürfen wir einen entscheidenden Faktor nicht vergessen: die rechtlichen Rahmenbedingungen. Datenqualität ist wichtig – aber Datenkonformität ist unerlässlich.

In der Welt der Daten und KI gibt es zwei rechtliche Rahmenwerke, die besondere Beachtung verdienen: die Datenschutz-Grundverordnung (DSGVO) und der neue EU AI Act. Lassen Sie uns diese nicht als lästige Hürden betrachten, sondern als Leitplanken, die uns helfen, verantwortungsvoll mit Daten umzugehen.

### 5.1 DSGVO-konforme Datenverarbeitung

Die DSGVO ist seit 2018 in Kraft und hat die Art und Weise, wie wir mit personenbezogenen Daten umgehen, grundlegend verändert. Für KI-Projekte ist sie besonders relevant, da KI-Systeme oft große Mengen personenbezogener Daten verarbeiten.

Die DSGVO basiert auf sechs Grundprinzipien, die auch für die Datenaufbereitung für KI-Anwendungen gelten:

#### 1. Rechtmäßigkeit, Verarbeitung nach Treu und Glauben, Transparenz

Für jede Verarbeitung personenbezogener Daten muss eine Rechtsgrundlage vorliegen – sei es eine Einwilligung, ein Vertrag, eine rechtliche Verpflichtung oder ein berechtigtes Interesse. Die Verarbeitung muss fair und für die betroffene Person nachvollziehbar sein.

Für KI-Anwendungen bedeutet dies: Klären Sie vor der Datenaufbereitung, ob Sie die Daten für den geplanten Zweck überhaupt verwenden dürfen. Informieren Sie die betroffenen Personen transparent über die KI-gestützte Verarbeitung ihrer Daten.

## **2. Zweckbindung**

Daten dürfen nur für festgelegte, eindeutige und legitime Zwecke erhoben werden und nicht in einer mit diesen Zwecken unvereinbaren Weise weiterverarbeitet werden.

© 2025 AS Digital Marketing | KI.Akademie - Modul 3: Datenbasis aufbauen - Qualität vor Quantität

Für KI-Anwendungen bedeutet dies: Definieren Sie klar, wofür Sie die Daten in Ihren KI-Systemen verwenden möchten, und halten Sie sich an diese Zwecke. Eine nachträgliche Zweckänderung erfordert in der Regel eine neue Rechtsgrundlage.

## **3. Datenminimierung**

Es dürfen nur die personenbezogenen Daten verarbeitet werden, die für den jeweiligen Zweck tatsächlich erforderlich sind – nicht mehr und nicht weniger.

Für KI-Anwendungen bedeutet dies: Prüfen Sie kritisch, welche Daten Sie wirklich für Ihre KI-Modelle benötigen. Oft können Sie mit weniger oder anonymisierten Daten arbeiten und trotzdem gute Ergebnisse erzielen.

## 4. Richtigkeit

Personenbezogene Daten müssen sachlich richtig und erforderlichenfalls auf dem neuesten Stand sein. Es sind alle angemessenen Maßnahmen zu treffen, damit unrichtige Daten gelöscht oder berichtigt werden.

Für KI-Anwendungen bedeutet dies: Die Datenqualität ist nicht nur für die Leistung Ihrer KI wichtig, sondern auch eine rechtliche Anforderung. Regelmäßige Datenbereinigung und -aktualisierung sind Pflicht.

## 5. Speicherbegrenzung

Personenbezogene Daten dürfen nur so lange gespeichert werden, wie es für die Zwecke, für die sie verarbeitet werden, erforderlich ist.

Für KI-Anwendungen bedeutet dies: Definieren Sie klare Löschfristen für Ihre Daten und implementieren Sie Prozesse zur regelmäßigen Überprüfung, ob Daten noch benötigt werden.

## 6. Integrität und Vertraulichkeit

Personenbezogene Daten müssen so verarbeitet werden, dass eine angemessene Sicherheit gewährleistet ist, einschließlich Schutz vor unbefugter oder unrechtmäßiger Verarbeitung und vor Verlust, Zerstörung oder Schädigung.

Für KI-Anwendungen bedeutet dies: Implementieren Sie angemessene technische und organisatorische Maßnahmen zum Schutz der Daten, insbesondere wenn Sie mit sensiblen personenbezogenen Daten arbeiten.

