

VDA QMC

Verband der Automobilindustrie
Qualitäts-Management-Center

Qualitätsmanagement in der Automobilindustrie

Künstliche Intelligenz im Qualitätsmanagement

1. Ausgabe, März 2026
Online-Download-Dokument

Qualitätsmanagement in der Automobilindustrie

Künstliche Intelligenz im Qualitätsmanagement

1. Ausgabe, März 2026

Online-Download-Dokument

Verband der Automobilindustrie e. V. (VDA)

ISSN 0943-9412

Copyright 2026 by

Verband der Automobilindustrie e. V. (VDA)
Qualitäts Management Center (QMC)
10117 Berlin, Behrenstraße 35

Online-Download-Dokument

Unverbindliche Empfehlung des VDA

Der Verband der Automobilindustrie (VDA) empfiehlt seinen Mitgliedern, den nachstehenden VDA-Band bei der Einführung und Aufrechterhaltung von QM-Systemen anzuwenden.

Haftungsausschluss

Dieser VDA-Band ist eine Empfehlung, die allen frei zur Anwendung steht. Wer sie anwendet, hat im konkreten Fall für die richtige Anwendung Sorge zu tragen.

Dieser VDA-Band berücksichtigt die zum Zeitpunkt der jeweiligen Ausgabe bekannten technischen Verfahrensweisen. Durch das Anwenden der VDA-Empfehlungen entzieht sich niemand der Verantwortung für sein eigenes Handeln. Alle handeln selbstverantwortlich.

Eine Haftung des VDA und der Personen, die an der Erstellung der VDA-Empfehlungen beteiligt sind, ist ausgeschlossen.

Wer bei der Anwendung dieser VDA-Empfehlung auf Unrichtigkeiten oder die Möglichkeit einer unrichtigen Auslegung stößt, wird gebeten, dies dem VDA umgehend mitzuteilen, damit etwaige Mängel beseitigt werden können.

Urheberrechtsschutz

Diese Schrift ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des VDA unzulässig und strafbar. Dies gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Übersetzungen

Das deutsche Dokument ist das Original. Bei Auslegungsfragen in anderen Sprachversionen ist auf die deutsche Version als Original Bezug zu nehmen. Diese Schrift wird auch in anderen Sprachen erscheinen. Der jeweils aktuelle Stand ist bei VDA QMC zu erfragen.

Vorwort

Künstliche Intelligenz (KI) ist längst keine Zukunftsvision mehr – sie ist Realität. In der Automobilindustrie eröffnet sie enorme Potenziale für Effizienz, Präzision und vorausschauende Steuerung. Gleichzeitig wirft ihr Einsatz neue Fragen auf: Wie funktioniert KI? Wo liegen Risiken? Wie können wir sie verantwortungsvoll und qualitätsgesichert nutzen und freigeben? Und welche Kompetenzen sind wichtig und müssen wir uns aneignen?

Der vorliegende VDA-Band „KI im Qualitätsmanagement“ bietet eine strukturierte Orientierung für Fach- und Führungskräfte in Qualitätssicherung, Produktion, Entwicklung, IT und Data Science. Ziel ist es, den Zugang zum Thema KI zu erleichtern – nicht durch technische Komplexität, sondern durch verständliche Begriffe, praxisnahe Beispiele und eine klare Struktur. Denn nur wer versteht, kann verantwortungsvoll mitgestalten.

Nutzen Sie diesen Band als praxisorientiertes Instrument, um Berührungspunkte mit dem Thema KI abzubauen, Risiken fundiert einschätzen zu lernen und sich zu kompetenten, verantwortungsvollen Anwender:innen weiterzuentwickeln. Die Integration von KI in industrielle Prozesse ist keine Frage des Ob, sondern des Wie. Umso wichtiger ist es, sich frühzeitig mit den Grundlagen vertraut zu machen und die eigene Gestaltungsfähigkeit zu stärken.

Inhaltsverzeichnis

Vorwort	7
Inhaltsverzeichnis	8
1 Einleitung	12
2 Terminologie	14
2.1 Einleitung	14
2.2 Bezug zu bestehenden Normen und Standards	14
2.2.1 DIN EN ISO/IEC 22989 – Konzepte und Terminologie für KI	14
2.2.2 EU AI Act – Anforderungen an das Qualitätsmanagement	15
2.2.3 ISO 9001 / IATF 16949 / VDA 6.x – Qualitätsmanagement in der Automobilindustrie	16
2.3 Grundbegriffe der Künstlichen Intelligenz	16
2.3.1 Künstliche Intelligenz (KI), Maschinelles Lernen (ML) und Deep Learning (DL)	17
2.3.2 Trainingsdaten, Modelle, Inferenz und Agenten	17
2.3.3 Überwachtes, teilüberwachtes, unüberwachtes und bestärkendes Lernen	18
2.3.4 Natural Language Processing (NLP), Language Models und Retrieval-Augmented Generation (RAG)	19
2.3.5 KI-gestützte Dialogsysteme („Chatbots“)	20
2.4 KI-spezifische Begriffe im Qualitätsmanagement	21
2.4.1 Anomalieerkennung	21
2.4.2 Auditierbarkeit	23
2.4.3 Bias	24
2.4.4 Blackbox	26
2.4.5 Confidence Score	26
2.4.6 Drift	27
2.4.7 Explainability	29
2.4.8 Fairness	29
2.4.9 Ground Truth	31
2.4.10 Halluzination	32
2.4.11 Kausalmodell	33

2.4.12	Predictive Quality	34
2.4.13	Prescriptive Quality	35
2.4.14	Robustheit.....	36
2.4.15	Vertrauenswürdigkeit (Trustworthiness).....	37
2.5	Regulatorische Begriffe im Kontext von KI-Systemen.....	38
2.5.1	High-Risk AI System	39
2.5.2	Konformitätsbewertung.....	39
2.5.3	Datenqualität und Daten-Governance	39
2.5.4	Auditability, Traceability und Transparency	40
2.5.5	Nicht-hochriskante KI-Systeme	41
2.6	Glossar KI im Qualitätsmanagement	42
3	KI im QM erfolgreich nutzen	43
3.1	Mindset, Arbeitskultur, Motivation und Change Management ...	46
3.2	Fähigkeiten und Kompetenzen	49
3.3	Daten	50
3.4	Organisationsstrukturen und Prozesse	51
3.5	Governance, Datenschutz, Standards und Regularien	52
3.6	Infrastruktur.....	54
3.7	Technische Schnittstellen und Integration.....	55
3.8	Einsatzpotenziale, Anwendungsfälle, Tools und Methoden	56
4	KI-Kompetenzen im QM	58
4.1	Haupt-Kompetenzen für KI im Qualitätsmanagement.....	59
4.2	Rollen im Qualitätsmanagement und relevante KI-Kompetenzen	61
4.2.1	Klassische Rollen im Qualitätsmanagement	62
4.2.2	Neue Rollen im Qualitätsmanagement.....	63
4.2.3	Rollenspezifische KI-Kompetenzen	64
5	KI-Systeme im QM freigeben	66
5.1	Schritt 1: Ermittlung der Projektrisikoklasse	67
5.2	Schritt 2: Risikobewertung des KI-Systems.....	73
5.2.1	Übersicht der Leitfragen zu den bewertungsrelevanten Anforderungen in Prozessphasen	73
5.2.2	Bewertungsrelevante Anforderungen an KI-Systeme	77

6	Handlungsempfehlungen und Anwendungsbeispiele.....	106
6.1	KI-gestützte optische Qualitätskontrolle	109
6.1.1	Beispiel	113
6.2	Regelorientierte KI-Agenten zur Unterstützung des 8D-Prozesses	116
6.2.1	Beispiel	119
6.3	KI-gestütztes Audit.....	125
6.3.1	Beispiel	128
6.4	KI-gestützte FMEA.....	132
6.4.1	Beispiel	134
6.5	Prädiktive Prozesslenkung	139
6.5.1	Beispiel	142
6.6	Vorbeugende Instandhaltung	147
6.6.1	Beispiel	150
6.7	Felddatenanalyse	155
6.7.1	Beispiel	158
6.8	Review von Entwicklungsarbeitsprodukten	161
6.8.1	Beispiel	163
6.8.2	Beispiel	163
6.8.3	Anmerkungen	166
6.9	VDA-Chatbot.....	167
6.9.1	Beispiel	168
6.9.2	Anmerkungen	170
6.10	Speech Mining für Arbeitsanweisungen	171
6.10.1	Beispiel	173
6.10.2	Anmerkung.....	174
6.11	Vergleichen von Dokumenten	175
6.11.1	Beispiel	177
6.11.2	Anmerkungen	178
6.12	Interaktives Lernen	179
7	Exkurs: Risikobasierte Bewertung von KI- Entwicklungswerkzeugen	183
7.1	Erläuterung der Grundkonzepte	184
7.2	Durchführung der risikobasierten Bewertung von KI- Entwicklungswerkzeugen	188

7.2.1	Schritt 1: Ermittlung von Risiken für ausgesuchte Entwicklungsaufgaben	190
7.2.2	Schritt 2: Werkzeugbewertung und Ermittlung des Qualifizierungsbedarfs	193
7.2.3	Liste potenzieller Fehlerzustände (beispielhaft)	197
7.3	Beispielhafte Anwendung der Methode	201
7.3.1	Kontext KI-gestützte SPC-Auswertung im Qualitätsmanagement	201
7.3.2	Kontext des fiktiven Entwicklungswerkzeugs <i>MLtoolExample</i>	201
7.3.3	Schritt 1 Ermittlung von Risiken für Entwicklungsaufgaben	202
7.3.4	Schritt 2 Werkzeugbewertung und Ermittlung des Qualifizierungsbedarfs	206
7.4	Zusammenfassende Einordnung im Kontext der Werkzeugqualifizierung	211

1 Einleitung

Der Band behandelt den praktischen Einsatz von Künstlicher Intelligenz im Qualitätsmanagement als Unterstützung und Hilfestellung in der Praxis. Der Schwerpunkt liegt darauf, konkrete KI-Systeme im Qualitätsmanagement zu beschreiben, freizugeben und ihre Risiken in der Anwendung gezielt zu reduzieren.

Bei identischen Anforderungen oder Anfragen an KI-Systeme können deren Ausgaben bzw. Ergebnisse variieren. Daher müssen sie durch geeignete und nachvollziehbare Mechanismen validierbar sein, beispielsweise durch definierte Regeln, festgelegte Grenzwerte und zugesicherte Antwortzeiten.

Der Band gliedert sich in folgende Kapitel, die jeweils zentrale Aspekte des KI-Einsatzes im Qualitätsmanagement beleuchten:

- **„Terminologie“** führt in die wesentlichen Begriffe der KI-Technologie ein und ordnet sie gezielt in den Kontext des Qualitätsmanagements ein.
- **„KI im QM erfolgreich nutzen“** beschreibt unter Berücksichtigung der geltenden regulatorischen Anforderungen sowie relevanter Standards den Einsatz von Künstlicher Intelligenz in Unternehmen. Der Band gibt einen Überblick über Methoden, Standards und Anwendungspotenziale entlang des Produktlebenszyklus.
- **„KI-Kompetenzen im QM“** beschreibt die für den Einsatz von Künstlicher Intelligenz im Qualitätsmanagement erforderlichen Rollen und Kompetenzanforderungen.
- **„KI-Systeme im QM freigeben“** beschreibt eine Methodik und Ermittlung der Risiken eines KI-Systems als Basis für eine Freigabe.
- **„Handlungsempfehlungen und Anwendungsbeispiele“** definiert in Form von Handlungsempfehlungen, wie KI-Anwendungen betrieben und angepasst und wie deren Ergebnisse fachlich interpretiert und bewertet werden können.

