Bei diesem Beitrag handelt es sich um einen wissenschaftlich begutachteten und freigegebenen Fachaufsatz („reviewed paper“).

doi.org/10.37544/1436-4980-2025-11-12-35
Datum der Einreichung: 15.09.2025
Datum der Annahme: 10.11.2025
Datum der Veröffentlichung: 17.12.2025

Künstliche Intelligenz im Fehlermanagement

Datenqualität bei der Informationsbereitstellung durch LLMs

C. Wachter, C. Schönekehs, M. Padrón Hinrichs, F. Sohnius, R. H. Schmitt

ZUSAMMENFASSUNG Die Qualität textueller Daten im Produktionskontext stellt einen noch unzureichend betrachteten Stellhebel zur Etablierung LLM-basierter Informationsbereitstellungssysteme dar. Dieser Beitrag zeigt Herausforderungen in der LLM-basierten Informationsverarbeitung hinsichtlich der Datenqualität auf und leitet Handlungsempfehlungen für die Gestaltung textueller Dokumente im Produktionskontext ab.

STICHWÖRTER
Datenqualität, Künstliche Intelligenz (KI), Qualitätsmanagement

AI in Failure Management - Data Quality for LLM-based Information Retrieval

ABSTRACT The quality of textual data in a production context represents an influencing factor for establishing LLM-based information retrieval systems that has not yet been sufficiently considered. This article outlines the limitations of LLM-based information processing regarding data quality and provides recommendations for the design of textual documents in a production context.

1 Einleitung

Im Kontext des Qualitätswesens kommt dem Fehlermanagement eine zentrale Bedeutung für die Steigerung der Wirtschaftlichkeit von produzierenden Unternehmen zu. Das Fehlermanagement umfasst unter anderem die Fehleridentifikation, die Ursachenanalyse, die Ableitung wirksamer Maßnahmen und die Sicherung des gewonnenen fehlerbezogenen Wissens. Ein effektives Fehlermanagement trägt zur Verbesserung der Prozessqualität, zur Vermeidung wiederkehrender Fehler und zur Steigerung der Gesamtleistung des Unternehmens bei. [1, 2] Von entscheidender Bedeutung ist eine konsequente Rückführung von Erfahrungswissen, das es ermöglicht, vergangenes Fehlerwissen gezielt für die Prävention bekannter Fehler und die Ursachenanalyse neuartiger Fehler einzusetzen [3]. Trotz der häufig umfangreichen historischen Fehlerdokumentationen, deren Berücksichtigung zur Effizienz der Fehlerursachenanalyse beitragen könnte, bleiben diese Informationen in Unternehmen häufig ungenutzt. Ein wesentlicher Grund dafür ist ihre oft unstrukturierte oder semistrukturierte Darstellung, die eine aufwendige, manuelle Auswertung erfordert und somit deren Integration in die Fehlermanagementprozesse erschwert. [4, 5]

Gleichzeitig gewinnen große Sprachmodelle (engl. Large Language Models, LLM) zunehmend an Bedeutung, da sie zur Verarbeitung textueller Daten eingesetzt werden können und eine Interaktion mittels natürlicher Sprache ermöglichen [6]. In der praktischen Anwendung stellen jedoch das probabilistische Verhalten von LLMs und deren Neigung zu Halluzinationen eine zentrale Herausforderung dar, wodurch ihre verlässliche Anwendung im industriellen Kontext eingeschränkt ist. Das probabilistische Verhalten von LLMs bezieht sich auf ihre Eigenschaft, bei der Generierung von Antworten nicht deterministisch zu sein, so dass eine gleichbleibende Anfrage kontextabhängig variierende Formulierungen oder Inhalte hervorbringen kann [7, 8]. Halluzinationen hingegen betiteln die Tendenz von LLMs zur Generation von faktisch falschen, erfundenen Informationen, insbesondere in Fällen, in denen das Modell nicht über die für die Beantwortung der Anfrage benötigten Informationen verfügt [6, 8].

Ansätze der domänenspezifischen Adaptationen der Modelle sind für eine Anwendung im Produktionskontext erforderlich. Eine umfangreiche Anpassung der Modellgewichte durch das Nachtrainieren mit zusätzlichen, fachspezifischen Daten stellt das Fine-Tuning dar, ist jedoch rechen- und datenintensiv sowie unzureichend flexibel gegenüber sich ändernden Informationsbeständen [9]. Eine für den Produktionskontext vorteilhaftere Alternative bietet das In-Context Learning (ICL). Bei diesem Ansatz erhält das LLM zur Verarbeitung einer Nutzeranfrage zusätzliche Daten, Regeln oder Beispiele bereitgestellt. Somit werden die für die jeweilige Aufgabe relevanten Informationen oder Muster extrahiert, ohne dass eine dauerhafte Anpassung der Modellparameter erforderlich ist. [6, 9] Diese Vorgehensweise ermöglicht eine schnelle Adaption an domänen- oder unternehmensspezifische Aufgaben. Im Kontext des Fehlermanagements ermöglicht ICL das Zurückgreifen auf historische Fehlerdoku-

https://doi.org/10.37544/1436-4980-2025-11-12-35 - am 18.12.2025, 13:23:07. https://www.inlibra.com/de/agb - Open Access -

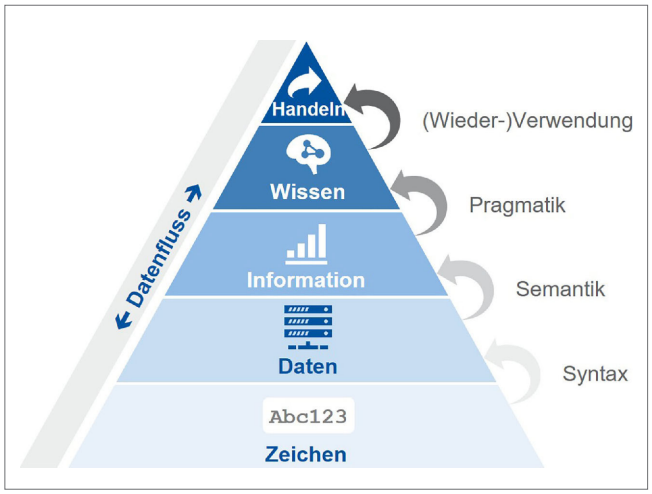

Bild 1 Wissenspyramide angelehnt an Aamodt [14].
Grafik: WZL der RWTH Aachen

mentationen, Wartungsprotokolle oder Arbeitsanweisungen und die Bereitstellung kontextrelevanter Informationen an Anwender:innen, um beispielsweise Fehlerursachen effektiver und effizienter ermitteln zu können. Der Ansatz ist jedoch durch die maximale Kontextlänge des Modells limitiert und anfällig für Qualitätsschwankungen der Modellantwort, sofern der bereitgestellte Kontext eine unzureichende Qualität aufweist [9].