*Ein Kunde von mir, ein mittelständisches Dienstleistungsunternehmen, wollte KI für die Analyse von Kundeninteraktionen einsetzen, um den Service zu verbessern. In den E-Mails und Chat-Protokollen fanden sich jedoch zahlreiche personenbezogene Daten, die für die Analyse nicht relevant waren.*

*Statt die Rohdaten direkt in die KI zu füttern, implementierten wir einen Pseudonymisierungsprozess: Namen wurden durch Codes ersetzt, E-Mail-Adressen und Telefonnummern entfernt, und andere identifizierende Merkmale wurden generalisiert. Die KI konnte trotzdem Muster in den Interaktionen erkennen, ohne dass die Identität der Kunden preisgegeben wurde.*

© 2023 AS Digital Marketing | KI-Akademie – Modul 9: Datenbasis aufbauen – Qualität vor Quantität

*Dieser Ansatz erfüllte nicht nur die DSGVO-Anforderungen, sondern führte auch zu besseren KI-Ergebnissen, da das Modell sich auf die relevanten Inhalte konzentrieren konnte, statt durch persönliche Details abgelenkt zu werden.*

Für die praktische Umsetzung der DSGVO-Anforderungen in KI-Projekten empfehle ich folgende Maßnahmen:

**Anonymisierung/Pseudonymisierung:** Wann immer möglich, anonymisieren oder pseudonymisieren Sie personenbezogene Daten vor der Verwendung in KI-Systemen. Dies reduziert das Risiko und kann in vielen Fällen die Anforderungen der DSGVO erheblich vereinfachen.

**Datenschutz-Folgenabschätzung (DSFA):** Für KI-Systeme mit hohem Risiko für die Rechte und Freiheiten natürlicher Personen ist eine DSFA durchzuführen. Dies hilft, Risiken frühzeitig zu erkennen und zu minimieren.

**Transparenz:** Informieren Sie betroffene Personen klar und verständlich über die KI-gestützte Verarbeitung ihrer Daten, einschließlich der Logik und der Tragweite automatisierter Entscheidungen.

**Auftragsverarbeitungsverträge (AVV):** Wenn Sie externe KI-Dienstleister einsetzen, schließen Sie AVVs ab, die die Verarbeitung personenbezogener Daten regeln.

**Drittlandtransfers:** Bei der Nutzung von Cloud-KI-Diensten aus Drittländern (z.B. USA) müssen angemessene Garantien für den Datenschutz geschaffen werden, etwa durch Standardvertragsklauseln.

**Dokumentation:** Dokumentieren Sie alle Datenverarbeitungsprozesse, einschließlich der Rechtsgrundlagen, Zwecke, Datenkategorien und Sicherheitsmaßnahmen.

**DSGVO-Tipp für KI-Tools:** Bei der Nutzung von Cloud-basierten Tools wie Google Sheets oder ChatGPT ist besondere Vorsicht geboten. Personenbezogene Daten sollten vor der Bearbeitung anonymisiert werden. Bevorzugen Sie Tools mit EU-Servern und prüfen Sie die Auftragsverarbeitungsverträge (AVV) auch für kostenlose Tools.

## 5.2 AI Act: Die neuen Spielregeln für KI

Während die DSGVO den Umgang mit personenbezogenen Daten regelt, adressiert der neue EU AI Act spezifisch die Entwicklung und Nutzung von KI-Systemen. Obwohl er noch nicht vollständig in Kraft ist, lohnt es sich, die Grundprinzipien zu verstehen und sich frühzeitig darauf vorzubereiten.

Der AI Act folgt einem risikobasierten Ansatz, der KI-Systeme in verschiedene Kategorien einteilt:

**Unannehmbares Risiko:** KI-Anwendungen, die als klare Bedrohung für die Sicherheit, die Lebensgrundlagen und die Rechte der Menschen angesehen werden, sind verboten. Dazu gehören KI-Systeme, die Menschen manipulieren oder Social Scoring-Systeme.

© 2025 AS Digital Marketing | KI.Akademie - Modul 3: Datenbasis aufbauen - Qualität vor Quantität  
**Hohes Risiko:** KI-Systeme, die in kritischen Infrastrukturen eingesetzt werden oder die Grundrechte beeinflussen können, unterliegen strengen Anforderungen an Transparenz, Dokumentation, menschliche Aufsicht und Robustheit.

**Begrenztes Risiko:** KI-Systeme mit spezifischen Transparenzanforderungen, wie Chatbots, müssen als KI gekennzeichnet werden, damit Nutzer informierte Entscheidungen treffen können.