Neben diesen Kerninhalten bietet der Exkurs **„Risikobasierte Bewertung von KI-Entwicklungswerkzeugen“** eine Beschreibung zur risikobasierten Bewertung von KI-Entwicklungswerkzeugen.

Nicht im Scope des Bandes sind Anwendungen von Fahrzeugfunktionen mit KI-Anteil sowie die praktische Umsetzung des EU AI Acts.

Dieser Band dient der fachlichen Orientierung und ersetzt keine juristische Bewertung oder regulatorische Einordnung von KI-Systemen.

2 Terminologie

2.1 Einleitung

Die fortschreitende Einbindung von Künstlicher Intelligenz (KI) in qualitätsrelevante Prozesse verändert nicht nur Methoden und Werkzeuge, sondern auch die Sprache des Qualitätsmanagements. Dies bringt neue Herausforderungen mit sich: Begriffe und Konzepte erhalten im KI-Kontext andere Bedeutungen, werden unterschiedlich interpretiert oder sind nicht eindeutig in bestehende QM-Methoden eingebettet.

Im vorliegenden Kapitel wird das Ziel verfolgt, eine praxisorientierte, einheitliche Verwendung von Begriffen für den Einsatz von KI im Qualitätsmanagement zu schaffen. Dies versteht sich als Ergänzung zu bestehenden Normen, insbesondere zur DIN EN ISO/IEC 22989 („Künstliche Intelligenz – Konzepte und Terminologie“) sowie zu qualitätsrelevanten Standards wie ISO 9001, IATF 16949 und VDA 6.x. Es werden zentrale KI-Begriffe in den Kontext des automobilien Qualitätsmanagements übertragen und konkrete Anwendungsbeispiele, Abgrenzungen und Hinweise zur Operationalisierung geboten.

Ziel ist es, durch eine gemeinsame Sprache die Kommunikation zwischen Fachbereichen zu erleichtern, die Auditierbarkeit von KI-Systemen zu verbessern und die Grundlage für zukünftige Standards und regulatorische Anforderungen zu legen.

2.2 Bezug zu bestehenden Normen und Standards

2.2.1 DIN EN ISO/IEC 22989 – Konzepte und Terminologie für KI

Die DIN EN ISO/IEC 22989:2022 definiert über 100 Begriffe und Konzepte im Zusammenhang mit Künstlicher Intelligenz und bietet eine international abgestimmte Grundlage für die Kommunikation über KI-Systeme. Sie beschreibt unter anderem den Lebenszyklus von KI-Systemen, die Rollen beteiligter Akteure (z. B. Anbieter, Nutzer:innen, Entwickler:innen) sowie zentrale Konzepte wie Bias, Explainability, Training, Inferenz und Modell.

Für das Qualitätsmanagement in der Automobilindustrie ist diese Norm ein wichtiger Referenzpunkt, da sie eine systematische Begriffswelt bereitstellt, die als Basis für die Bewertung, Auditierung und Weiterentwicklung von KI-

Anwendungen dienen kann. Allerdings bleibt sie bewusst branchenneutral und bietet keine spezifischen Anwendungsbeispiele oder Interpretationen für qualitätsrelevante Prozesse in der Automobilindustrie – hier setzt der vorliegende Band an.

2.2.2 EU AI Act – Anforderungen an das Qualitätsmanagement

Der europäische AI Act (EU 2024/1689) verpflichtet Anbieter und Hersteller – abhängig von der Risikoklasse des Systems – zur Umsetzung spezifischer Anforderungen an das Qualitätsmanagement. Für sog. High-Risk-Systeme (s. Kapitel 2.5.1) gelten besonders umfangreiche Vorgaben, doch auch Systeme niedrigerer Risikoklassen müssen ausgewählte Anforderungen der Verordnung erfüllen, etwa zu Transparenz, Robustheit oder Datenqualität. Artikel 17¹ der Verordnung fordert für High-Risk-Systeme ein dokumentiertes, systematisches Qualitätsmanagementsystem (QMS) bzw. die Integration der Anforderungen in ein bestehendes QMS. Dazu gehören unter anderem:

- Strategien zur regulatorischen Konformität
- Verfahren zur Entwicklung, Prüfung und Validierung von KI-Systemen
- Datenmanagementprozesse (inkl. Datenqualität, Labeling, Speicherung)
- Risikomanagement und Post-Market-Monitoring
- Nachvollziehbare Dokumentation und Kommunikationsprozesse
- Überwachung des Systems im Betrieb (Monitoring)

Diese Anforderungen überschneiden sich in vielen Punkten mit bestehenden QM-Systemen nach ISO 9001 oder IATF 16949, gehen jedoch in Bezug auf KI-spezifische Aspekte wie Datenverarbeitung, Modellvalidierung und algorithmische Transparenz deutlich darüber hinaus. Der vorliegende Band soll helfen, diese Anforderungen in die Sprache und Praxis des automobilen Qualitätsmanagements zu übersetzen.

¹ [Article 17: Quality Management System | EU Artificial Intelligence Act.](#)

2.2.3 ISO 9001 / IATF 16949 / VDA 6.x – Qualitätsmanagement in der Automobilindustrie

Die Norm ISO 9001:2015 bildet die Grundlage für Qualitätsmanagementsysteme weltweit. Sie betont Kundenorientierung, risikobasiertes Denken und kontinuierliche Verbesserung. Die IATF 16949:2016 ergänzt diese Anforderungen um automobilspezifische Aspekte, darunter:

- Produktsicherheit und Rückverfolgbarkeit
- Fehlervermeidung statt Fehlerentdeckung
- Lieferantenentwicklung und -bewertung
- Prozessfähigkeitsanalysen und FMEA

Ergänzend dazu konkretisieren die VDA-Bände 6.x die System-, Prozess- und Produktanforderungen für die deutsche Automobilindustrie.

Mit dem Einsatz von KI in qualitätsrelevanten Prozessen – etwa zur automatisierten Fehlerklassifikation, Anomalieerkennung oder Vorhersage von Qualitätsabweichungen – entsteht für diese etablierten Systeme ein neuer Regelungsbedarf, u. a.:

- Auditierbarkeit von KI-Modellen im Rahmen von System-, Prozess- und Produktaudits
- Erklärungspflicht bei automatisierten Entscheidungen (z. B. bei Ausschussklassifikation)
- Bewertung der Datenqualität und -herkunft als Teil der Prozess- und Produktsicherheit
- Integration von KI in bestehende Prüfstrategien, Dienstleistungsbewertungen und Lieferantenbewertungen

2.3 Grundbegriffe der Künstlichen Intelligenz

Dieses Kapitel stellt zentrale Begriffe und Konzepte von KI vor, die für das Qualitätsmanagement in der Automobilindustrie relevant sind. Ziel ist es, ein gemeinsames Verständnis zu schaffen, das die Kommunikation zwischen Qualitätsmanagement, Data Science, IT und Produktion erleichtert.

2.3.1 Künstliche Intelligenz (KI), Maschinelles Lernen (ML) und Deep Learning (DL)

Künstliche Intelligenz (KI) bezeichnet maschinengestützte Systeme, die für einen unterschiedlich autonomen Betrieb ausgelegt sind und die aus Eingaben ableiten, wie Ausgaben wie z. B. Vorhersagen, Inhalte, Empfehlungen oder Entscheidungen erzeugt werden, die physische oder virtuelle Umgebungen beeinflussen können (vgl. Art. 3 Nr. 1 EU AI Act). Im Kontext dieses Bandes umfasst der Begriff insbesondere Systeme, die Aufgaben erfüllen, die typischerweise menschliche Intelligenz erfordern – etwa Text- und Bild-generierung, Mustererkennung oder Entscheidungsfindung

Maschinelles Lernen (ML) ist ein Teilgebiet der Künstlichen Intelligenz, bei dem Systeme aus Daten lernen, ohne explizit programmiert zu werden. ML-Modelle erkennen Muster und treffen Vorhersagen auf Basis historischer Daten. ML-Modelle erkennen Muster und treffen Vorhersagen auf Basis historischer Daten.

Deep Learning (DL) ist eine spezielle Form des Maschinellen Lernens, die auf künstlichen, tiefen neuronalen Netzen basiert. DL wird häufig für komplexe Aufgaben wie Bild- oder Spracherkennung eingesetzt.

Bezug zum Qualitätsmanagement (QM): In der Qualitätskontrolle können ML-Modelle z. B. Fehlerbilder klassifizieren, Anomalien in Prozessdaten erkennen oder Qualitätsabweichungen vorhersagen.

2.3.2 Trainingsdaten, Modelle, Inferenz und Agenten

Trainingsdaten sind strukturierte oder unstrukturierte Daten,² mit denen ein KI-Modell „angelernt“ wird. Die Qualität und Repräsentativität dieser Daten ist entscheidend für die spätere Modellleistung.

Hinweis: Trainingsdaten enthalten in der Regel auch die als korrekt geltenden Referenzwerte (Ground Truth), die für das Lernen und die spätere Validierung des Modells entscheidend sind. Ground Truth wird in Kapitel 2.4.9 separat erläutert.

² „Strukturierte Daten“ sind klar organisierte Informationen, meist in Tabellenform oder Datenbanken, z. B. Messwerte mit Zeitstempeln. „Unstrukturierte Daten“ sind weniger formalisiert und liegen z. B. als Bilder, Texte, Audioaufnahmen oder Videos vor.

Ein trainiertes KI-**Modell** bezeichnet die mathematische Struktur, die nach dem Training in der Lage ist, Vorhersagen zu treffen oder Entscheidungen zu unterstützen.

Inferenz ist der Prozess, bei dem ein trainiertes Modell auf neue, unbekannte Daten angewendet wird, um eine Vorhersage oder Klassifikation zu erzeugen.

Agenten in KI-Systemen sind autonome oder teilautonome Softwarekomponenten, die Aufgaben ausführen, Entscheidungen treffen oder mit ihrer Umgebung interagieren. Sie nutzen trainierte Modelle, um Informationen zu verarbeiten und zielgerichtet zu handeln – etwa durch das Erkennen von Mustern, das Treffen von Entscheidungen oder das Auslösen von Aktionen. In komplexen Anwendungen können mehrere Agenten miteinander kommunizieren und kooperieren, um ein übergeordnetes Ziel zu erreichen.

Bezug zum Qualitätsmanagement (QM): Im Qualitätsmanagement können trainierte Modelle z. B. in der visuellen End-of-Line-Prüfung eingesetzt werden, um automatisch zu entscheiden, ob ein Bauteil den Qualitätsanforderungen entspricht. Agenten können darüber hinaus in der Produktionsüberwachung eingesetzt werden, um kontinuierlich Daten zu analysieren, Anomalien zu erkennen und automatisch Maßnahmen zur Qualitätssicherung einzuleiten.

2.3.3 Überwachtes, teilüberwachtes, unüberwachtes und bestärkendes Lernen

Überwachtes Lernen (Supervised Learning): Ein Modell wird anhand von Daten, die mit bekannten Ergebnissen (Labels) versehen sind – z. B. „i. O.“ oder „n. i. O.“³ – trainiert.