Um das Risiko von Halluzinationen zu reduzieren und die Zuverlässigkeit der auf ICL basierenden Informationsbereitstellungssystemen zu steigern, stellt die Qualität der Kontextdaten einen maßgeblichen Stellhebel dar. Eine hohe Qualität der Inputdaten gewährleistet, dass der im ICL herangezogene Kontext einen positiven Einfluss auf die Genauigkeit der Modellantwort sowie die Robustheit gegenüber fehlerhaften oder irrelevanten Ausgaben aufweisen kann. [9, 10]

Für die im Produktionsumfeld typischerweise un- beziehungsweise semi-strukturierten Dokumente stellt sich daher die Frage, wie ihre Datenqualität angemessen bestimmt werden kann. Etablierte Normen wie die ISO 8000-8 oder ISO 25012 richten sich überwiegend an numerische, strukturierte Datensätze und bieten somit nur eingeschränkte Orientierung für textuelle Daten [11, 12]. Zudem sind die Anforderungen an die Datenqualität, die durch eine LLM-basierte Datenverarbeitung zu tragen kommen, unzureichend berücksichtigt [13].

Dieser Beitrag erörtert, welche spezifischen Limitationen LLMs bei der Verarbeitung textueller Daten aufzeigen und wie sich diese auf eine praxisnahe Bewertung der Datenqualität im Fehlermanagement auswirken. Durch ein umfassendes Verständnis des Qualitätsbegriffs in diesem Kontext lassen sich wirksame Maßnahmen zur Verbesserung der Datenqualität ableiten und ein Beitrag zur Etablierung LLM-basierter Informationsbereitstellung leisten.

2 Datenqualität im Kontext der LLM-basierten Datenverarbeitung

2.1 Grundlagen der Datenqualität

Für die Abgrenzung der Begriffe der Datenqualität existieren verschiedene Modelle, die in Abhängigkeit des Anwendungsfalls verwendet werden. Für die praxisnahe Bewertung der Datenqualität zur Nutzung mit LLMs sind vor allem die Begriffe Zeichen, Daten, Informationen und Wissen relevant, da für den Begriff der Datenqualität für textuelle Daten keine allumfassende Definition existiert. In der ISO 8000-8 [12] werden grundlegende Konzepte der Datenqualität aufgegriffen und die Begriffe der syntaktischen, semantischen und pragmatischen Datenqualität definiert. Die Begriffe Zeichen, Daten, Informationen und Wissen lassen sich anhand der Wissenspyramide, **Bild 1**, nach *Aamodt* [14] voneinander abgrenzen. Im Gegensatz dazu berücksichtigt die DIKW-Pyramide (Data–Information–Knowledge–Wisdom) nach Ackoff [15] die Ebene der Zeichen als syntaktische Grundlage zur Bildung von Daten nicht. Die Wissenspyramide nach *Aamodt* stellt zugleich einen Bestandteil der Wissenstreppe nach North [16] dar. Letztere erweitert das Modell um die Ebenen Kompetenz und Wettbewerbsfähigkeit, die vor allem für das strategische Wissensmanagement in Unternehmen relevant sind und daher im Rahmen dieses Beitrags nicht weiter betrachtet werden.

Daten selbst sind in der ISO 8000-8 als interpretierbare Darstellungen von formalisierten Informationen definiert, die geeignet sind zur Kommunikation, Interpretation oder Verarbeitung. Zur Bewertung der Qualität dieser Daten eignet sich ein Qualitätsbericht, der aus der Bewertung der syntaktischen, semantischen und pragmatischen Datenqualität besteht. Die syntaktische Qualität wird verstanden als der Grad, in dem Daten mit der spezifizierten Syntax (beispielsweise in den Metadaten) übereinstimmen. Somit werden aus Zeichen Daten durch die Anwendung von Syntax-Regeln zur Anordnung der Zeichen, indem die Vokabeln und grammatikalischen Regeln einer Sprache genutzt werden. Auf der darüberliegenden Ebene wird unter der semantischen Qualität die einzigartige und eindeutige Zuordnung von Daten zu ihren Entitäten verstanden. Dadurch wird Daten eine Bedeutung gegeben, indem sie durch ihre unmissverständliche Zuordnung interpretiert werden können, so dass Informationen entstehen. Die pragmatische Datenqualität wird interpretiert als die Übereinstimmung mit nutzungsabhängigen Anforderungen. Das bedeutet, dass Daten nicht nur als spezifische Informationen repräsentiert werden, sondern auch einen darüberliegenden Zweck im Rahmen des Systems erfüllen, auf dessen Grundlage Entscheidungen durch die Wissensverwendung getroffen werden können.

2.2 Herausforderungen in der LLM-basierten Datenverarbeitung

Um eine Bewertung der Qualität textueller Daten im Rahmen der LLM-basierten Informationsbereitstellung zu ermöglichen, werden im folgenden Abschnitt die Herausforderungen hinsichtlich der LLM-basierten Datenverarbeitung erörtert. Durch eine Einordnung der Herausforderungen in die Ebenen der Datenqualität erfolgt die Ableitung von Anforderungen an die Datenqualität und die Definition möglicher Datenqualitätsmetriken.

Auf der syntaktischen Ebene wird die LLM-basierte Datenverarbeitung durch grammatikalische Fehler, Rechtschreib- sowie Tastaturfehler und verarbeitungsbedingte Fehler beeinflusst. Der Einfluss grammatikalischer oder typografischer Fehler auf die LLM-generierte Antwort ist vergleichsweise gering, da diese eine hohe Repräsentation in den Trainingsdaten des Modells aufweisen. [17, 18] Verarbeitungsbedingte Abweichungen, die beispielsweise durch die Verwendung von Methoden wie Optical Charac-

https://doi.org/10.37544/1436-4980-2025-11-12-35 - am 18.12.2025, 13:23:07. https://www.inlibra.com/de/agb - Open Access -