**Minimales Risiko:** Die große Mehrheit der KI-Systeme fällt in diese Kategorie und kann ohne zusätzliche rechtliche Verpflichtungen entwickelt und genutzt werden.

Für die Datenaufbereitung sind besonders die Anforderungen an Hochrisiko-KI-Systeme relevant:

- Verwendung hochwertiger Trainingsdaten zur Minimierung von Risiken und diskriminierenden Ergebnissen
- Dokumentation der verwendeten Daten und Prozesse
- Detaillierte technische Dokumentation mit allen notwendigen Informationen
- Aufzeichnung der Aktivitäten des Systems für die Rückverfolgbarkeit
- Transparenz gegenüber den Nutzern über Fähigkeiten und Grenzen des Systems
- Menschliche Aufsicht zur Minimierung von Risiken

Auch wenn Ihr KI-System nicht als Hochrisiko eingestuft wird, ist es ratsam, die Grundprinzipien des AI Acts zu beachten – sie repräsentieren Best Practices für verantwortungsvolle KI-Entwicklung und -Nutzung.

**Vorbereitung auf den AI Act:** Bereiten Sie sich frühzeitig auf diese Anforderungen vor, indem Sie Ihre KI-Anwendungen inventarisieren und deren Risikopotenzial bewerten. Dokumentieren Sie, welche Daten wie verarbeitet werden, und aktualisieren Sie Ihr Datenverarbeitungsverzeichnis entsprechend.

Die rechtssichere Umsetzung von KI-Projekten mag auf den ersten Blick komplex erscheinen, aber sie bietet auch Chancen: Unternehmen, die Datenschutz und KI-Ethik ernst nehmen, gewinnen das Vertrauen ihrer Kunden und Mitarbeiter. Zudem führen die Anforderungen an Datenqualität und Dokumentation oft zu besseren, robusteren KI-Systemen.

Betrachten Sie DSGVO und AI Act nicht als Hindernisse, sondern als Qualitätssiegel für Ihre Daten und KI-Anwendungen. Ein strukturierter, dokumentierter Ansatz zur Datenaufbereitung erfüllt nicht nur rechtliche Anforderungen, sondern verbessert auch die Leistung und Akzeptanz Ihrer KI-Systeme.

*"Datenschutz und KI-Ethik sind keine Gegensätze zur Innovation – sie sind ihre Voraussetzung in einer vertrauenswürdigen digitalen Welt."*

Im nächsten Kapitel werden wir uns konkreten Praxisbeispielen zuwenden – Erfolgsgeschichten von Unternehmen, die durch hochwertige Datenaufbereitung beeindruckende KI-Ergebnisse erzielt haben.



---

## 6. Praxisbeispiele und Anwendungsfälle

---

Theorie ist wichtig, aber nichts überzeugt so sehr wie erfolgreiche Praxisbeispiele. In diesem Kapitel möchte ich Ihnen Geschichten von Unternehmen erzählen, die durch gezielte Datenaufbereitung beeindruckende Ergebnisse mit KI erzielt haben. Diese Beispiele stammen aus meiner Beratungspraxis, wobei ich die Namen und einige Details geändert habe, um die Vertraulichkeit zu wahren.

### 6.1 Branchenspezifische Erfolgsgeschichten

#### Handwerk: Vom Zettelchaos zur intelligenten Ressourcenplanung

Stellen Sie sich einen Elektrohandwerksbetrieb mit 15 Mitarbeitern vor. Aufträge wurden auf Papierzetteln notiert, Kundendaten existierten teilweise im CRM, teilweise in Excel und teilweise nur in den Köpfen langjähriger Mitarbeiter. Die Einsatzplanung erfolgte mit Magnettafel und Bauchgefühl.

Der Inhaber wollte KI nutzen, um die Ressourcenplanung zu optimieren – doch die ersten Versuche scheiterten kläglich. Die KI konnte mit den fragmentierten, inkonsistenten Daten nichts anfangen.

Wir begannen mit einer gründlichen Bestandsaufnahme und identifizierten vier kritische Datenbereiche: Kundendaten, Auftragsdaten, Mitarbeiterdaten (Qualifikationen, Verfügbarkeiten) und historische Zeiterfassungsdaten. In einem strukturierten Prozess wurden diese Daten bereinigt, standardisiert und in einer relationalen Datenbank zusammengeführt.