Unüberwachtes Lernen (Unsupervised Learning): Ein Modell wird mit Daten ohne Labels trainiert, um Muster oder Strukturen zu erkennen – z. B. Clusterbildung bei Prozessdaten.

Teilüberwachtes Lernen (Semi-Supervised Learning): Das Modell wird anhand einer Kombination aus gelabelten und ungelabelten Daten trainiert.

³ „i. O.“ steht für „in Ordnung“ und bezeichnet ein Produkt oder Bauteil, das den Qualitätsanforderungen entspricht. „n. i. O.“ steht für „nicht in Ordnung“ und kennzeichnet ein Produkt oder Bauteil, das die Qualitätsanforderungen nicht erfüllt.

Dies ist besonders nützlich, wenn die manuelle Kennzeichnung großer Datenmengen aufwendig oder kostenintensiv ist. Das Modell nutzt die wenigen vorhandenen Labels, um Muster auch in den unbeschrifteten Daten zu erkennen.

Bestärkendes Lernen (Reinforcement Learning): Der Agent lernt durch Versuch und Irrtum, indem er Rückmeldungen aus seiner Umgebung und von der Bedienperson erhält, z. B. zur Optimierung von Prüfstrategien.

Bezug zum Qualitätsmanagement (QM): Überwachtes Lernen ist besonders relevant für Klassifikationsaufgaben, z. B. bei der automatisierten Bauteilbewertung. Teilüberwachtes Lernen kann eingesetzt werden, wenn nur ein Teil der Qualitätsdaten manuell klassifiziert wurde, etwa bei neuen Produktvarianten. Unüberwachtes Lernen eignet sich für die Anomalieerkennung, z. B. bei Sensorwerten. Bestärkendes Lernen kann zur Optimierung von Prüfstrategien oder zur dynamischen Anpassung von Qualitätsprozessen genutzt werden.

2.3.4 Natural Language Processing (NLP), Language Models und Retrieval-Augmented Generation (RAG)

Natural Language Processing (NLP) ist ein Teilbereich der Künstlichen Intelligenz, der sich mit der automatisierten Verarbeitung, Analyse und Generierung natürlicher Sprache beschäftigt. Ziel ist es, Maschinen in die Lage zu versetzen, menschliche Sprache zu verstehen und sinnvoll darauf zu reagieren – sowohl in geschriebener als auch gesprochener Form.

Language Models sind KI-Modelle, die darauf trainiert werden, Sprache zu verstehen, zu analysieren oder zu erzeugen. Diese Modellklasse existiert schon seit vielen Jahren. **Large Language Models (LLMs)** sind eine weiterentwickelte und stark vergrößerte Form dieser Sprachmodelle. Sie werden auf sehr großen Textmengen trainiert und nutzen tiefe neuronale Netze, um kontextbezogene Antworten zu generieren, Texte zu analysieren oder neue Inhalte zu erstellen. LLMs bilden die Grundlage für generative KI-Anwendungen (GenAI).

Retrieval-Augmented Generation (RAG) erweitert die Fähigkeiten von LLMs, indem es die generative Komponente mit einem gezielten Informationsabruf aus externen Datenquellen kombiniert. Während ein LLM auf sei-

nem Trainingswissen basiert, ermöglicht RAG den Zugriff auf aktuelle, geprüfte Inhalte – z. B. QM-Datenbanken, Normendokumente oder Reklamationshistorien. Dadurch können Antworten nicht nur sprachlich plausibel, sondern auch inhaltlich korrekt und kontextbezogen sein.

Bezug zum Qualitätsmanagement (QM): Im Qualitätsmanagement können NLP, LLMs und RAG genutzt werden, um textbasierte Informationen wie Prüfberichte, Reklamationen oder Normendokumente effizient zu verarbeiten. Sie unterstützen die automatisierte Erstellung von Berichten, die Analyse von Fehlerursachen und die Bereitstellung interaktiver Assistenzsysteme (z. B. Chatbots), die Fachkräfte bei der Bearbeitung von 8D-Reports oder der Ursachenanalyse unterstützen. RAG ist besonders relevant, wenn KI-Systeme auf aktuelle, validierte QM-Daten zugreifen müssen, um normgerechte Formulierungen oder fundierte Handlungsvorschläge zu liefern.

2.3.5 KI-gestützte Dialogsysteme („Chatbots“)

KI-gestützte Dialogsysteme – häufig als „**Chatbots**“ bezeichnet – sind eine spezielle Form von Sprachassistenzsystemen, die auf Künstlicher Intelligenz basieren. Im Gegensatz zu regelbasierten Chatbots, die auf vordefinierten Entscheidungsbäumen und festen Antwortmustern beruhen, reagieren KI-gestützte Systeme flexibel und kontextbezogen auf Nutzereingaben. Sie ermöglichen eine interaktive Kommunikation mit Nutzer:innen und können im Qualitätsmanagement vielfältige Aufgaben übernehmen, etwa bei der Bearbeitung von Reklamationen, der Ursachenanalyse oder der Dokumentation von Maßnahmen.

Technologische Grundlagen wie Natural Language Processing (NLP), Large Language Models (LLMs) und Retrieval-Methoden (z. B. Retrieval-Augmented Generation, RAG) werden in Kapitel 2.3.2 erläutert. NLP ermöglicht das Verstehen von Sprache, LLMs generieren kontextbezogene Antworten auf Basis großer Textmengen, und RAG ergänzt diese Fähigkeit durch den Zugriff auf aktuelle, geprüfte Datenquellen wie QM-Datenbanken oder Normendokumente, sodass Antworten nicht nur sprachlich plausibel, sondern auch inhaltlich korrekt sind.

Kurz zusammengefasst: NLP = Sprache verstehen, LLM = Antworten generieren, RAG = Unternehmenswissen einbinden.

Bezug zum Qualitätsmanagement (QM): Im Qualitätsmanagement bieten solche Systeme Potenzial zur Effizienzsteigerung, insbesondere bei repetitiven Aufgaben oder bei der Unterstützung von Fachkräften in komplexen Entscheidungsprozessen. Gleichzeitig stellen sie neue Anforderungen an die Validierung, Nachvollziehbarkeit und Akzeptanz: Die Grenze zwischen assistierender Unterstützung und automatisierter Entscheidung muss klar definiert sein, insbesondere wenn Vorschläge für Maßnahmen oder Ursachenanalysen generiert werden. Ein KI-gestützter Chatbot kann z. B. bei der Bearbeitung von 8D-Berichten unterstützen, ähnliche Fehlerfälle identifizieren oder normgerechte Formulierungen vorschlagen. Voraussetzung für den erfolgreichen Einsatz ist eine transparente Dokumentation der Datenquellen, eine klare Rollenverteilung zwischen Mensch und Maschine sowie eine Schulung der Anwender:innen im Umgang mit KI-Applikationen im Qualitätsmanagement.

Wichtig: Bei generativen Systemen besteht das Risiko sogenannter Halluzinationen – also der Erzeugung plausibel klingender, aber faktisch falscher Inhalte. Dieses Risiko muss bei der Validierung und im Betrieb berücksichtigt werden (siehe Kapitel 2.4.10 „Halluzination“). Darüber hinaus bestehen Risiken wie Fehlinterpretationen, Bias in den Daten oder mangelnde Nachvollziehbarkeit der Antworten. Diese Aspekte müssen bei der Konzeption und im Betrieb berücksichtigt werden.

2.4 KI-spezifische Begriffe im Qualitätsmanagement

Dieses Kapitel erläutert zentrale Begriffe aus der KI, die im Kontext des Qualitätsmanagements eine besondere Bedeutung haben. Ziel ist es, ein gemeinsames Verständnis für Begriffe zu schaffen, die in der Praxis häufig verwendet, aber unterschiedlich interpretiert werden – insbesondere an den Schnittstellen zwischen Data Science, Produktion und Qualitätssicherung.

Jeder Begriff wird mit einer Definition, einem QM-spezifischen Anwendungsbezug und ggf. mit Verweisen auf Normen oder regulatorische Anforderungen erläutert.

2.4.1 Anomalieerkennung

Definition (ISO/IEC 22989):

Anomalieerkennung (engl. Anomaly Detection) bezeichnet die Identifikation von Datenpunkten, Mustern oder Ereignissen, die von der erwarteten Norm

abweichen. Sie kann auf statistischen Methoden, maschinellem Lernen oder hybriden Ansätzen basieren.

Relevanz im QM-Kontext:

In der Qualitätsüberwachung kann Anomalieerkennung helfen, ungewöhnliche Prozessverläufe, fehlerhafte Bauteile oder Sensorabweichungen frühzeitig zu erkennen – oft bevor ein klassischer Regelverstoß oder ein Ausschuss auftritt –, und kann daher gut z. B. bei der statistischen Prozesslenkung (SPC) eingesetzt werden. Sie ist besonders nützlich in komplexen, datenreichen Prozessen, in denen klassische Grenzwertlogik an ihre Grenzen stößt.

Typische Herausforderungen:

- Prozessinstabilitäten (z. B. Werkzeugverschleiß, Temperaturschwankungen)
- Sensorfehler oder Kalibrierprobleme
- Menschliche Eingriffe oder Bedienfehler
- Neue oder seltene Fehlerbilder, die im Training nicht enthalten waren

Maßnahmen zur Sicherstellung:

- Auswahl geeigneter Algorithmen (z. B. Isolation Forest, Autoencoder, statistische Verfahren)
- Kombination mit Domänenwissen zur Vermeidung von Fehlalarmen
- Etablierung eines Feedbackprozesses zur kontinuierlichen Verbesserung
- Visualisierung und Kontextualisierung der Anomalien für Fachanwender:innen

Verwandte Begriffe:

Predictive Quality, Drift, Datenqualität, SPC, Prozessüberwachung

2.4.2 Auditierbarkeit

Definition:

Auditierbarkeit beschreibt die Fähigkeit, Entscheidungen, Prozesse und Ergebnisse eines KI-Systems nachvollziehbar zu dokumentieren, zu bewerten und zu überprüfen – sowohl intern als auch durch externe Stellen (z. B. Kunden, Zertifizierungsstellen, Behörden).

Relevanz im QM-Kontext:

In der Automobilindustrie ist Auditierbarkeit ein zentrales Prinzip – etwa im Rahmen von Systemaudits (ISO 9001, IATF 16949), Prozessaudits (VDA 6.3) und Produktaudits (VDA 6.5) sowie bei Lieferantenaudits. Wenn KI-Systeme qualitätsrelevante Entscheidungen treffen oder Entscheidungsvorschläge machen (z. B. Ausschussklassifikation, Lieferantenbewertung), müssen diese Entscheidungen rückverfolgbar, erklärbar und dokumentiert sein. Dies ist auch eine zentrale Anforderung im EU AI Act für sogenannte High-Risk AI Systems.