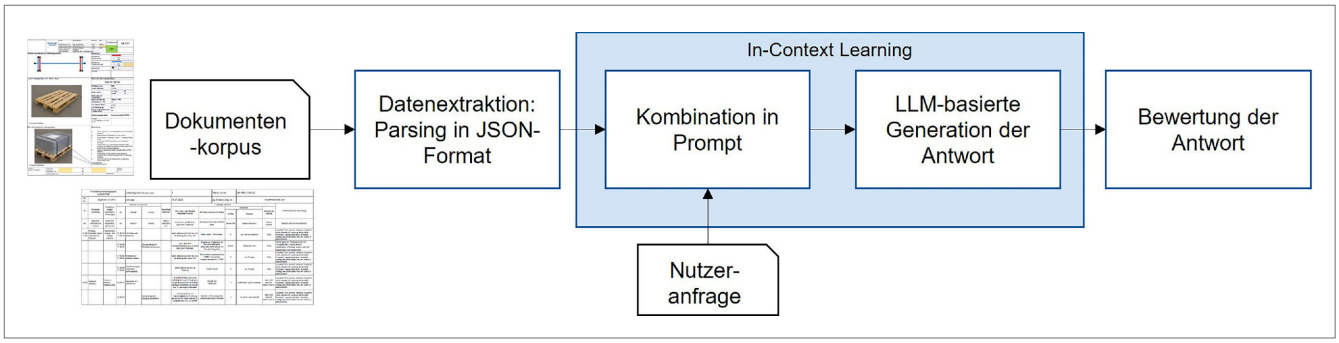

Bild 2 Datenpipeline des industriellen Anwendungsfalls. *Grafik: WZL der RWTH Aachen*

ter Recognition (OCR) entstehen, erhöhen das Risiko von Tokenisierungsfehlern in der LLM-basierten Verarbeitung [19]. Eine Betrachtung der Qualität auf syntaktischer Ebene kann durch die Bewertung der Genauigkeit der Inputdaten für das Modell erfolgen. Die syntaktische Genauigkeit ist definiert als „die Nähe der Datenwerte zu einer Menge von Werten, die in einem Bereich definiert sind, der als syntaktisch korrekt angesehen wird" [11].

Auf semantischer Ebene weisen LLMs Limitationen hinsichtlich der Verarbeitung von mehrdeutigen Begriffen, terminologischer Heterogenität, nicht im Trainingswortschatz enthaltener Wörter und nicht-maschinenlesbarer Layouts auf [13, 19, 20].

Dabei handelt es sich beispielsweise um domänenspezifische Terminologie, Fremdwörter, Abkürzungen oder unternehmensspezifische Ausdrücke. Modelle ohne domänenspezifisches Fine-Tuning weisen eine geringere Genauigkeit in der Interpretation unbekannter Wörter auf. Der negative Effekt verstärkt sich bei mehrdeutigen Begriffen, die eine unterschiedliche Bedeutung im ursprünglichen Trainingsdatensatz des Modells, dem intrinsischen Wissen, aufweisen. Ein Beispiel hierfür ist die Anwendung der Abkürzung SOP: In einem Kontext bedeutet sie „Start of Production", während sie in einem anderen „Standard Operating Procedure" entspricht. Terminologische Heterogenität innerhalb oder zwischen Dokumenten (das bedeutet, unterschiedliche Namen für dieselben Entitäten) kann eine Limitation bei der Kombination von Informationen darstellen und somit die Fähigkeit des Modells zur Informationsintegration einschränken. Nichtmaschinenlesbare Layouts können in der Verarbeitung zu Informationsverlust führen oder ihren Kontextbezug durch eine Verschiebung verlieren. Dies wiederum führt zu einer fehlerhaften semantischen Interpretation durch das LLM.

Qualitätscharakteristika, die zur Bewertung der semantischen Qualität hinzugezogen werden könnten, sind die Konsistenz, Vollständigkeit und Verständlichkeit. Konsistenz adressiert den Grad an Widersprüchen und Zusammenhängen innerhalb oder zwischen verschiedenen Textdokumenten [11]. Die Qualität der Eingabedaten wird durch die Anzahl von inkonsistent benannten Eingabedaten beeinflusst und limitiert dadurch die Interpretationsfähigkeiten des LLMs. Die Rate der Konsistenz kann auf Textebene aus der Anzahl der einheitlich benannten Entitäten und der Gesamtzahl der Entitäten bestimmt werden [21]. Weiterhin bezieht sich Konsistenz ebenso auf die Layoutgestaltung von Dokumenten. Eine einheitliche Zuordnung von Key-Value-Paaren ist notwendig, um Fehler in der Informationsextraktion zu minimieren. Die Vollständigkeit von textuellen Daten kann durch die Anzahl von zur Verfügung gestellten Definitionen bewertet werden, die sich auf die Limitationen von unternehmensspezifischer Terminologie und den Begriffen beziehen, die nicht im Vokabular enthalten sind. Somit lässt sich der Grad der Vollständigkeit durch die Anzahl der definierten Abkürzungen im Verhältnis zu den genutzten Abkürzungen bestimmen [21]. Verständlichkeit wird als Grad definiert, inwieweit Daten geeignet sind, um von Nutzern interpretiert zu werden [11]. Dies bezieht sich auf die Verwendung von mehrdeutigen Begriffen, wie zum Beispiel die doppelte Verwendung von eindeutig zuordbaren Keys.

3 Bewertung der Datenqualität am Anwendungsfall

Im Rahmen dieses Beitrags wird die beispielhafte Bewertung der Datenqualität an einem Dokumentenkorpus mit reduziertem Umfang aufgezeigt, um die Konzepte aus syntaktischer, semantischer und pragmatischer Qualität zu veranschaulichen. Es handelt sich um zwei reale, jedoch für die Darstellung verfremdete Dokumente eines Automobilzulieferers, die eine Verpackungsvorschrift (VV) und einen Produktionslenkungsplan (PLP) umfassen. Beide Dokumente beziehen sich auf dasselbe Bauteil und bilden einen der letzten Prozessschritte vor der Auslieferung ab. Der PLP enthält darüber hinaus den zum Verpackungsprozess gehörigen Kontrollprozess sowie einen Reaktionsplan bei Abweichungen und ist daher für eine Betrachtung im Rahmen des Fehlermanagements von hoher Relevanz.