Besonders herausfordernd waren die Auftragsdaten – jeder Mitarbeiter hatte seine eigene Art, Aufträge zu beschreiben. Wir entwickelten ein standardisiertes Kategorisierungssystem mit klaren Auftragsstypen, erforderlichen Qualifikationen und typischen Zeitaufwänden.

Das Ergebnis? Eine KI-gestützte Ressourcenplanungssoftware, die Aufträge basierend auf Mitarbeiterqualifikationen, Standorten und historischen Zeitdaten optimal plant. Der Betrieb konnte 23% mehr Aufträge pro Tag abwickeln, die Fahrtzeiten um 18% reduzieren und die Kundenzufriedenheit durch pünktlichere Termine deutlich steigern.

© 2025 AS Digital Marketing | KI.Akademie - Modul 3: Datenbasis aufbauen - Qualität vor Quantität

### **Einzelhandel: Personalisierte Empfehlungen durch saubere Produktdaten**

Ein mittelständischer Online-Händler für Bürobedarf wollte KI für personalisierte Produktempfehlungen nutzen. Sein Produktkatalog umfasste über 5.000 Artikel – mit inkonsistenten Kategorien, fehlenden Attributen und uneinheitlichen Produktbeschreibungen.

Die ersten KI-Empfehlungen waren bestenfalls zufällig, schlimmstenfalls absurd: Druckerpatronen für inkompatible Drucker, Bürostühle als Zubehör zu Heftgeräten, Luxus-Füllfederhalter für Kunden, die ausschließlich Basisbürobedarf kauften.

Wir starteten ein umfassendes Projekt zur Produktdatenbereinigung:

1. Entwicklung einer konsistenten Kategorisierungshierarchie mit klaren Definitionen
2. Definition von Pflichtattributen für jede Produktkategorie
3. Standardisierung von Produktbeschreibungen mit einheitlicher Struktur und Terminologie
4. Explizite Erfassung von Kompatibilitätsinformationen und Produktbeziehungen
5. Anreicherung mit Nutzungskontexten und typischen Anwendungsfällen

Die Arbeit war intensiv – etwa 80% der Projektzeit floss in die Datenaufbereitung, nur 20% in die eigentliche KI-Implementierung. Doch die Ergebnisse waren beeindruckend: 34% höhere Conversion-Rate bei Produktempfehlungen, 27% Steigerung des durchschnittlichen Warenkorbs und deutlich reduzierte Retouren durch passendere Empfehlungen.

Der Händler erkannte: Die Investition in Datenqualität zahlte sich nicht nur für die KI aus, sondern verbesserte auch viele andere Geschäftsprozesse – von der Lagerhaltung bis zum Produktmarketing.

### **Dienstleistung: Wissensmanagement durch strukturierte Dokumentation**

Eine Steuerberatungskanzlei mit 25 Mitarbeitern stand vor einer typischen Herausforderung: Das Wissen war ungleich verteilt, Expertenwissen zu speziellen Steuerfragen war oft nur in den Köpfen erfahrener Mitarbeiter vorhanden. Die Kanzlei wollte KI nutzen, um dieses Wissen zugänglicher zu machen und neue Mitarbeiter schneller einzuarbeiten.

Die Ausgangssituation war ernüchternd: Fachwissen war verteilt auf E-Mails, Word-Dokumente, handschriftliche Notizen und verschiedene Fachanwendungen. Eine einheitliche Struktur oder Terminologie existierte nicht.

Wir entwickelten einen strukturierten Ansatz zur Wissenserfassung:

1. Definition einer klaren Taxonomie für steuerliche Fachthemen
2. Entwicklung von Vorlagen für die Dokumentation von Fachwissen
3. Systematische Erfassung von Fallbeispielen, Entscheidungen und

#### Lösungsansätzen

© 2025 AS Digital Marketing | KI.Akademie - Modul 3: Datenbasis aufbauen - Qualität vor Quantität

4. Verknüpfung mit relevanten Gesetzestexten, Urteilen und Kommentaren

Die strukturierten Wissensdokumente bildeten die Grundlage für ein KI-gestütztes Assistenzsystem, das Mitarbeitern bei spezifischen Steuerfragen relevante Informationen, Präzedenzfälle und Lösungsansätze vorschlägt. Neue Mitarbeiter konnten schneller produktiv werden, und auch erfahrene Berater profitierten von der schnellen Verfügbarkeit relevanter Informationen.