Typische Herausforderungen:

- Fehlende oder unvollständige Dokumentation von Trainingsdaten, Modellversionen oder Entscheidungslogik
- Einsatz von „Blackbox“-Modellen ohne erklärbare Entscheidungswege
- Keine klaren Verantwortlichkeiten für KI-Systeme im QM-System
- Fehlende Integration in bestehende Auditprozesse

Maßnahmen zur Sicherstellung:

- Einführung eines KI-Lebenszyklusmanagements mit Versionierung, Dokumentation und Nachvollziehbarkeit
- Einsatz von Explainability-Methoden zur Unterstützung der Auditfähigkeit
- Verankerung von KI-Systemen im Qualitätsmanagementhandbuch
- Schulung von Auditor:innen im Umgang mit KI-Systemen
- Nutzung von Audit-Trails, Logging und automatisierter Dokumentation

Verwandte Begriffe:

Explainability, Transparenz, Re-Training, Validierung, EU AI Act

Bezug zu Kapitel 2.5.4 Auditability:

Kapitel 2.4.2 beschreibt „Auditierbarkeit“ im Sinne des Qualitätsmanagements – also die Fähigkeit, KI-gestützte Entscheidungen im Rahmen von QM-Audits (z. B. nach VDA 6.3) nachvollziehbar zu dokumentieren.

In Kapitel 2.5.4 wird der Begriff „Auditability“ im regulatorischen Kontext des EU AI Act betrachtet, wo er als Pflichtenforderung für High-Risk-Systeme definiert ist. Beide Perspektiven ergänzen sich und sollten im Sprachgebrauch klar voneinander abgegrenzt werden, indem QM-Dokumentationen explizit auf die Auditziele (System-, Prozess-, Produktaudit) verweisen und regulatorische Anforderungen unter dem Begriff „Auditability“ separat in Compliance-Dokumenten verankert werden.

2.4.3 Bias

Definition (ISO/IEC 22989):

Bias bezeichnet eine systematische Verzerrung in Daten, Modellen oder Entscheidungsprozessen, die zu fehlerhaften oder unfairen Ergebnissen führen kann.

Relevanz im QM-Kontext:

Im Qualitätsmanagement kann Bias in unterschiedlichen Formen auftreten. Ein Beispiel ist Daten- oder Label-Bias, wenn Trainingsdaten für ein KI-Modell zur Fehlerklassifikation überwiegend aus einer bestimmten Produktionslinie stammen und dadurch andere Linien schlechter erkannt werden. Dies kann zu einer verzerrten Fehlererkennung führen und die Wirksamkeit von Korrekturmaßnahmen erheblich beeinträchtigen. Ebenso können Messmethoden selbst verzerrt sein (Mess- oder Feature-Bias), Modellarchitekturen bestimmte Muster bevorzugen (algorithmischer Bias) oder sich durch Rückkopplungseffekte im Betrieb neue Verzerrungen ergeben (Feedback-Loop / Deployment Bias).

Typische Herausforderungen:

- Daten- und Label-Bias: nicht repräsentative Daten (z. B. nur Tagesschicht), historische Fehlerklassifikationen mit menschlichem Bias, Ungleichgewicht in der Fehlerverteilung (Class Imbalance)

- Mess- und Feature-Bias: Merkmale oder Messmethoden sind selbst verzerrt (z. B. Sensoren mit systematischer Abweichung, ungeeignete Feature-Auswahl)
- Algorithmischer Bias: Modellarchitektur oder Loss-Funktionen benachteiligen bestimmte Gruppen oder sind ungeeignet für den Einsatzzweck
- Feedback-Loop / Deployment Bias: Das Modell beeinflusst die Daten, die es später selbst verarbeitet (z. B. bei Predictive Maintenance, wo Entscheidungen zukünftige Datenverteilungen verändern)

Maßnahmen zur Sicherstellung:

- Datenanalyse hinsichtlich Repräsentativität und Qualität
- Einsatz von Fairness-Metriken und Bias-Detektionsverfahren
- Validierung durch unabhängige QM-Teams
- Prüfung der Messmethoden und Feature-Auswahl auf Verzerrungen
- Review der Modellarchitektur und Loss-Funktionen hinsichtlich Fairness und Eignung
- Monitoring nach Deployment, um Feedback-Loops zu erkennen und zu korrigieren

Abgrenzung zum Bias im QM-Prüfprozessmanagement:

Der Begriff „Bias“ wird im Qualitätsmanagement traditionell auch für systematische Messabweichungen verwendet. Das VDA QMC Glossary definiert Bias als:

„Bias/BI der Messung bzw. Schätzwert einer systematischen Messabweichung.“

Diese Definition bezieht sich auf die Genauigkeit von Messprozessen und ist von dem hier behandelten Bias im KI-Kontext klar zu unterscheiden.

Verwandte Begriffe:

Fairness, Ground Truth, Data Drift

2.4.4 Blackbox

Definition:

Als Blackbox wird ein KI-System bezeichnet, dessen interne Funktionsweise für Anwender:innen nicht transparent oder nachvollziehbar ist. Die Eingaben und Ausgaben sind bekannt, aber die Entscheidungslogik bleibt verborgen.

Relevanz im QM-Kontext:

Blackbox-Modelle (z. B. komplexe neuronale Netze) erschweren die Auditierbarkeit und die Ursachenanalyse bei Reklamationen. Sie stehen im Gegensatz zu erklärbaren Modellen und sind besonders kritisch bei sicherheits- oder qualitätsrelevanten Anwendungen.

Typische Herausforderungen:

- Fehlende Transparenz für Auditor:innen und QM-Teams
- Schwierige Fehlerursachenanalyse
- Regulatorische Risiken bei High-Risk-Systemen (EU AI Act)

Maßnahmen zur Sicherstellung:

- Einsatz von XAI-Methoden zur teilweisen Öffnung der Blackbox
- Dokumentation der Modellarchitektur und Entscheidungslogik
- Kombination von Blackbox-Modellen mit erklärbaren Komponenten (Hybridansätze)

Verwandte Begriffe:

Explainability, Transparenz, Auditierbarkeit

2.4.5 Confidence Score

Definition:

Ein Confidence Score gibt an, mit welcher Wahrscheinlichkeit ein Modell seiner eigenen Vorhersage vertraut.⁴ Er ist ein Maß für die Unsicherheit der Entscheidung: Je höher der Score, desto geringer die Unsicherheit.

⁴ Hinweis: Der Begriff „Confidence Score“ ist derzeit nicht in ISO/IEC-Normen formal definiert, wird jedoch in der industriellen Praxis und in KI-Frameworks weit verbreitet verwendet.

Wichtig: Der Confidence Score ist nicht mit der Modellgenauigkeit (Accuracy) gleichzusetzen. Ein Modell kann eine Vorhersage mit 99 % Vertrauen treffen und dennoch eine falsche Entscheidung liefern.

Relevanz im QM-Kontext:

Ein Modell klassifiziert ein Teil z. B. als „i. O.“ mit 92 % Sicherheit. Bei niedrigen Confidence Scores steigt die Unsicherheit, sodass eine manuelle Nachprüfung erforderlich sein kann – insbesondere bei sicherheitskritischen Bauteilen. Die Berücksichtigung von Unsicherheit ist auch für die Prüfprozesseignung relevant (vgl. VDA-Band 5).

Typische Herausforderungen:

- Fehlinterpretation durch Anwender:innen
- Keine Schwellenwerte definiert
- Unsicherheit nicht dokumentiert oder nicht berücksichtigt

Maßnahmen zur Sicherstellung:

- Festlegung von Schwellenwerten für Nachprüfung unter Berücksichtigung der Unsicherheit
- Visualisierung der Scores und Unsicherheiten in Dashboards
- Kombination mit Explainability-Methoden zur besseren Interpretation
- Dokumentation der Unsicherheitsbewertung für Audits und Prüfprozesseignung

Verwandte Begriffe:

Unsicherheit, Entscheidungslogik, Auditierbarkeit

2.4.6 Drift

Definition (ISO/IEC 22989):

Drift bezeichnet die Veränderung von Datenverteilungen oder Zusammenhängen über die Zeit, wodurch die Leistung eines KI-Modells beeinträchtigt werden kann. Man unterscheidet typischerweise zwischen Data Drift (Veränderung der Eingabedaten) und Concept Drift (Veränderung der Beziehung zwischen Eingabe und Zielwert).

Relevanz im QM-Kontext:

Ein Modell, das auf Prozessdaten eines bestimmten Materials oder Maschinenzustands trainiert wurde, kann bei Änderungen in der Produktion (z. B. Materialcharge, Werkzeugverschleiß, neue Schichtbesetzung) an Genauigkeit verlieren. Drift kann zu Fehlklassifikationen, verzögerten Reaktionen oder falschen Entscheidungen führen.

Typische Herausforderungen:

- Prozessveränderungen (z. B. neue Chargen, neue Lots [Material], neue Lieferanten, Maschinenupdates)
- Saisonale Effekte oder Schichtwechsel
- Sensoralterung oder Kalibrierabweichungen
- Änderungen im Prüfprozess oder in der Datenvorverarbeitung

Maßnahmen zur Sicherstellung:

- Monitoring von Modellmetriken über die Zeit (z. B. Genauigkeit, Fehlerquote)
- Einsatz von Drift-Detektionsalgorithmen (z. B. Population Stability Index, Kolmogorov-Smirnov-Test)
- Re-Training oder Modellanpassung bei signifikanter Drift
- Dokumentation und Nachvollziehbarkeit von Änderungen im Prozessumfeld

Drift in anderen QM-Kontexten:

Im klassischen Qualitätsmanagement wird „Drift“ häufig für physikalische oder messtechnische Veränderungen verwendet, z. B. Temperaturschwankungen oder Abnutzungseffekte bei Prüfmitteln. Diese Bedeutung unterscheidet sich von der hier behandelten Drift im KI-Kontext, die sich auf Veränderungen in Daten oder Modellbeziehungen bezieht.

Verwandte Begriffe:

Re-Training, Modellvalidierung, Datenqualität, Robustheit

2.4.7 Explainability

Definition (EU AI Act, ISO/IEC 24029):

Explainability beschreibt die Fähigkeit, Entscheidungen eines KI-Systems für Menschen nachvollziehbar zu machen.

Relevanz im QM-Kontext:

In Audits oder bei Reklamationen muss nachvollziehbar sein, warum ein KI-System ein Teil als „n. i. O.“ klassifiziert hat. Explainability ist auch eine regulatorische Anforderung für High-Risk-Systeme gemäß EU AI Act.

Typische Herausforderungen:

- Komplexe Modelle (z. B. Deep Learning) sind schwer erklärbar
- Fehlende Dokumentation der Entscheidungslogik
- Unverständliche Visualisierungen oder Scores

Maßnahmen zur Sicherstellung:

- Einsatz erklärbarer Modelle (z. B. Entscheidungsbäume)
- Visualisierung von Entscheidungswegen
- Integration von XAI-Methoden (Explainable AI)⁵

Verwandte Begriffe:

Auditierbarkeit, Confidence Score, Transparenz, Blackbox

2.4.8 Fairness

Definition (ISO/IEC 22989):

Fairness bezeichnet die Eigenschaft eines KI-Systems, keine systematische Benachteiligung bestimmter Datenkategorien, Datengruppen oder Prozessbeteiligter zu verursachen. Sie ist eng mit den Konzepten von Bias und Transparenz verbunden.

⁵ Explainable AI (XAI) bezeichnet Methoden und Techniken, die darauf abzielen, die Entscheidungen und die Funktionsweise von KI-Systemen für Menschen verständlich und nachvollziehbar zu machen. Ziel ist es, Transparenz zu schaffen, Vertrauen zu fördern und regulatorische Anforderungen (z.B. EU AI Act) zu erfüllen – insbesondere bei komplexen Modellen wie Deep Learning.