Zunächst ist die Datenpipeline des Anwendungsfalls, **Bild 2**, zu erläutern. Die Daten werden in einem ersten Schritt mittels eines Parsing-Scripts extrahiert und in ein maschinenlesbares Format, hier eine JSON-file, übertragen. Aufgrund der unterschiedlichen Beschaffenheiten der Dokumente werden zwei verschiedene Extraktionsmethoden angewandt, die in der Bewertung der semantischen Qualität weiter erläutert werden. Mittels eines In-Context Learning Ansatzes werden die Nutzeranfrage sowie die extrahierten Daten im Prompt kombiniert und dem LLM zur Generierung der Antwort bereitgestellt. Letztendlich wird die auf den in dem Prompt bereitgestellten Daten generierte Antwort dem Nutzer ausgegeben. Die Zuverlässigkeit von In-Context-Learning-basierten Anwendungen kann jedoch durch die Länge des dem Modell bereitgestellten Textes beeinflusst werden. Mit zunehmender Kontextgröße steigt die Wahrscheinlichkeit von Halluzinationen, da Modelle dazu neigen, Informationen in langen Texten zu übersehen („lost in the middle"-Prinzip). Um diesem Effekt entgegenzuwirken, kann bei großen Kontextmengen eine Unterteilung in kleinere Abschnitte (Chunks) und der Einsatz eines Retrieval-Augmented-Generation-Ansatzes (RAG)

https://doi.org/10.37544/1436-4980-2025-11-12-35 - am 18.12.2025, 13:23:07. https://www.inlibra.com/de/agb - Open Access -

Produktionslenkungsplan:

Nr.	Prozessbe-schreibung	Produktion-Anlage/ Maschine(n), Werkzeug(e)	Merkmal Characteristic			Besonderes Merkmal*
			Nr.	Produkt	Prozess	
No.	Operation, description of process	Production equipment/ device, tool	No.	Product	Process	Special characteristic*
17.40 / 17.50	Prüfung Fertigteil/ series inspection of final part	Feinschneid-anlage / fine cutting machine	17.40.10 / 17.50.10	Teildicke / part thickness		

Verpackungsvorschrift:

Kunde:	Zelvix Motion	Baureihe:	EF2	CC-Merkmal	AB-3371
Bauteilbezeichnung:	Top AccuPressS	OEM:	Orvian		
Kunden-Teilenummer:	G087 123 32_01 03	Index:		nein	
Interne Teilenummer:	AB-3371-2345-42	Index:	Z-09		-
Interne Materialnr.:	12123412				
Materialkurztext:	EdgeForm, 3371 AccuPressS EF2				

Legende:

Key Value

Bild 3 Gegenüberstellung zweier Ausschnitte der Formatierung in PLP und VV. *Grafik: WZL der RWTH Aachen*

sinnvoll sein. Im vorliegenden Anwendungsfall ist dies jedoch nicht erforderlich, da die verwendeten LLMs über ein Kontextfenster von etwa 128 000 Token verfügen, während die beiden bereitgestellten Dokumente zusammen lediglich etwa 5 720 Zeichen beziehungsweise rund 2 700 Token (modellabhängig) umfassen. Diese Textmenge liegt somit deutlich unter dem maximalen Kontextlimit, weshalb die Kontextlänge in dieser Untersuchung nicht weiter berücksichtigt wird.

3.1 Bewertung der syntaktischen Qualität

Die Bewertung der syntaktischen Qualität umfasst die Betrachtung der beiden Originaldokumente hinsichtlich ihrer syntaktischen Genauigkeit. Eine quantifizierte Bewertung wird durch die Anzahl an Grammatik-, Tipp- oder Rechtschreibfehlern ermöglicht. Die Wortfehlerrate (engl. Word-Error-Rate, WER) eignet sich, um die syntaktischen Fehler auf einem Zeichenlevel zu beurteilen. Für beide Dokumente beträgt die WER null. Betrachtet man jedoch den zweiten Pipeline-Schritt, die JSON-Datenextrakte, wird deutlich, dass an dieser Stelle Qualitätsdefizite entstehen können. So kann die Datenextraktion in Sonderfällen von bspw. Umlauten zur Entstehung von Zeichenfehlern führen. Derartige verarbeitungsbedingte Fehler sind im Trainingskorpus des LLM seltener vertreten und steigern daher das Risiko von Tokenisierungsfehlern. Die Wahl einer geeigneten Parsing-Methode kann dieses Risiko reduzieren. Im vorliegenden Anwendungsfall führte die Anwendung der Libraries pandas und openpyxl zu keinen syntaktischen Fehlern.

3.2 Bewertung der semantischen Qualität

Weiterhin ist die semantische Qualität der Originaldokumente zu bewerten. Es ist die Konsistenz in der Formatierung zu betrachten, da diese die Komplexität des Extraktionsprozesses beziehungsweise die Qualität der als JSON gespeicherten Daten beeinflusst. Insbesondere Dokumente aus dem Produktionskontext sind durch eine komplexe und strukturierte Formatierung charakterisiert, deren enthaltene Informationen für eine adäquate Verarbeitung und Auswertung zu berücksichtigen sind. So ist beispielsweise die Struktur der Tabelle, die die Daten des PLP enthält, entscheidend, um eine korrekte Interpretation der Inhalte zu ermöglichen. Inhalte werden in der Form von Key-Value-Paaren, also Zuordnungen zwischen einer eindeutigen Entität (Key) und dem dazugehörigen Wert (Value), dargestellt. **Bild 3** verdeutlicht die Positionierung und Zuordnung von Key-Value-Paaren im Dokument. Uneinheitlichkeit in der Formatierung beziehungsweise Positionierung der Key-Value-Einträge erfordert, wie im Beispiel der VV, komplexere Extraktionsmethoden und erhöht das Risiko verarbeitungsbedingter Fehler.

Der Begriff der Konsistenz bezieht sich ebenso auf die Verwendung heterogener Terminologien. Die Heterogenität kann sich auf die Bezeichnungen innerhalb eines Dokuments oder über einen Dokumentenkorpus hinweg beziehen. In der VV ist aufzuführen, dass die Konsistenz in der Benennung der Entität des Bauteils vergleichsweise hoch ist: Die Bauteilbezeichnung „Top AccuPressS“ wird lediglich ein Mal auf der linken Seite der VV aufgeführt, siehe **Bild 4** (gelbe Markierung). Weitere Referenzen zur Entität des Bauteils werden durch die Bezeichnung „Platine“ auf der rechten Seite hergestellt. Der Begriff „Platine“ mit Referenz auf die Bauteilbezeichnung wird durchgängig im Dokument verwendet.

Die Konsistenz in der Benennung der Entität des Bauteils ist weiterhin dokumentenübergreifend in Bezug auf den PLP zu betrachten. Da der PLP ein produktübergreifendes Dokument darstellt und die Beschreibungen für mehrere Produkte verwendet werden, wird hier der Entität „Platine“ der Begriff „Fertigteil“, „HU“ oder „Behälter“ zugeordnet. Eine Referenz zur Bauteilbezeichnung ist nicht zu erkennen. Stattdessen sind die „interne Materialnummer“ und die „Teil-Nr.“ als Zuordnung aufgeführt (siehe grüne Markierung in Bild 4). Weiterhin unterscheiden sich die Bezeichnungen der Identifikationsnummern in beiden Dokumenten (in Bild 4 in Blau dargestellt). In der VV ist die „Interne Materialnr.“ aufgeführt, während im Prozesslenkungsplan dieselbe Nummer unter „EdgeForm“ aufgeführt ist. Daher ist die dokumentenübergreifende Konsistenz als gering zu bewerten.