Ein besonderer Erfolg: Die Bearbeitungszeit für komplexe Steuerfälle sank um durchschnittlich 22%, während die Qualität der Beratung durch konsistentere Anwendung von Fachwissen stieg.

---

## 6.2 Vom Datenchaos zum KI-Erfolg: Zwei Wege

Die vorgestellten Beispiele zeigen zwei grundlegende Ansätze zur Verbesserung der Datenqualität für KI-Anwendungen:

### **Der umfassende Ansatz: Datenqualität als strategisches Projekt**

Beim umfassenden Ansatz wird die Verbesserung der Datenqualität als strategisches Projekt mit klarem Anfang und Ende betrachtet. Typische Schritte sind:

1. Umfassende Bestandsaufnahme aller relevanten Datenquellen
2. Entwicklung eines Zielbildes für die Datenstruktur und -qualität
3. Systematische Bereinigung und Strukturierung aller Daten
4. Migration in neue, optimierte Datenhaltungssysteme
5. Implementierung von Prozessen zur Qualitätssicherung

Dieser Ansatz eignet sich besonders für Unternehmen mit überschaubaren Datenmengen oder bei grundlegenden Veränderungen wie der Einführung neuer Systeme. Er erfordert eine höhere initiale Investition, bietet aber einen "sauberen Schnitt" und eine solide Grundlage für zukünftige Datenarbeit.

Ein Beispiel ist der Elektrohandwerksbetrieb, der seine gesamte Datenlandschaft neu strukturierte, bevor er KI für die Ressourcenplanung einsetzte.

### **Der iterative Ansatz: Schrittweise Verbesserung mit schnellen Erfolgen**

Beim iterativen Ansatz wird die Datenqualität schrittweise verbessert, beginnend mit den Bereichen, die den größten Nutzen versprechen. Typische Schritte sind:

1. Identifikation des kritischsten Datenbereichs für die geplante KI-Anwendung
2. Fokussierte Bereinigung und Strukturierung dieses Bereichs
3. Implementierung der KI-Anwendung mit den verbesserten Daten

4. Messung der Ergebnisse und Identifikation des nächsten Verbesserungsbereichs
5. Kontinuierliche Erweiterung auf weitere Datenbereiche

Dieser Ansatz eignet sich besonders für Unternehmen mit großen, komplexen Datenlandschaften oder wenn schnelle Erfolge wichtig sind. Er erfordert weniger initiale Investition und liefert früher sichtbare Ergebnisse, kann aber zu "Insellösungen" führen, die später integriert werden müssen.

Ein Beispiel ist der Online-Händler, der zunächst seine Produktdaten optimierte, um personalisierte Empfehlungen zu verbessern, und erst später andere Datenbereiche wie Kundendaten und Verkaufshistorien anging.

### **Welcher Ansatz ist der richtige für Sie?**

Die Wahl zwischen umfassendem und iterativem Ansatz hängt von verschiedenen Faktoren ab:

- Größe und Komplexität Ihrer Datenlandschaft
- Verfügbare Ressourcen (Zeit, Budget, Personal)
- Dringlichkeit der KI-Implementierung
- Unternehmenskultur und Veränderungsbereitschaft

Oft ist eine Kombination beider Ansätze sinnvoll: Eine klare strategische Vision für die gesamte Datenlandschaft, umgesetzt in iterativen Schritten mit messbaren Erfolgen.

---

## 7. Aktionsplan und nächste Schritte

---

Theorie und Praxisbeispiele sind wichtig, aber am Ende zählt nur eines: Was Sie konkret in Ihrem Unternehmen umsetzen. In diesem Kapitel möchte ich Ihnen einen praktischen Aktionsplan an die Hand geben, mit dem Sie Ihre Daten Schritt für Schritt KI-ready machen können.

### 7.1 Daten-Audit: Wo stehen Sie heute?

Jede Reise beginnt mit einem klaren Verständnis des Ausgangspunkts. Bevor Sie in die Datenbereinigung und -strukturierung einsteigen, sollten Sie eine Bestandsaufnahme Ihrer aktuellen Datensituation durchführen – ein Daten-Audit.