Relevanz im QM-Kontext:

In qualitätsrelevanten Anwendungen kann mangelnde Fairness z. B. dazu führen, dass bestimmte Prozesse, Schichten oder Produktvarianten systematisch schlechter bewertet werden – nicht aufgrund objektiver Qualitätsdaten, sondern wegen unausgewogener Trainingsdaten oder unreflektierter Modelllogik.

Abgrenzung zu Bias:

- **Bias** bezeichnet eine Verzerrung in den (Trainings-)Daten, Modellen oder Prozessen und ist **messbar** (z. B. durch statistische Analysen)
- **Fairness** beschreibt die Gerechtigkeit der Ergebnisse eines Modells und ist bewertbar (z. B. durch Fairness-Metriken wie Equal Opportunity oder Demographic Parity)

Beispiel:

Ein Modell für Kredite kann Bias enthalten, weil historische Daten mehr Kredite an Männer beinhalten. Fairness bedeutet zu prüfen, ob das Modell trotzdem gleiche Chancen für Männer und Frauen sicherstellt.

Typische Herausforderungen:

- Schwierigkeit, Fairness im Modell-Output zu prüfen (z. B. gleiche Chancen für alle Gruppen)
- Auswahl geeigneter Fairness-Metriken und deren Interpretation
- Abwägung zwischen Fairness und anderen Zielen (z. B. Genauigkeit, Effizienz)
- Transparente Kommunikation der Fairness-Bewertung gegenüber Auditor:innen und Stakeholdern

Maßnahmen zur Sicherstellung:

- Einsatz von Fairness-Metriken (z. B. Equal Opportunity, Demographic Parity)
- Analyse und Ausbalancierung der Trainingsdaten
- Sensitivitätsanalysen und Gegenbeispiele im Modelltest
- Dokumentation von Annahmen und Modellgrenzen

Verwandte Begriffe:

Bias, Datenqualität, Lieferantenbewertung, Explainability

2.4.9 Ground Truth

Definition (ISO/IEC 25012):

Ground Truth bezeichnet die als korrekt geltenden Referenzdaten, mit denen ein KI-Modell trainiert oder validiert wird.

Relevanz im QM-Kontext:

In der Qualitätsprüfung kann Ground Truth z. B. durch manuell geprüfte Fehlerbilder oder durch Messdaten aus zertifizierten Prüfmitteln definiert sein. Die Qualität der Ground Truth ist entscheidend für die Modellgüte und die spätere Auditierbarkeit.

Hinweis: Ground Truth wird häufig auch in Trainingsdaten verwendet, ist jedoch nicht identisch mit dem Begriff „Trainingsdaten“. Während Trainingsdaten die gesamte Datenbasis für das Lernen eines Modells darstellen, bezeichnet Ground Truth die als korrekt geltenden Referenzwerte, die für Training, Validierung und Auditierung entscheidend sind (siehe Kapitel 2.3.2 „Trainingsdaten, Modelle, Inferenz und Agenten“).

Typische Herausforderungen:

- Fehlerhafte oder uneinheitliche Labeling-Prozesse
- Unvollständige oder nicht repräsentative Daten
- Abweichungen zwischen den als korrekt definierten Referenzdaten (Ground Truth) und den tatsächlichen Prozessbedingungen, z. B. wenn sich Produktionsparameter ändern oder neue Varianten auftreten
- Fehlende Verfügbarkeit oder unzureichende Menge an Trainingsdaten, die für die Erstellung und Validierung einer belastbaren Ground Truth erforderlich sind

Maßnahmen zur Sicherstellung:

- Einsatz von Expertenlabeling
- Validierung durch unabhängige QM-Teams

- Dokumentation der Herkunft und Qualität der Ground Truth

Verwandte Begriffe:

Bias, Datenqualität, Validierung

2.4.10 Halluzination**Definition (ISO/IEC 22989):**

Eine Halluzination bezeichnet die Erzeugung von Ausgaben durch ein KI-System, die zwar plausibel erscheinen, aber faktisch falsch, unbegründet oder nicht durch die Eingabedaten gedeckt sind.

Relevanz im QM-Kontext:

In qualitätskritischen Anwendungen kann eine Halluzination dazu führen, dass ein KI-System falsche Handlungsempfehlungen oder Klassifikationen generiert – z. B. eine fehlerhafte Diagnose im Predictive Quality oder eine falsche Ursache in einer Reklamationsanalyse. Dies gefährdet die Prozesssicherheit und die Auditierbarkeit.

Typische Herausforderungen:

- Fehlende Validierung der generierten Inhalte
- Übermäßiges Vertrauen in KI-Ausgaben durch Anwender:innen
- Einsatz von Modellen in Kontexten, für die sie nicht trainiert wurden
- Unzureichende Datenbasis oder mangelhaftes Prompt-Design bei generativen KI-Systemen

Maßnahmen zur Sicherstellung:

- Einsatz von Verifikationsmechanismen (z. B. Cross-Checks, Plausibilitätsprüfungen)
- Begrenzung des Einsatzes generativer KI auf nicht sicherheitskritische Bereiche
- Schulung der Anwender:innen im Umgang mit KI-Ausgaben
- Kombination mit klassischen QM-Methoden zur Validierung (z. B. Stichprobenprüfung)

Verwandte Begriffe:

Unsicherheit, Confidence Score, Explainability, Bias

2.4.11 Kausalmodell**Definition:**

Ein Kausalmodell beschreibt die kausalen Beziehungen zwischen Variablen in einem System. Sie können in Form von mathematischen Gleichungen oder visuell als Diagramme sowie in qualitativer und quantitativer Form repräsentiert werden.

Relevanz im QM-Kontext:

Diese Modelle können verwendet werden, um die Auswirkungen kontrollierbarer Interventionen und Ursache-Wirkungs-Beziehungen zu verstehen. In Kombination mit statistischen Daten können diese Modelle die kausale Inferenz von Beziehungen aus Daten unterstützen, die über Assoziationen hinausgehen. Einige QM-Methoden, wie die FMEA und die Ishikawa-Analyse, basieren auf qualitativen Betrachtungen von Kausalzusammenhängen.

Typische Herausforderungen:

- Entwicklung vollständiger und genauer Modelle
- Herausforderung bei der Prüfung und Validierung von Annahmen
- Unvollständige oder nicht repräsentative Daten

Maßnahmen zur Sicherstellung:

- Einsatz von Expertenwissen
- Visualisierung von Wirkungsketten
- Evaluation von Modellen und Sensitivitätsanalysen

Verwandte Begriffe:

Modelle, Inferenz, Ground Truth, Explainability, Transparenz, Auditierbarkeit, Drift

2.4.12 Predictive Quality

Definition:

Predictive Quality⁶ bezeichnet den Einsatz von Datenanalyse und maschinellem Lernen zur frühzeitigen Vorhersage von produkt- und prozessbezogenen Qualitätsmerkmalen. Ziel ist es, auf Grundlage der Vorhersagen Entscheidungen zu treffen und Maßnahmen einzuleiten, bevor mögliche Qualitätsabweichungen zu Ausschuss, Nacharbeit oder Kundenreklamationen führen.

Relevanz im QM-Kontext:

Predictive Quality ermöglicht eine proaktive Qualitätssicherung: Statt auf Fehler zu reagieren, können Unternehmen auf Basis von Prozess-, Maschinen- oder Umweltdaten präventiv eingreifen. Dies erhöht die Prozessstabilität, reduziert Kosten und verbessert die Kundenzufriedenheit.

Typische Herausforderungen:

- Fehlende oder unstrukturierte Datenhistorie und mangelnde Datenqualität
- Komplexe, nichtlineare Zusammenhänge zwischen Prozessparametern und Qualitätsmerkmalen
- Geringe Akzeptanz bei Fachabteilungen aufgrund fehlender Erklärbarkeit der Modelle („Blackbox“-Effekt)
- Schwierige Integration in bestehende IT- und Produktionssysteme (z. B. MES, ERP)

Maßnahmen zur Sicherstellung:

- Aufbau robuster Datenpipelines und Sicherstellung der Datenqualität
- Auswahl geeigneter ML-Modelle, z. B. Random Forest, Gradient Boosting, Long Short-Term Memory (LSTM)

⁶ Hinweis: Der Begriff „Predictive Quality“ ist derzeit nicht normativ definiert, wird jedoch in der industriellen Praxis verwendet, um datenbasierte Verfahren zur Vorhersage von Qualitätsabweichungen zu beschreiben. Er steht im Kontext von Industrie 4.0, datengetriebenem Qualitätsmanagement und KI-gestützter Prozessüberwachung.

- Enge Zusammenarbeit zwischen Data Science, Produktion und Qualitätssicherung
- Visualisierung von Vorhersagen und Handlungsempfehlungen für Anwender:innen
- Pilotierung mit klaren KPIs (z. B. Ausschussquote, Falschalarmrate, Frühwarnzeit, ROI)

Verwandte Begriffe:

Anomalieerkennung, Bias, Datenqualität, Drift, Explainability, Prozessfähigkeit, Regressionsanalyse, Re-Training, Condition Monitoring

2.4.13 Prescriptive Quality

Definition:

Prescriptive Quality bezeichnet den Einsatz von Datenanalyse, maschinellem Lernen und Optimierungsverfahren zur Vorhersage von Qualitätsabweichungen und zur Ableitung konkreter Handlungsempfehlungen, um deren Eintreten zu verhindern. Im Gegensatz zu Predictive Quality, das lediglich Vorhersagen trifft, kombiniert Prescriptive Quality Prognosen mit Vorschlägen für präventive Maßnahmen, die unter menschlicher Aufsicht umgesetzt werden können.

Relevanz im QM-Kontext:

Prescriptive Quality ermöglicht eine proaktive und handlungsorientierte Qualitätssicherung. Unternehmen können nicht nur erkennen, dass ein Risiko besteht, sondern auch, wie es vermieden werden kann – z. B. durch Anpassung von Prozessparametern, Wartungsmaßnahmen oder Materialwechsel. Dies erhöht die Prozessstabilität, reduziert Ausschuss und Nacharbeit und unterstützt eine kontinuierliche Verbesserung.

Typische Herausforderungen:

- Hohe Komplexität bei der Ableitung von Handlungsempfehlungen aus Vorhersagen
- Sicherstellung der Umsetzbarkeit und Praxistauglichkeit der vorgeschlagenen Maßnahmen

- Akzeptanz bei Fachabteilungen (Vertrauen in KI-generierte Empfehlungen)
- Integration in bestehende Entscheidungs- und Steuerungsprozesse

Maßnahmen zur Sicherstellung:

- Entwicklung von Entscheidungsmodellen, die Empfehlungen transparent und nachvollziehbar machen
- Kombination von KI-Methoden mit Domänenwissen zur Validierung der Vorschläge
- Klare Definition der Rolle von menschlicher Aufsicht („Human Oversight“) bei der Umsetzung
- Pilotprojekte mit messbaren KPIs (z. B. Reduktion von Ausschuss, Verbesserung der Prozessfähigkeit)
- Visualisierung von Empfehlungen und deren erwarteten Auswirkungen für Anwender:innen

Verwandte Begriffe:

Predictive Quality, Explainability, Condition Monitoring, Prozessoptimierung, Human Oversight, Re-Training

2.4.14 Robustheit

Definition (ISO/IEC 22989):

Robustheit beschreibt die Fähigkeit eines KI-Systems, auch unter veränderten Bedingungen, Störungen oder Ausreißern stabile und verlässliche Ergebnisse zu liefern.