Als weiteres Qualitätskriterium ist die Vollständigkeit zu betrachten. Vollständigkeit bezieht sich auf die Bereitstellung von Definitionen interner Abkürzungen oder organisations-spezifischer Terminologie. Zu nennen ist beispielsweise die Verwendung von Nomenklaturen in Bauteilbezeichnungen, deren Interpretation ohne die Bereitstellung einer Definition das intrinsische Wissen der Prozessexperten erfordert. So kann ein LLM ohne explizite Definition weder die semantische Bedeutung einer Teilenummer wie „AB-001-2345-42“ noch die in der Bezeichnung „AA Lenkung“ enthaltene Information erschließen, da diese auf implizitem Expertenwissen beruht.

Verständlichkeit bezieht sich auf die Interpretierbarkeit der domänenspezifischen Terminologie und die Verwendung mehrdeutiger Begrifflichkeiten. Die Verständlichkeit hinsichtlich der Verwendung domänenspezifischer Terminologie kann lediglich

https://doi.org/10.37544/1436-4980-2025-11-12-35 - am 18.12.2025, 13:23:07. https://www.inlibra.com/de/agb - Open Access -

Kunde:	Zelvix Motion	Baureihe:	EF2	CC-Merkmal	AB-3371
Bauteilbezeichnung:	Top AccuPressS	OEM:	Orvian		
Kunden-Teilenummer:	G087 123 32_01 03	Index:		nein	
Interne Teilenummer:	AB-3371-2345-42	Index:	Z-09		-
Interne Materialnr.:	12123412				
Materialkurztext:	EdgeForm, 3371 AccuPressS EF2				

Name / Nr. des Ladungsträgers		
Palette 450 x 1000 x 300		
SAP-Bezeichnung:	P900	
Platinen-Stapelhöhe:	500 mm	
Anzahl Platinen	pro Stapel:	
	Gesamt:	
Anzahl Stapel pro Ladungsträger:	1	
Gewicht des Gebindes:	1000 kg + TARA	
Stapelfaktor (voll / leer):	1+0	
max. Stapelschiefstand:	10 mm	
max. Platinenversatz:	5 mm	

Produktionslenkungsplan Control Plan		Änderungsindex Revision index :	1	Teil-Nr. Part No. :	AB-0001-2345-42
Nr. No.:	EdgeForm: 12123412	vom date :	01.07.2025	Zg.-Ä-Stand Draw. rel .:	TLD-DFR-08/03.05.2017

Nr.	Prozessbe-schreibung	Produktion-Anlage/ Maschine(n), Werkzeug(e)	Merkmal Characteristic			Besonderes Merkmal*	Methoden Methods		Stichprobe		Methode der Lenkung	Reaktionsplan bei Abweichung
			Nr.	Produkt	Prozess		Prod./Proz., Spezifikation/ Parameter/Toleranz	Prüf-/Messmethode/ Prüfmittel	Umfang	Frequenz		
17.40/ 17.50	Prüfung Fertigteil/ series inspection of final part	Feinschneid-anlage / fine cutting machine	17.40.10/ 17.50.10	Teildicke/ part thickness			siehe Zeichnung & akt. AA/ acc. to drawing and actual WI	Mikrometer / Micrometer	5	pro Schicht (alle 8h)	CAQ	suspekte Teile sperren, weiteres Vorgehen siehe aktuelle AA Lenkung fehlerhafter Produkte / *suspekt parts block, for further strategy see current status from AA control of faulty products*
34.20	Versand/ shipping	Versand-bereich shipping area	34.20.10	verpackte HU/ packed HU			ohne Beschädigungen bzw. Auffälligkeiten am Produkt gem. Verpackungsvorschrift/ without damages anomalies on product acc. to packaging instruction	Visuell (E)/ visual (D)	1	je Behälter/ each container	wenn niO sperren/ block if not ok	suspekte Teile sperren, weiteres Vorgehen siehe aktuelle AA Lenkung fehlerhafter Produkte / *suspekt parts block, for further strategy see current status from AA control of faulty products*

Bild 4 Bewertung der dokumentenübergreifenden Konsistenz im Anwendungsbeispiel. *Grafik: WZL der RWTH Aachen*

durch die Bewertung der LLM-generierten Antwort ermittelt werden. Im PLP und der VV sind die domänenspezifischen Abkürzungen „OEM“, „CAQ“ und „HU“ aufgeführt. Als zweideutige Terminologie kann die doppelte Verwendung des Keys „Index“ in der VV genannt werden.

3.3 Bewertung der pragmatischen Qualität

Die Bewertung der pragmatischen Qualität erfolgt durch die Bewertung der auf einer Nutzeranfrage und den bereitgestellten Kontextdaten generierten Modellantwort. Weiterhin ist der reale Einfluss der semantischen Defizite auf die Qualität beziehungsweise Sinnhaftigkeit der Modellantwort zu untersuchen. Dazu werden Nutzeranfragen beispielhaft zu den identifizierten Defiziten in Konsistenz, Vollständigkeit und Verständlichkeit definiert. Dabei ist darauf zu achten, dass die Terminologie, die in den Anfragen verwendet wird, mit der aus dem beigefügten Kontext übereinstimmt, um einen geringen Einfluss auf das Untersuchungsergebnis zu haben. Zur Analyse der Antworten wurden drei unterschiedliche Modelle betrachtet: Phi3:3.8b, Ollama3.2:3b und GPT-4o. Letzteres dient als Referenzmodell, da es die höchste Anzahl an Parametern aufweist. Die weiteren Modelle wurden trotz der geringeren Anzahl an Parametern berücksichtigt, da sie insbesondere lokal gehostet werden können und einen geringen Ressourcenaufwand erfordern und somit die Implementationshürde für beispielsweise KMU reduzieren. Die Tabelle 1 führt die Nutzeranfragen sowie die Antworten der drei Modelle in den drei Qualitätsmetriken auf. Korrekte Keywords sind in grün aufgeführt, während inkorrekte Antworten rot markiert sind.