Ein Daten-Audit ist wie eine Inventur für Ihre digitalen Bestände. Es hilft Ihnen, einen Überblick über Ihre Datenquellen, deren Qualität und die größten Herausforderungen zu gewinnen. Hier ist eine strukturierte Vorgehensweise:

#### Schritt 1: Identifizieren Sie alle relevanten Datenquellen

Erstellen Sie eine Liste aller Systeme und Orte, an denen geschäftsrelevante Daten gespeichert sind. Denken Sie dabei an:

- CRM-Systeme und Kundendatenbanken
- ERP-Systeme und Warenwirtschaft
- Excel-Tabellen und lokale Datenbanken
- E-Mail-Systeme und Kommunikationsplattformen
- Dokumentenmanagementsysteme und Fileserver
- Cloud-Speicher und SaaS-Anwendungen
- Papierbasierte Aufzeichnungen und Formulare

Vergessen Sie nicht die "versteckten" Datenquellen – Excel-Tabellen auf lokalen Rechnern, persönliche Notizen oder Wissen, das nur in den Köpfen langjähriger Mitarbeiter existiert.

## **Schritt 2: Bewerten Sie die Qualität jeder Datenquelle**

© 2025 AS Digital Marketing | KI.Akademie - Modul 3: Datenbasis aufbauen - Qualität vor Quantität

Für jede identifizierte Datenquelle sollten Sie die Qualität anhand verschiedener Kriterien bewerten:

- **Vollständigkeit:** Sind alle notwendigen Informationen vorhanden?
- **Genauigkeit:** Sind die Daten korrekt und aktuell?
- **Konsistenz:** Werden einheitliche Formate und Bezeichnungen verwendet?
- **Struktur:** Folgen die Daten einem klaren Schema?
- **Zugänglichkeit:** Wie leicht können die Daten abgerufen und verarbeitet werden?
- **Dokumentation:** Sind die Datenstrukturen und -inhalte dokumentiert?

Eine einfache Bewertungsskala von 1 (sehr schlecht) bis 5 (ausgezeichnet) kann helfen, einen schnellen Überblick zu gewinnen.

## Schritt 3: Identifizieren Sie die größten Probleme und Herausforderungen

Basierend auf Ihrer Bewertung können Sie nun die größten Probleme und Herausforderungen identifizieren:

- Welche Datentypen sind am chaotischsten?
- Wo gibt es die meisten Duplikate oder Inkonsistenzen?
- Welche Daten sind besonders veraltet oder unvollständig?
- Wo fehlen wichtige Verknüpfungen zwischen Datensätzen?
- Welche Daten sind für Ihre geplanten KI-Anwendungen besonders kritisch?

Diese Analyse hilft Ihnen, Prioritäten zu setzen und die Bereiche zu identifizieren, die zuerst angegangen werden sollten.

## Schritt 4: Setzen Sie Prioritäten basierend auf Geschäftswert und Aufwand

Nicht alle Datenprobleme sind gleich wichtig oder gleich aufwändig zu lösen. Bewerten Sie jedes identifizierte Problem anhand von zwei Dimensionen:

- **Geschäftswert:** Wie wichtig sind diese Daten für Ihre KI-Anwendungen und Geschäftsprozesse?
- **Aufwand:** Wie komplex und zeitaufwändig ist die Bereinigung und Strukturierung dieser Daten?

Diese Bewertung können Sie in einer einfachen Matrix visualisieren:

<b>Hoher Wert, Geringer Aufwand</b>  Sofort angehen!  Diese "Low-Hanging Fruits" bieten schnelle Erfolge mit hohem ROI.	<b>Hoher Wert, Hoher Aufwand</b>  Strategisch planen  Diese Bereiche erfordern sorgfältige Planung und ausreichende
---	---

Ressourcen.

### **Geringer Wert, Geringer Aufwand**

Nebenbei erledigen

Diese Aufgaben können bei Gelegenheit oder in Leerlaufzeiten angegangen werden.

### **Geringer Wert, Hoher Aufwand**

Zurückstellen

Diese Bereiche sollten nur angegangen werden, wenn alle wichtigeren Aufgaben erledigt sind.

Mit dieser Priorisierung haben Sie einen klaren Fahrplan für Ihre Datenqualitätsinitiative.

## **Daten-Audit-Checkliste**

Verwenden Sie die folgende Checkliste, um Ihr Daten-Audit zu strukturieren:

- Haben Sie alle relevanten Datenquellen identifiziert?
- Haben Sie die Qualität jeder Datenquelle anhand klarer Kriterien bewertet?
- Haben Sie die größten Probleme und Herausforderungen identifiziert?
- Haben Sie die Probleme nach Geschäftswert und Aufwand priorisiert?
- Haben Sie einen klaren Fahrplan für die nächsten Schritte?