Relevanz im QM-Kontext:

Für KI-Anwendungen bedeutet Robustheit die Widerstandsfähigkeit des Modells gegenüber Datenrauschen, schwankenden Umgebungsbedingungen oder neuen Varianten. Dabei unterscheidet sich dieser Begriff von der im VDA-Band „Robuster Produktionsprozess“ verwendeten Definition, die die Gesamtprozessstabilität beschreibt – also einen Betrieb, der gemäß Planung gesichert ist, unempfindlich gegen Störungen bleibt und Produkte nach Spezifikation termingerecht sowie in korrekter Menge liefert

(⇒ Kunde). Im KI-Kontext geht es hingegen primär um die Leistungsfähigkeit des Modells unter variierenden Eingangsbedingungen. Beide Konzepte ergänzen sich: Ein robustes KI-System unterstützt die Prozessrobustheit, ersetzt sie aber nicht.

Typische Herausforderungen:

- Überanpassung an Trainingsdaten (Overfitting)
- Empfindlichkeit gegenüber kleinen Datenabweichungen
- Fehlende Tests unter realen Produktionsbedingungen
- Unzureichende Modellpflege bei Prozessänderungen

Maßnahmen zur Sicherstellung:

- Einsatz robuster Modellarchitekturen und Regularisierung
- Testen mit Stördaten, Ausreißern und realen Prozessabweichungen
- Monitoring und kontinuierliche Validierung im Betrieb
- Kombination mit klassischen QM-Methoden (z. B. SPC)

Verwandte Begriffe:

Drift, Re-Training, Validierung, Prozessfähigkeit, Resilienz

2.4.15 Vertrauenswürdigkeit (Trustworthiness)

Definition (ISO/IEC 22989):

Vertrauenswürdigkeit bezeichnet die Eigenschaft eines KI-Systems, so gestaltet zu sein, dass es berechtigtes Vertrauen bei Anwender:innen und Stakeholdern schafft. Sie umfasst Aspekte wie Transparenz, Robustheit, Fairness, Sicherheit und Nachvollziehbarkeit.

Relevanz im QM-Kontext:

Im Qualitätsmanagement ist Vertrauenswürdigkeit entscheidend für die Akzeptanz von KI-Systemen. Sie beeinflusst die Bereitschaft, KI-gestützte Entscheidungen zu übernehmen, und ist Voraussetzung für Audits und regulatorische Konformität (z. B. EU AI Act für High-Risk-Systeme).

Typische Herausforderungen:

- Fehlende Transparenz bei komplexen Modellen
- Unklare Verantwortlichkeiten für KI-Entscheidungen
- Risiken durch Halluzinationen oder Bias
- Mangelnde Integration in bestehende QM-Prozesse

Maßnahmen zur Sicherstellung:

- Einsatz erklärbarer Modelle und Dokumentation der Entscheidungslogik
- Etablierung von Governance-Strukturen für KI-Systeme
- Validierung und Monitoring über den gesamten Lebenszyklus
- Schulung von Anwender:innen und Auditor:innen

Verwandte Begriffe:

Explainability, Auditierbarkeit, Fairness, Robustheit, Sicherheit

2.5 Regulatorische Begriffe im Kontext von KI-Systemen

Die zunehmende Regulierung von KI-Systemen – insbesondere durch den EU AI Act (EU 2024/1689) – bringt neue Begriffe mit sich, die auch im Qualitätsmanagement verstanden und verwendet werden müssen. Dieses Kapitel stellt zentrale Begriffe aus dem regulatorischen Umfeld vor und erläutert ihre Bedeutung im Kontext der KI-Terminologie. Ziel ist es, ein gemeinsames Sprachverständnis zu schaffen, das die Kommunikation insbesondere zwischen Fachbereichen, Auditor:innen und Regulierungsstellen erleichtert.

Hinweis zur Anwendung: Dieses Kapitel – wie auch der gesamte Band – dient ausschließlich der fachlichen und sprachlichen Orientierung im Umgang mit KI-Systemen im Qualitätsmanagement. Es ersetzt keine juristische Bewertung oder regulatorische Klassifikation. Für die Einordnung eines konkreten KI-Systems sind stets der Verwendungszweck, die betroffenen Grundrechte und die technische Ausgestaltung zu prüfen.

2.5.1 High-Risk AI System

Definition (EU AI Act, Art. 6 & Anhang III):

Ein KI-System, das ein hohes Risiko für Gesundheit, Sicherheit oder Grundrechte darstellt.

Sprachliche Einordnung:

Der Begriff „High-Risk“ ist nicht als technisches Attribut zu verstehen, sondern als rechtliche Klassifikation. Er bezieht sich auf den Verwendungszweck und das Anwendungsumfeld eines Systems.

Abgrenzung:

Ein technisches System kann je nach Einsatzkontext als High-Risk oder Non-High-Risk gelten.

Beispielhafte Formulierung:

„Das KI-System zur automatisierten Ausschussklassifikation fällt gemäß Anhang III des AI Act unter die Kategorie High-Risk.“

2.5.2 Konformitätsbewertung

Definition (EU AI Act, Art. 19–24):

Verfahren zur Überprüfung, ob ein KI-System die Anforderungen des AI Act erfüllt.

Sprachliche Besonderheit:

Der Begriff ist aus der Produktsicherheitsregulierung bekannt (z. B. CE-Kennzeichnung), wird im KI-Kontext jedoch auf algorithmische Systeme übertragen.

Typische Missverständnisse:

„Konformitätsbewertung“ ist nicht gleichzusetzen mit einem klassischen QM-Audit – sie umfasst auch technische Dokumentation, Risikobewertung und ggf. externe Prüfstellen.

2.5.3 Datenqualität und Daten-Governance

Definition (EU AI Act, Art. 10):

Anforderungen an die Qualität, Repräsentativität und Dokumentation der verwendeten Daten.

Sprachliche Einordnung:

Der Begriff „Datenqualität“ wird im AI Act nicht normativ definiert, sondern durch Anforderungen operationalisiert (z. B. „frei von Verzerrungen“, „repräsentativ“).

Begriffliche Abgrenzung:

„Datenqualität“ im Sinne des AI Act ist nicht identisch mit klassischen QM-Kriterien wie Messgenauigkeit, Accuracy, Precision, Messunsicherheit oder Kalibrierung.

2.5.4 Auditability, Traceability und Transparency

Definition:

Im EU AI Act werden mehrere Begriffe verwendet, die sich auf die Nachvollziehbarkeit und Prüfbarkeit von KI-Systemen beziehen:

- **Auditability:** Fähigkeit zur externen Überprüfung eines KI-Systems durch Dritte (z. B. Behörden, Zertifizierungsstellen)
- **Traceability:** Rückverfolgbarkeit von Daten, Modellen und Entscheidungen über den gesamten Lebenszyklus
- **Transparency:** Offenlegung der Funktionsweise, Logik und Grenzen eines KI-Systems

Sprachliche Einordnung:

Diese Begriffe werden im Deutschen häufig unter dem Oberbegriff „Transparenz“ oder „Nachvollziehbarkeit“ zusammengefasst, haben jedoch im regulatorischen Kontext unterschiedliche Bedeutungen. Eine präzise Verwendung unterstützt die Kommunikation mit Regulierungsbehörden, Auditor:innen und internen Stakeholdern.

Tabelle 2-1: Abgrenzung der Begriffe Auditability, Traceability und Transparency

Begriff	Fokus	Typische Inhalte	Empfohlene deutsche Übersetzung
Auditability	Prüfbarkeit durch Dritte	Dokumentation, Prüfpfade, externe Bewertung	Auditierbarkeit ⁷
Traceability	Datenfluss & Modellhistorie	Datenquellen, Modellversionen, Labeling-Prozesse	Rückverfolgbarkeit
Transparency	Verständlichkeit & Offenlegung	Zweck, Funktionsweise, Grenzen des Systems	Transparenz

2.5.5 Nicht-hochriskante KI-Systeme

Definition (implizit im AI Act):

KI-Systeme, die nicht unter Anhang III des EU AI Act fallen und somit nicht den Anforderungen für High-Risk-Systeme unterliegen.

Sprachliche Einordnung:

Es gibt keine offizielle Bezeichnung wie „Low-Risk“ oder „Unregulated AI“ – in der Praxis haben sich Begriffe wie „Non-High-Risk AI“ oder „Limited-Risk AI“ etabliert.

⁷ Kapitel 3.4.8 beschreibt „Auditierbarkeit“ im Sinne des Qualitätsmanagements – also die Fähigkeit, KI-gestützte Entscheidungen im Rahmen von QM-Audits (z. B. nach VDA 6.3) nachvollziehbar zu dokumentieren.

In Kapitel 5.4 wird der Begriff im regulatorischen Kontext des EU AI Act betrachtet, wo er Teil der Anforderungen an High-Risk-Systeme ist. Beide Perspektiven ergänzen sich und sollten im Sprachgebrauch klar voneinander abgegrenzt werden.

Empfehlung zur Sprachverwendung:

- Begriffe wie „ungefährlich“ oder „nicht reguliert“ vermeiden
- Formulierungen verwenden wie: „Dieses System fällt nicht unter die High-Risk-Klassifikation gemäß AI Act“

2.6 Glossar KI im Qualitätsmanagement

Definitionen und Begriffe aus den VDA-Publikationen werden im übergreifenden Online-Glossar des VDA QMC zur Verfügung gestellt:

<https://vda-qmc-learning.de/module/glossar/glossar.php>

3 KI im QM erfolgreich nutzen

Die Nutzung von KI in Organisationen, insbesondere im Qualitätsmanagement, bietet vielfältige Möglichkeiten zur Verbesserung der Produkt- und Dienstleistungsqualität. Aufgrund immer kürzerer Produktlebenszyklen und der sich ändernden Marktbedingungen ist der Einsatz von KI zur Unterstützung von Qualitätsmanagementmethoden unerlässlich, um mit diesen Entwicklungen Schritt zu halten.

Der Einsatz von KI kann unter anderem den folgenden Nutzen entlang des Produktentstehungsprozesses generieren:

- **Effizientere Datenanalyse:** KI kann Daten aus multimodalen Quellen (z. B. Texte, Bilder, Zeitreihen und Sensoren) verarbeiten und analysieren, die bisher durch klassische Methoden nicht auswertbar waren.
- **Neue Zugriffsoptionen und Interfaces auf interne und externe Daten:** Intuitive Interaktion über semantische Suche und natürliche Sprache ermöglicht einen deutlich schnelleren und breiteren Zugang zu Wissen, auch für Nutzer:innen ohne technische Expertise.
- **Verbesserung und Automatisierung von Prozessen:** Durch den Einsatz von KI können Verbesserungspotenziale in Prozessen (z. B. Entwicklungs-, Business- und Produktionsprozesse) identifiziert und entsprechende Maßnahmen vorgeschlagen werden. Dies kann zu einer höheren Produktivität und geringeren Kosten führen.
- **Automatisierte Inspektionen:** KI-gestützte Analysen können zur automatischen Inspektion von Bauteilen eingesetzt werden, um die Geschwindigkeit, Genauigkeit und Robustheit der Prüfungen zu erhöhen.
- **Fehlererkennung und -prognose:** KI-gestützte Systeme können Muster bzw. Anomalien in Daten analysieren und potenzielle Qualitätsabweichungen vorhersagen (Predictive Quality), bevor sie zu komplexeren Problemen und höheren Kosten führen. Dies ermöglicht eine vorausschauende Qualitätskontrolle und -steuerung.