Im oberen Teil der **Tabelle** sind drei Anfragen zur Qualitätsmetrik der Konsistenz aufgeführt. Die erste Frage bezieht sich auf die Abweichungen in der Formatierung in der VV. Da der zum Key „CC-Merkmal“ zugehörige Wert nicht im gleichen Muster dokumentiert wurde, ist keines der Modelle in der Lage, die Anfrage korrekt zu beantworten. Vielmehr wird ein fehlerhafter Wert, der jedoch dem Muster entsprechen würde, abgerufen. Die zweite Frage der Tabelle beruft sich auf die terminologische Heterogenität hinsichtlich der Entität „Platine“. Zwei der drei Modelle haben die Verbindung zwischen der Bauteilbezeichnung und der Entität „Platine“ in der VV ziehen können. Die dritte Frage bezieht sich auf die dokumentenübergreifende Konsistenz in der Benennung der Teilenummer. Der Begriff „Fertigteil“ aus dem PLP konnte von allen LLMs korrekt auf die in der VV dokumentierte Teilenummer beziehungsweise die im PLP aufgeführte Abkürzung bezogen werden.

Eine beispielhafte Anfrage hinsichtlich der Vollständigkeit ist im zweiten Abschnitt der Tabelle aufgeführt. Der unternehmensinterne Begriff „AA Lenkung“ wurde in der Anfrage als „Arbeitsanweisung Lenkung“ aufgeführt, um zu sehen, ob die Modelle in der Lage sind, dieses interne Wissen mit dem Kontext zu verknüpfen. Keines der Modelle war in der Lage, Abkürzung im bereitgestellten Kontext mit der Anfrage zu verknüpfen. Es ist zu

https://doi.org/10.37544/1436-4980-2025-11-12-35 - am 18.12.2025, 13:23:07. https://www.inlibra.com/de/agb - Open Access -

Tabelle Analyse der pragmatischen Qualität hinsichtlich Konsistenz, Vollständigkeit und Verständlichkeit.

Metrik	Nutzeranfrage	LLM		
		Phi3:3.8b	Ollama3.2:3b	GPT-4o
	Weist das Bauteil aus der Verpackungsvorschrift ein CC-Merkmal auf?	based on the information you've shared, yes it appears that the part from the packaging instructions does have a Control Characteristic (CC).	Ja, das Bauteil „Top AccuPressS" aus der Verpackungsvorschrift weist ein CC-Merkmal auf. Dies wird in der Verpackungsvorschrift unter dem Punkt „CC-Merkmal" angegeben: „CC-Merkmal: AB-3371".	Ja, das Bauteil aus der Verpackungsvorschrift weist ein CC-Merkmal auf: AB-3371.
Konsistenz	Was ist die Materialnummer der Platinen?	1234567890	Die Materialnummer der Platinen ist 12123412.	Die Materialnummer der Platinen ist 12123412, wie im Verpackungsdokument für „Top AccuPressS" angegeben.
	Welche Teilenummer ist dem Fertigteil zuzuordnen?	Gemäß den Verpackungsanweisungen ist „AB-3371-2345-42" wahrscheinlich die Teilenummer, die der spezifischen Produktvariante zugewiesen wurde.	Basierend auf den Informationen aus der Verpackungs-vorschrift kann die Teilenummer dem Fertigteil „AB-3371-2345-42" zugewiesen werden. Dies steht in der 4. Spalte, unter der „Interne Materialnr." als „12123412" geführt ist.	Dem Fertigteil ist die interne Teilenummer AB-0001-2345-42 aus dem ersten Dokument zuzuordnen
Vollständigkeit	Wann ist die Arbeitsanweisung „Lenkung" zu berücksichtigen?	Die Anweisung für die Aufbewahrung oder Lenkung (Au) kann bei der Vorbereitung und Verpackung von Produkten, wie in den Angaben zum Skizzenbeispiel angegeben, relevant sein. [...]	Basierend auf der bereitgestellten Kontextinformationen scheint die „Lenkung" in Zusammenhang mit der Verarbeitung von defekten Teilen oder Produkten zu sein. Die genaue Antwort ist jedoch nicht direkt im Text enthalten.[...]	Die Arbeitsanweisung „Lenkung" ist immer dann zu berücksichtigen, wenn das Fertigteil in den Fertigungs- oder Prüfprozess eintritt, insbesondere in den folgenden Fällen: Vor Beginn der Produktion:[...] Bei Qualitätskontrollen / Prüfungen:[...]
Verständlichkeit	Wer ist der Original Equipment Manufacturer der Platinen?	The original equipment manufacturer (OEM) of the Top AccuPressS platens is Orvian.	Der Original Equipment Manufacturer (OEM) der Platinen laut den Packaging-Anweisungen ist „Orvian".	Der Original Equipment Manufacturer (OEM) der Platinen laut dem von dir bereitgestellten zweiten Dokument ist Orvian.
	Wie lautet der Index des Bauteils „Top AccuPressS"?	Der Index des Bauteils „Top AccuPressS" ist „Z-09".	Der Index des Bauteils „Top AccuPressS" ist „--".	Der Index des Bauteils „Top AccuPressS" lautet Z-09.

sehen, dass die Modelle in diesem Fall halluzinieren und erfundene Antworten generieren.

Im unteren Abschnitt der Tabelle sind zwei Fragen hinsichtlich der Verständlichkeit aufgeführt. Zunächst wird die Interpretation des domänen-spezifischen Begriffs „Original Equipment Manufacturer" abgefragt. Die in den Daten aufgeführte Abkürzung scheint im Trainingskorpus der drei Modelle vorhanden gewesen zu sein. Die doppelte Aufführung des Begriffs „Index" wird in der zweiten Anfrage thematisiert. Hier ist zu sehen, dass das Ollama-Modell die Anfrage fehlerhaft beantwortet. Welcher Wert mit dem Key „Index" vom Modell korrekt identifiziert wird, ist somit nicht eindeutig.

Durch die beispielhafte Untersuchung wird ersichtlich, dass eine hohe semantische Qualität in den Bereichen Konsistenz, Vollständigkeit und Verständlichkeit Voraussetzungen für hohe pragmatische Qualität und die Vernetzung von Wissen ist. Im folgenden Abschnitt werden daher Maßnahmen zur Erhöhung der Datenqualität erörtert.