## 7.2 30-Tage-Plan zur Datenqualitätsverbesserung

Mit den Erkenntnissen aus Ihrem Daten-Audit können Sie nun einen konkreten Aktionsplan für die nächsten 30 Tage erstellen. Dieser Plan sollte ambitioniert, aber realistisch sein – mit klaren Zielen, Verantwortlichkeiten und Meilensteinen.

Hier ist ein Beispielplan, den Sie an Ihre spezifische Situation anpassen können:

### Woche 1: Analyse und Planung

#### *Tag 1-2: Team und Ressourcen*

- Stellen Sie ein kleines, schlagkräftiges Team zusammen
- Definieren Sie klare Rollen und Verantwortlichkeiten
- Stellen Sie sicher, dass alle notwendigen Zugriffsrechte vorhanden sind
- Installieren Sie benötigte Tools (OpenRefine, LibreOffice Base)

#### *Tag 3-5: Detaillierte Analyse und Planung*

- Vertiefen Sie die Analyse der priorisierten Datenbereiche
- Entwickeln Sie ein Zielbild für die Datenstruktur
- Definieren Sie klare Qualitätskriterien und Erfolgskennzahlen
- Erstellen Sie einen detaillierten Projektplan mit Meilensteinen

### Woche 2: Pilotprojekt - Datenbereinigung

#### *Tag 6-7: Vorbereitung und Export*

- Wählen Sie einen überschaubaren Datensatz für das Pilotprojekt
- Exportieren Sie die Daten in ein bearbeitbares Format
- Erstellen Sie Sicherungskopien der Originaldaten
- Dokumentieren Sie den Ausgangszustand für spätere Vergleiche

### *Tag 8-10: Grundlegende Bereinigung*

- Identifizieren und entfernen Sie Duplikate
- Standardisieren Sie Formate und Schreibweisen

© 2025 AS Digital Marketing | KI-Akademie - Modul 8: Datenbasis aufbauen - Qualität vor Quantität

- Behandeln Sie fehlende oder offensichtlich falsche Werte
- Führen Sie erste Qualitätsprüfungen durch

### **Woche 3: Pilotprojekt - Strukturierung**

#### *Tag 11-12: Datenmodellierung*

- Definieren Sie ein klares Datenmodell mit Entitäten und Beziehungen
- Legen Sie Datentypen und Validierungsregeln fest
- Entwickeln Sie eine konsistente Terminologie und Namenskonventionen
- Dokumentieren Sie das Datenmodell für zukünftige Referenz

#### *Tag 13-15: Strukturierung und Validierung*

- Überführen Sie die bereinigten Daten in die neue Struktur
- Erstellen Sie Verknüpfungen zwischen zusammengehörigen Datensätzen
- Führen Sie umfassende Validierungstests durch
- Beheben Sie identifizierte Probleme und Inkonsistenzen

## Woche 4: Implementierung und Ausweitung

### Tag 16-18: Integration und Tests

- Integrieren Sie die strukturierten Daten in Ihre Produktivsysteme
- Führen Sie erste Tests mit Ihren geplanten KI-Anwendungen durch
- Messen Sie die Verbesserung anhand der definierten Erfolgskennzahlen
- Sammeln Sie Feedback von Anwendern und Stakeholdern

### Tag 19-20: Dokumentation und Prozessoptimierung

- Dokumentieren Sie den gesamten Bereinigungsprozess
- Entwickeln Sie Richtlinien für die zukünftige Datenpflege
- Schulen Sie relevante Mitarbeiter in den neuen Prozessen
- Implementieren Sie Qualitätssicherungsmaß
- Implementieren Sie Qualitätssicherungsmaßnahmen für neue Dateneingaben

### Tag 21-30: Ausweitung und nächste Schritte

- Identifizieren Sie den nächsten Datenbereich für die Bereinigung
- Wenden Sie die entwickelten Methoden und Prozesse an
- Beginnen Sie mit der Planung langfristiger Datenqualitätsinitiativen
- Feiern Sie die erreichten Erfolge und kommunizieren Sie sie im Unternehmen

*Ein Kunde von mir, ein mittelständischer Großhändler, folgte einem ähnlichen 30-Tage-Plan für seine Produktdaten. Am Ende des Monats hatte er nicht nur einen sauberen, strukturierten Produktkatalog, sondern auch neue Prozesse, die sicherstellten, dass die Datenqualität langfristig erhalten blieb.*