- **Predictive Maintenance:** KI kann dabei unterstützen, Wartungsbedarfe vorherzusagen, indem sie Daten über den Zustand von Maschinen analysiert. Dadurch kann die Ausfallzeit minimiert und die Lebensdauer der Maschinen verlängert werden.
- **Kundenfeedback-Analyse:** KI-Tools können Kundenfeedback und Daten aus dem Feld wie sozialen Medien oder Umfragen analysieren, um Trends und Probleme zu erkennen, die auf Qualitätsmängel hinweisen könnten.

Die gezielte und erfolgreiche Nutzung von KI im Qualitätsmanagement erfordert die Berücksichtigung verschiedener Aspekte in Bezug auf **Menschen, Technik und Organisation**.

- **Mensch:** Die Mitarbeitenden sind der Kern jeder Organisation. Ihre Akzeptanz und Fähigkeit, mit neuen Technologien umzugehen, sind entscheidend für den Erfolg der KI-Implementierung. Schulungen und Weiterbildung sind notwendig, um sicherzustellen, dass die Mitarbeitenden die neuen Systeme verstehen und effektiv nutzen können. Schließlich sind Motivation und eine positive Einstellung gegenüber Technologie Schlüsselfaktoren.
- **Technik:** Die technische Infrastruktur muss die Voraussetzungen erfüllen, KI-Anwendungen zu unterstützen und in der Organisation zu verbreiten. Dazu gehören geeignete Hardware, Softwarelösungen und Plattformen sowie Datenbanken zur Speicherung und Verarbeitung großer Datenmengen. Definierte Schnittstellen und Prozesse ermöglichen die Einbettung von KI in der Organisation. Tools und Methoden befähigen zum bedarfsgerechten Einsatz von KI.
- **Organisation:** Gezielte Anwendungsfälle bilden die Grundlage für die Definition von Anforderungen der benötigten Ressourcen. Klare Organisationsstrukturen sind notwendig, um Verantwortlichkeiten zu definieren und die Kommunikation zwischen verschiedenen Abteilungen und Fachbereichen zu fördern. Governance und Standards ermöglichen den nutzenstiftenden Einsatz von KI unter Einhaltung von Datenschutz und Regularien (z. B. EU AI Act, DSGVO).

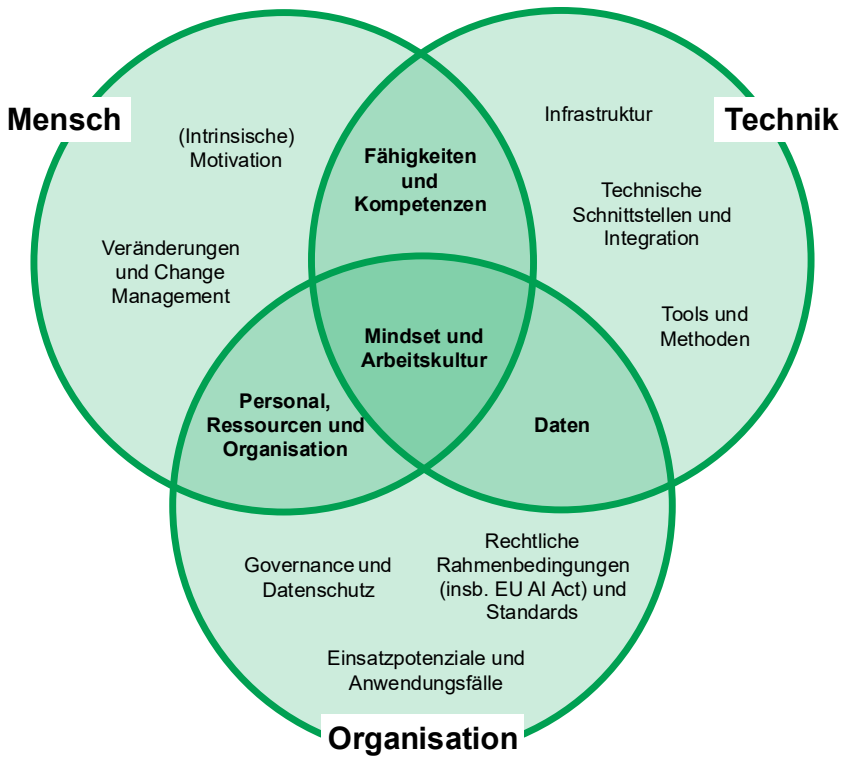


Abbildung 3-1: Einordnung von Aspekten, die bei der erfolgreichen Nutzung von KI in der Organisation zu berücksichtigen sind.

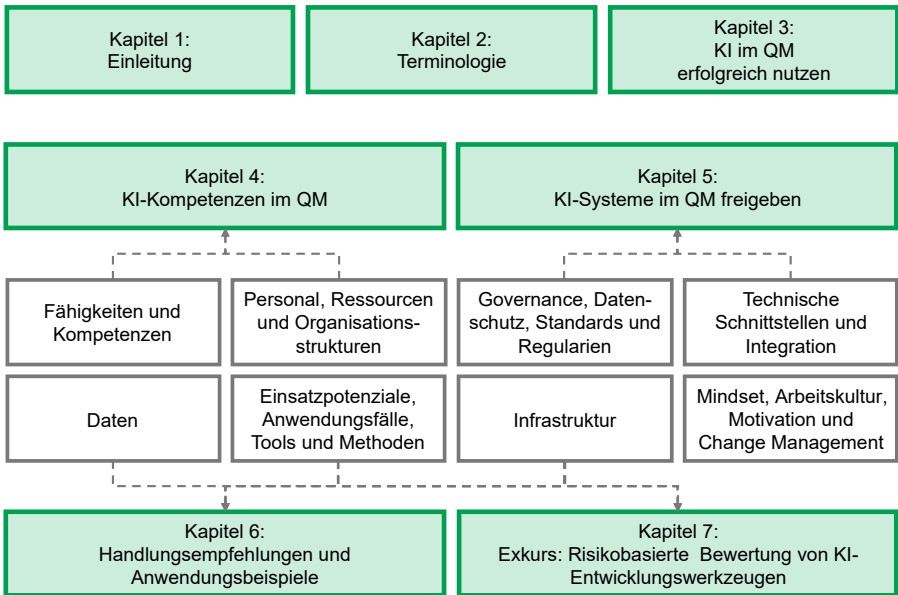


Abbildung 3-2: Erfolgsfaktoren für die Nutzung von KI und Übersicht der Kapitelstruktur

Der folgende Abschnitt gibt einen Überblick über diese Aspekte, die in acht Blöcke unterteilt sind, in denen die Erfolgsfaktoren und Herausforderungen für die erfolgreiche Nutzung von KI in der Organisation hervorgehoben werden. Gleichzeitig bilden diese Blöcke auch die Grundlage für eine eingehende Betrachtung in den folgenden Kapiteln.

3.1 Mindset, Arbeitskultur, Motivation und Change Management

Die erfolgreiche Einführung von KI im Qualitätsmanagement hängt nicht allein von technischen Lösungen ab, sondern in hohem Maße auch vom Mindset, der Arbeitskultur und der Veränderungsbereitschaft in einer Organisation. Unter diesen Begriffen versteht man die gemeinsame Haltung, Werte und Verhaltensweisen, die bestimmen, wie Menschen mit neuen Technologien, Unsicherheiten und Lernprozessen umgehen. Dazu gehören Offenheit, Lernbereitschaft, Vertrauen in Daten, interdisziplinäre Zusammenarbeit und der Mut, etablierte Prozesse zu hinterfragen. Eine innovationsfreundliche Arbeitskultur fördert die Akzeptanz von KI.

Diese Faktoren sind entscheidend, da KI-Systeme nicht einfach bestehende Abläufe automatisieren, sondern Arbeitsweisen, Rollen und Entscheidungsprozesse grundlegend verändern können. Nur wenn Mitarbeitende die Vorteile verstehen, Verantwortung übernehmen und Vertrauen in KI-basierte Entscheidungen entwickeln, kann das volle Potenzial ausgeschöpft werden. Die Zusammenarbeit zwischen Mensch und Maschine verlangt neue Formen von Kommunikation, Führung und Motivation. Damit wird die Weiterentwicklung des Mindsets zu einem zentralen Erfolgsfaktor auf dem Weg zu einer zukunftsfähigen Qualitätskultur.

Erfolgsfaktoren

Herausforderungen

- **Offenheit und Lernbereitschaft:** Zulassung von neuen Ideen und Perspektiven sowie Motivation und Wille, sich neues Wissen anzueignen und weiterzuentwickeln
- **Iterative Verbesserung:** Akzeptanz, dass KI-Lösungen nicht sofort perfekt funktionieren und eine iterative Verbesserung notwendig ist
- **Transparenz:** Klare Kommunikation, wie und warum KI eingesetzt wird und was sie kann und (noch) nicht kann
- **Partizipation und Mitgestaltung:** Aktive Einbeziehung der Mitarbeitenden in die Entwicklung und Einführung von KI-Lösungen
- **Wertschätzung von Erfahrungswissen:** Kombination von KI und menschliche Expertise und Intuition als Stärke begreifen
- **Führungskräfte als Wegbereiter:** Veränderung vorleben, Orientierung geben und Offenheit für den Dialog
- **Zugang zu KI-Coaches:** Ermöglichen von lokalen Multiplikatoren, die Akzeptanz fördern, Anwender:innen unterstützen und Best Practices im Unternehmen verbreiten
- **Angst vor Job-, Autonomie- oder Kontrollverlust:** „Ich werde überflüssig“, „KI ersetzt mich“, „KI entscheidet“
- **Misstrauen in die Technologie:** Fehlendes Verständnis für die Funktionsweise der KI kann zu Ablehnung oder Unsicherheit führen
- **Überhöhte Erwartungen und blindes Vertrauen:** „KI löst alles“, Wissenslücke oder fehlendes Verständnis bei Entscheidern auf allen Ebenen
- **Unklarer persönlicher Nutzen:** Fehlende Erkennung des direkten Vorteils für den eigenen Arbeitsalltag
- **Einstiegshürde durch neue Tools:** Erhöhter Aufwand für Datenpflege, Schulungen und Einarbeitungszeit
- **Unklare Verantwortlichkeiten:** Nicht eindeutig definierte Rollen und Zuständigkeiten führen zu Unsicherheit im Umgang mit der Technologie („Was darf ich mit der KI tun, was ist erlaubt, was nicht?“)
- **Erzwingen von Technologie:** Einführung ohne Einbeziehung der Mitarbeiter:innen
- **Widerstand gegenüber Veränderungen:** Festhalten an vertrauten Abläufen und Routinen

3.2 Fähigkeiten und Kompetenzen

Der Einsatz von KI erweitert das Kompetenzspektrum im Qualitätsmanagement deutlich. Neben klassischem Qualitätswissen werden nun Datenanalyse und algorithmisches Verständnis zu zentralen Fähigkeiten. KI-Systeme fordern von Fachkräften nicht nur technisches Fachwissen, sondern auch die Fähigkeit, Modelle zu interpretieren, deren Ergebnisse zu analysieren und interdisziplinär zu agieren. So entsteht ein neues Kompetenzprofil, das Qualitätsmanagement stärker mit Data Science, IT und intelligenten Prozessen verbindet.