4 Maßnahmen zur Verbesserung der Datenqualität

Maßnahmen zur Verbesserung der Qualität textueller Daten im Kontext der LLM-basierten Informationsbereitstellung lassen sich in die drei Kategorien Konsistenz, Vollständigkeit und Verständlichkeit zusammenfassen, **Bild 5**. Zur Steigerung der Konsistenz sind Standardisierungskonzepte hinsichtlich organisationsinterner Dokumententypen, deren Layouts und der verwendeten Terminologie in Betracht zu ziehen. Die Einführung solcher Standards erfordert die Berücksichtigung einer abteilungsübergreifenden Perspektive sowie eine technische Betrachtung der Anforderungen an eine maschinenlesbare Layoutgestaltung. Aspekte zur Erhöhung der kontextuellen Vollständigkeit adressieren die hinreichende Bereitstellung von Kontextinformationen in den jeweiligen Dokumenten, die Verfügbarkeit von Metadaten sowie die Definition von organisations-spezifischen Abkürzungen. Die Anreicherung der Dokumente mit diesen Informationen zielt die Einordnung der Inhalte in den organisationalen Kontext ab. Eine Steigerung der Datenqualität in der Dimension der Verständlichkeit hingegen zielt auf eine Einordnung der Inhalte in den domänen-spezifischen Kontext und die Verknüpfung innerhalb des

https://doi.org/10.37544/1436-4980-2025-11-12-35 - am 18.12.2025, 13:23:07. https://www.inlibra.com/de/agb - Open Access -

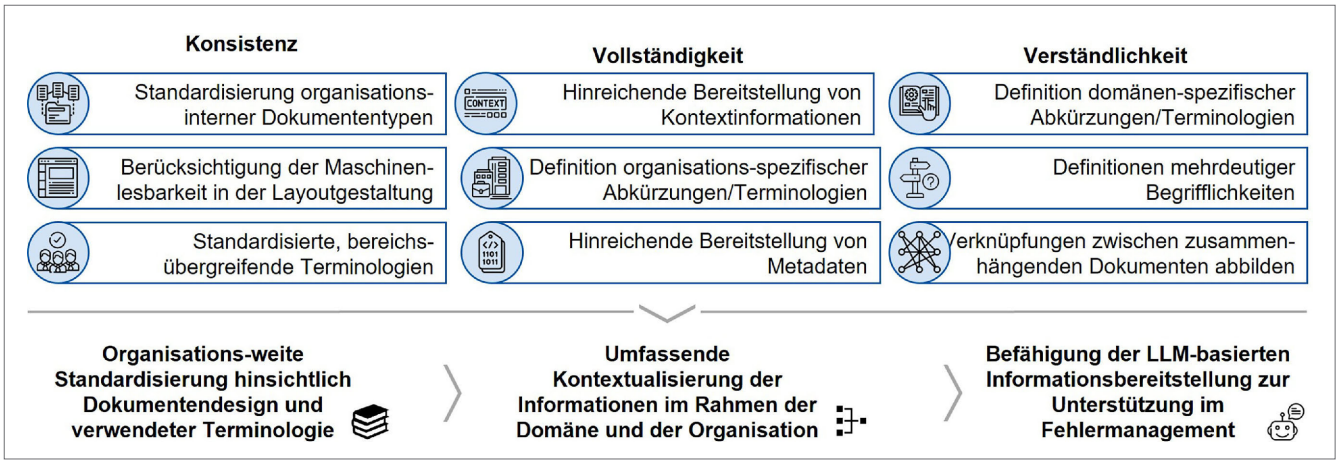

Bild 5 Handlungsempfehlungen zur Steigerung der Qualität textueller Daten einer Organisation. *Grafik: WZL der RWTH Aachen*

Dokumentenkorpus ab. Somit sind domänen-spezifische und mehrdeutige Begrifflichkeiten zu identifizieren und Definitionen bereitzustellen. Weiterhin können Methoden der Informationsmodellierung zur Verknüpfung der Dokumente hinzugezogen werden, um Unklarheiten hinsichtlich der Zusammenhänge zwischen einzelnen Informationsentitäten zu eliminieren.

Durch eine Standardisierung der Dokumentengestaltung und internen Terminologien, der Bereitstellung von hinreichenden Kontextinformationen sowie der Einordnung von Informationen in den organisationalen und domänen-spezifischen Kontext kann eine umfassende Steigerung der Qualität textueller Daten erreicht werden.

5 Zusammenfassung und Fazit

Der Beitrag untersuchte die Bedeutung der Datenqualität textueller Dokumente im Produktionskontext als zentrale Voraussetzung für die Etablierung LLM-basierter Informationsbereitstellungssysteme. Es zeigte sich, dass die semantische Qualität der Dokumente eine Voraussetzung für das Erreichen hoher pragmatischer Qualität in den LLM-generierten Antworten darstellt. Von besonderer Relevanz sind dabei die Dimensionen Konsistenz, Vollständigkeit und Verständlichkeit: Inkonsistenzen im Layout oder in Terminologien, fehlende Kontextinformationen durch Abkürzungen oder organisationsspezifische Begrifflichkeiten sowie Mehrdeutigkeiten oder domänenspezifischen Formulierungen erschweren die Informationsverarbeitung. Anhand eines industriellen Praxisbeispiels wurden das Vorgehen zur Qualitätsbewertung textueller Daten sowie die Auswirkungen von Qualitätsdefiziten auf die LLM-basierte Informationsbereitstellung aufgezeigt.

Zur Verbesserung der Datenqualität wurden Maßnahmen wie die organisationsweite Standardisierung von Dokumentenlayouts und Terminologien sowie die umfassende Kontextualisierung der Informationen im organisatorischen und domänenspezifischen Umfeld abgeleitet. Die Umsetzung der Maßnahmen zur Qualitätssteigerung in den Dimensionen Konsistenz, Vollständigkeit und Verständlichkeit weist ein hohes Potenzial zur Etablierung LLM-basierter Informationsbereitstellung im Produktionskontext auf und könnte einen maßgeblichen Beitrag zur Effektivitätssteigerung in Fehlermanagementprozessen leisten.

Ausblickend sind Parallelen zu Herausforderungen bei industriellen Sensordaten aufzuzeigen, für die mit den FAIR-Prinzipien (Findable, Accessible, Interoperable, Reusable) bereits ein etabliertes Rahmenwerk zur Sicherstellung von Datenqualität existiert. Dennoch fehlen in der Praxis häufig konsistente und interoperable Metadaten [22]. Ohne eine klare Kontextualisierung und Standardisierung sind Sensordaten, wie textuelle Dokumente, oft nicht interpretierbar oder wiederverwendbar. Die Übertragung der FAIR-Prinzipien zur Qualitätssteigerung von Sensordaten auf textuelle Daten eröffnet daher erhebliche Potenziale. Langfristig könnte so eine gemeinsame Grundlage für die durchgängige Nutzbarkeit heterogener industrieller Daten geschaffen werden, die nicht nur LLM-basierte Informationsbereitstellung, sondern auch umfassend datengetriebene Innovationen im Produktionskontext ermöglicht.