*Der wichtigste Erfolgsfaktor war nicht die Technologie, sondern die Einbindung der Mitarbeiter. Indem er ihnen den Wert hochwertiger Daten vermittelte und sie aktiv in den Prozess einbezog, schuf er eine "Datenqualitätskultur", die weit über das initiale Projekt hinaus Bestand hatte.*

© 2025 AS Digital Marketing | KI.Akademie - Modul 3: Datenbasis aufbauen - Qualität vor Quantität

#### **Community-Aufgabe bis Modul 4:**

Wählen Sie einen überschaubaren Datensatz aus Ihrem Unternehmen – eine Kundenliste, einen Produktkatalog oder eine andere Datensammlung, die für Ihre KI-Anwendungen relevant ist. Wenden Sie die gelernten Methoden an, um diesen Datensatz zu bereinigen und zu strukturieren.

Dokumentieren Sie Ihren Prozess und die Ergebnisse: Was hat sich verbessert? Welche Herausforderungen sind aufgetreten? Welche Erkenntnisse haben Sie gewonnen?

Bringen Sie Ihre Erfahrungen zum nächsten Modul mit – sie werden eine wertvolle Grundlage für die praktische Arbeit mit KI-Anwendungen sein.

Der Weg zu hochwertigen, KI-ready Daten ist keine einmalige Reise, sondern ein kontinuierlicher Prozess. Mit jedem Schritt, den Sie auf diesem Weg gehen, verbessern Sie nicht nur die Grundlage für Ihre KI-Anwendungen, sondern schaffen auch Mehrwert für viele andere Bereiche Ihres Unternehmens.

Denken Sie daran: Qualität vor Quantität. Es ist besser, einen kleinen Datensatz in exzellenter Qualität zu haben als einen großen Datensatz voller Fehler und Inkonsistenzen. Beginnen Sie klein, lernen Sie aus Ihren Erfahrungen und bauen Sie schrittweise auf Ihren Erfolgen auf.

*"Der beste Zeitpunkt, Ihre Datenqualität zu verbessern, war vor zehn Jahren. Der zweitbeste Zeitpunkt ist heute."*

---

# Ausblick auf Modul 4: KI-Anwendungen im Alltag

---

Nachdem wir uns in diesem Modul intensiv mit der Vorbereitung Ihrer Daten beschäftigt haben, werden wir im nächsten Modul den nächsten logischen Schritt gehen: die praktische Implementierung von KI-Anwendungen im Unternehmensalltag.

Modul 4 "KI-Anwendungen im Alltag - Praktische Implementierung" wird sich mit folgenden Themen beschäftigen:

- **Auswahl der richtigen KI-Tools** für Ihre spezifischen Anforderungen
- **Integration von KI in bestehende Geschäftsprozesse** mit minimaler Disruption
- **Praktische Anwendungsfälle** für verschiedene Unternehmensbereiche
- **Messung und Optimierung** des ROI Ihrer KI-Investitionen
- **Change Management** und Mitarbeiterakzeptanz

Wir werden mit Ihren bereinigten Datensätzen arbeiten und konkrete KI-Anwendungen umsetzen. Bringen Sie Ihre aufbereiteten Daten mit, um direkt in die praktische Anwendung einzusteigen.

**Termin:** 15. Oktober 2025, 17:00-18:00 Uhr

**Ort:** Online via BigBlueButton

**Vorbereitung:** Bitte bearbeiten Sie die Community-Aufgabe und bringen Sie Ihre Erfahrungen und Fragen mit.

Ich freue mich darauf, Ihre Fortschritte zu sehen und gemeinsam mit Ihnen den nächsten Schritt auf Ihrer KI-Reise zu gehen!

© 2025 AS Digital Marketing | KI-Akademie - Modul 3: Datenbasis aufbauen - Qualität vor Quantität

### Arno Schimmelpfennig



KI-Experte & Digitalisierungsberater

 Amalienstraße 11, 96047 Bamberg

 0172 271 26 00

 [info@arno-schimmelpfennig.de](mailto:info@arno-schimmelpfennig.de)

 [as-digitalmarketing.de](http://as-digitalmarketing.de)

 [linkedin.com/in/arno-schimmelpfennig](https://linkedin.com/in/arno-schimmelpfennig)

## Vielen Dank für Ihre Teilnahme!

Ich freue mich auf das nächste Modul und die gemeinsame Arbeit an Ihren KI-Projekten.