FÖRDERHINWEIS

Diese Arbeit wurde im Rahmen des Forschungsprojekts „Pre-PAIR - Predictive Failure Management with AI in Production“ (Projektnummer 13IK037I) durchgeführt. Besonderer Dank gilt der Europäischen Union und dem Bundesministerium für Wirtschaft und Energie (BMWE) für die Förderung des Projektes.

LITERATUR

[1] Glöckner, H.: System Dynamics Modeling of Reactive Failure Management in Production Systems. Aachen: Apprimus 2020

[2] Lee, J.; Siahpour, S.; Jia, X. et al.: Introduction to resilient manufacturing systems. Manufacturing Letters 32 (2022) 32, S. 24–27. Internet: https://doi.org/10.1016/j.mfglet.2022.02.002. Zugriff am 05.12.2025

[3] Wachter, C.; Beckschulte, S.; Padrón Hinrichs, M. et al.: Strategies for Resilient Manufacturing: A Systematic Literature Review of Failure Management in Production. Procedia CIRP 130 (2024), S. 1393–1402. Internet: https://doi.org/10.1016/j.procir.2024.10.257. Zugriff am 05.12.2025

[4] Beckschulte, S.; Padrón Hinrichs, M.; Pirrone, L. et al.: Manufacturing Data Analytics Study 2023 - Empirical Industry Study. Aachen: Apprimus 2023

[5] Soares Ito, A.; Ylipää, T.; Gullander, P. et al.: Prioritisation of root cause analysis in production disturbance management. International Journal of Quality & Reliability Management 39 (2022) 5, S. 1133–1150. Internet: https://doi.org/10.1108/IJQRM-12-2020-0402. Zugriff am 05.12.2025

[6] Gao, Y.; Xiong, Y.; Gao, X. et al.: Retrieval-Augmented Generation for Large Language Models: A Survey (2024). Internet: https://arxiv.org/pdf/2312.10997v5. Zugriff am 05.12.2025

https://doi.org/10.37544/1436-4980-2025-11-12-35 - am 18.12.2025, 13:23:07. https://www.inlibra.com/de/agb - Open Access -

[7] Ghahramani, Z.: Probabilistic machine learning and artificial intelligence. Nature 521 (2015) 7553, pp. 452–459. https://doi.org/10.1038/nature14541. Zugriff am 05.12.2025

[8] Wang, C.; Liu, X.; Yue, Y. et al.: Survey on Factuality in Large Language Models: Knowledge, Retrieval and Domain-Specificity (2023). Internet: https://arxiv.org/pdf/2310.07521v3. Zugriff am 03.12.2025

[9] Jiang, M.; Lin, B. Y.; Wang, S. et al.: Knowledge-augmented Methods for Natural Language Processing. Singapore: Springer Nature Singapore; Imprint Springer 2024

[10] Barnett, S.; Kurniawan, S.; Thudumu, S. et al.: Seven Failure Points When Engineering a Retrieval Augmented Generation System. Proceedings of the IEEE/ACM 3rd International Conference on AI Engineering (2024), S. 194–199. Internet: https://doi.org/10.1145/3644815.3644945. Zugriff am 05.12.2025

[11] ISO/IEC 25012, Ausgabe 2008

[12] ISO 8000-8. Data quality - Part 8: Information and data quality: Concepts and measuring. Ausgabe 2015

[13] Günther, L. C.; Colangelo, E.; Wiendahl, H.-H. et al.: Data quality assessment for improved decision-making: a methodology for small and medium-sized enterprises. Procedia Manufacturing 29 (2019), S. 583–591. Internet: https://doi.org/10.1016/j.promfg.2019.02.114. Zugriff am 05.12.2025

[14] Aamodt, A.; Nygård, M.: Different roles and mutual dependencies of data, information, and knowledge — An AI perspective on their integration. 1995

[15] Ackoff, R. L.: From data to wisdom. In: Journal of Applied Systems Analysis, 16 (1989), S. 3–9

[16] North, K.: Wissensorientierte Unternehmensführung, 7. Aufl., Wiesbaden: Springer Gabler (2021)

[17] Wang, B.; Wei, C.; Liu, Z. et al.: Resilience of Large Language Models for Noisy Instructions (2024). Internet: https://arxiv.org/pdf/2404.09754. Zugriff am 03.12.2025

[18] Zhang, X.; Yang, C.; Yan, H. et al.: Does Correction Remain A Problem For Large Language Models? (2023). Internet: https://arxiv.org/pdf/2308.01776. Zugriff am 03.12.2025

[19] Chen, J.; Lin, H.; Han, X. et al.: Benchmarking Large Language Models in Retrieval-Augmented Generation (2023). Internet: https://arxiv.org/pdf/2309.01431v2. Zugriff am 03.12.2025

[20] Kamm, S.; Jazdi, N.; Weyrich, M.: Knowledge Discovery in Heterogeneous and Unstructured Data of Industry 4.0 Systems: Challenges and Approaches. Procedia CIRP 104 (2021), S. 975–980. Internet: https://10.1016/j.procir.2021.11.164. Zugriff am 03.12.2025

[21] Batini, C.; Cappiello, C.; Francalanci, C. et al.: Methodologies for data quality assessment and improvement. ACM Computing Surveys 41 (2009) 3, S. 1–52. Internet: https://doi.org/10.1145/1541880.1541883. Zugriff am 05.12.2025

[22] Bodenbenner, M.; Montavon, B.; Schmitt, R. H.: FAIR sensor services - Towards sustainable sensor data management. Measurement: Sensors 18 (2021), S. 100206. Internet: https://doi.org/10.1016/j.measen.2021.100206. Zugriff am 05.12.2025

Charlotte Wachter
charlotte.wachter@wzl-iqs.rwth-aachen.de

Chris Schönekehs

Marcos Padrón Hinrichs

Felix Sohnius

Robert H. Schmitt

Werkzeugmaschinenlabor WZL der RWTH Aachen
Campus-Boulevard 30, 52074 Aachen

Robert H. Schmitt

Fraunhofer Institut für Produktionstechnologie IPT
Steinbachstr. 17, 52074 Aachen

LIZENZ

Dieser Fachaufsatz steht unter der Lizenz Creative Commons Namensnennung 4.0 International (CC BY 4.0)

https://doi.org/10.37544/1436-4980-2025-11-12-35 - am 18.12.2025, 13:23:07. https://www.inlibra.com/de/agb - Open Access -