Erfolgsfaktoren

- **Verständnis von KI-Grundlagen:** Kenntnisse in Fachbereichen wie Statistik, Datenanalysen, ML und Algorithmen
- **KI-Tools und Plattformen:** Grundlegende Kenntnisse in Nutzung und Aufbau von KI-Tools
- **Interdisziplinäres Fachwissen:** Kombination aus verschiedenen Disziplinen, bspw. IT, Recht, Technik und Ingenieurwesen
- **Freiraum für kontinuierliches Lernen:** Erweiterung von Fähigkeiten und Aufbau von Kompetenzen
- **Schulungskonzepte:** Vermittlung über verschiedene Medien und Formate wie Online-Kurse, Schulungen und Workshops

Herausforderungen

- **Fehlende Datenkompetenz:** Fachliche Einschränkungen bei der Erstellung, Interpretation, Auswertung, Evaluation und sinnvollen Nutzung von Daten
- **Technologischer Wandel:** Schnelle Entwicklungen bei digitalen und KI-basierten Technologien erfordern laufende Weiterentwicklung von Kompetenzen, Methoden, Tools und Rollen.
- **Technologische Komplexität:** Technische Komplexität und Weitläufigkeit des KI-Themenfeldes

3.3 Daten

Daten bilden die zentrale Grundlage für den erfolgreichen Einsatz von KI im Qualitätsmanagement. Unter diesem Begriff versteht man nicht nur Mess- und Prüfergebnisse, sondern auch Prozess-, Produkt- und Kontextinformationen, die Qualität bewertbar machen. Ihre Verfügbarkeit, Qualität und Struktur entscheiden darüber, wie zuverlässig KI-Modelle arbeiten können. Gleichzeitig verändert die Einführung von KI den Umgang mit Daten selbst, von der Datenerfassung über die Aufbereitung bis hin zur Interpretation. Konzepte und Strategien zur Sammlung, Speicherung und Verwaltung von Daten sind unerlässlich für eine hohe Datenqualität. Damit wird ein systematisches Datenmanagement zum Schlüsselfaktor, um aus Informationen belastbare Erkenntnisse und nachhaltige Verbesserungen abzuleiten.

Erfolgsfaktoren

- **Datenqualität:** Vollständigkeit, Korrektheit, Konsistenz, Aktualität und Relevanz der Daten. Referenzen: ISO 8000, ISO/IEC 25012
- **Datenmanagement:** Erfassung, Speicherung und Pflege von Daten zur Verbesserung der Datenqualität
- **Standardisierung:** Einheitlichkeit der Formate zur Erfassung, zum Austausch und Management von Daten
- **Integration:** Integration und Aggregation aus unterschiedlichen Quellen
- **Automatisierung:** Verringerung der Störfaktoren, die auf die Qualität einwirken
- **Systemwartung:** Regelmäßige Überprüfung der Datenqualität und -

Herausforderungen

- **Variabilität der Anforderungen an Datenqualität:** Vielfältigkeit und Variabilität der Datenqualitätskriterien für KI je nach Anwendungsfall
- **Datenverfügbarkeit und Skalierbarkeit:** Unzureichende Datenmenge für KI-Anwendungen
- **Planung der erforderlichen Daten:** Ungenaue Einschätzung der benötigten Datenmengen und -qualität
- **Datenzugriff:** Zugangsbeschränkungen oder Datenschutzrichtlinien für erforderliche Daten
- **Aktualität der Daten:** Schnelle Veränderung der Qualitätsanforderungen aufgrund dynamischer Marktbedingungen und immer kürzerer Produktlebenszyklen

Erfolgsfaktoren

Herausforderungen

Stabilität (Veränderung der Variabilität über die Zeit)

3.4 Organisationsstrukturen und Prozesse

Neben personellen Ressourcen und technischen Kompetenzen hängt die erfolgreiche Einführung von KI im Qualitätsmanagement maßgeblich von ausreichend Kapazitäten, klaren Verantwortlichkeiten und Strukturen ab, die agile Zusammenarbeit und kontinuierliches Lernen ermöglichen. KI verändert dabei die Art, wie Teams organisiert, Ressourcen verteilt und Entscheidungswege gestaltet werden. Dadurch entstehen neue Anforderungen an Führung, Rollenverständnis und interdisziplinäre Zusammenarbeit, die für eine nachhaltige Integration von KI-Systemen entscheidend sind.

Erfolgsfaktoren

Herausforderungen

- **Ressourcen:** Genügende Ressourcen für die Entwicklung und Nutzung von KI-Systemen
- **Motivation und Mindset:** Motivation und Bereitschaft, sich mit einer neuen, KI-gesteuerten Umgebung auseinanderzusetzen und das System zu verbessern
- **Kompetenzen:** Befähigung zur Nutzung einer neuen, KI-gesteuerten Umgebung durch Schulungen
- **Organisation:** Ausrichtung und Anpassung der Organisationsstrukturen und Rollen an neue KI-Systeme
- **Silo-Denken:** Fehlendes Know-how oder fehlende Zusammenarbeit, um eine effiziente KI-gesteuerte Umgebung aufzubauen
- **Unzureichende Integration in Organisationsstruktur:** Geringer Digitalisierungsgrad und Integration von KI-Tools in Qualitätsmanagementprozessen
- **Unzureichende Integration in Organisationskultur:** Misstrauen, überhöhte Erwartungen oder Widerstand gegenüber KI-Technologien

Erfolgsfaktoren

- **Infrastruktur:** Vertrautheit mit den Schnittstellen zum KI-System zwecks Datenaufbereitung
- **Zielgerichtete Kommunikation:** Klare Informationen und regelmäßiger Austausch auch über neue Kommunikationswege, -kanäle und -formate
- **KI- und Daten-Governance:** Klare und schnelle Entscheidungen, erreichbare Ansprechpersonen

Herausforderungen

3.5 Governance, Datenschutz, Standards und Regularien

Governance und Datenschutz beziehen sich auf die Rahmenbedingungen, die sicherstellen, dass KI verantwortungsvoll entwickelt und eingesetzt wird. Es sind Mechanismen erforderlich, um Compliance-Vorgaben zu definieren, die den rechtlichen Anforderungen entsprechen. Insbesondere der Umgang mit sensiblen Daten erfordert umfassende Datenschutzrichtlinien. Dies bedeutet, dass Unternehmen klare Richtlinien benötigen, wie Daten erhoben, gespeichert und genutzt werden. Die Einhaltung nationaler und internationaler Standards und Regularien ist essenziell für die sichere Entwicklung und Nutzung von KI, um unvorhergesehene Risiken für Qualität oder Sicherheit zu vermeiden. Diese Aspekte sind entscheidend für das Vertrauen in KI-Systeme; ohne diese können rechtliche Risiken und Reputationsschäden entstehen.

Erfolgsfaktoren

- **Klare Verantwortlichkeiten:** Definition von Rollen und Zuständigkeiten für die Entwicklung, den Betrieb und die Überwachung von KI-Anwendungen

Herausforderungen

- **Unklare regulatorische Lage:** Unsicherheit über geltende Vorschriften, z. B. EU AI Act, DSGVO und Produkthaftung

Erfolgsfaktoren

- **Regulatorische Konformität:** Entwicklung von KI-Systemen von Anfang an unter Berücksichtigung regulatorischer Anforderungen
- **Datenschutzkonforme Datenverarbeitung:** Einhaltung von Datenschutzrichtlinien bspw. durch Anonymisierung, Pseudonymisierung oder Einwilligungsmanagement
- **Standardisierte Prozesse für KI-Validierung:** Überwachung der KI-Systeme bspw. durch regelmäßige Modellüberprüfung, Audit-Trails und Dokumentation
- **Einsatz von etablierten Normen und Frameworks:** Orientierung an Regelwerken und Normen wie ISO 9001, ISO/IEC 27001, ISO/IEC 23894 (KI-Risiken) und EU AI Act
- **Transparenz und Nachvollziehbarkeit:** Erklärbarkeit und Dokumentation der Modellergebnisse und der KI-Entscheidungen
- **Ethikrichtlinien für KI:** Berücksichtigung und Einhaltung von Aspekten wie Fairness, Nichtdiskriminierung und Mensch-zentrierte Entwicklung
- **Interdisziplinäre Governance-Teams:** Einbindung von verschiedenen Fachbereichen wie QM, IT und Legal

Herausforderungen

- **Fehlende interne Richtlinien für KI:** Fehlende Prozesse für die Auswahl, den Einsatz und die Kontrolle von KI-Systemen
- **Datenschutzkonflikte:** Ungenehmigte Nutzung von Daten, bspw. personenbezogene Daten für Trainingszwecke
- **Mangelnde Auditierbarkeit:** Geringe Nachvollziehbarkeit und fehlende Dokumentation der KI-Modelle
- **Vertrauensverlust durch Intransparenz:** Geringe Nachvollziehbarkeit oder hohe Unsicherheit der Ergebnisse und Entscheidungen von KI-Modellen
- **Fehlende einheitliche Standards für KI:** Fehlende etablierte Benchmarks oder Validierungsmethoden für KI-Modelle
- **Ungeklärte Haftungsfragen:** Fehlende Definition der Verantwortung bei Ungenauigkeiten und Fehlentscheidungen der KI-Modelle
- **Langsame Anpassung bestehender QM-Systeme:** Mangelnde Integration in bestehende QM-Methoden, bspw. FMEA (Fehlermöglichkeits- und Einflussanalyse) oder CAPA (Corrective and Preventive Action) nicht auf KI-Risiken ausgelegt

3.6 Infrastruktur

Infrastruktur bezieht sich auf die grundlegenden Systeme und Technologien, die erforderlich sind, um KI erfolgreich in Organisationen zu implementieren. Eine moderne IT-Infrastruktur bildet das Rückgrat für die Entwicklung und den Betrieb von KI-Anwendungen. Sie umfasst alles von Hardware wie leistungsstarken Servern und spezialisierten Prozessoren über Softwarelösungen und Konzepte, die den Austausch und die Verarbeitung von Daten sowie die Ausführung von Algorithmen ermöglichen, bis hin zu Cloud-Plattformen und Datenräumen sowie technischen Lösungen und Strategien für Datenhoheit und Cyber-Security. Dies umfasst auch verschiedene Modelle, die zum Inhalt haben, wie und in welchem Umfang die IT-Infrastruktur selbst (On-Premise) oder von einem Dienstleister verwaltet wird, z. B. Infrastructure as a Service (IaaS), Platform as a Service (PaaS) und Software as a Service (SaaS). Darüber hinaus spielt die Netzwerkinfrastruktur eine entscheidende Rolle bei der Gewährleistung einer stabilen Verbindung zwischen verschiedenen Systemen und der schnellen Übertragung großer Datenmengen. Ein zentraler Erfolgsfaktor für moderne Infrastruktur ist die nahtlose Integration verschiedener Systeme innerhalb der Organisation sowie mit externen Partnern (z. B. für den Austausch von Qualitätsdaten).

Erfolgsfaktoren

- **Dateninfrastruktur:** Erfassung (Sensorik), Speicherung und Management von Daten in KI-fähigen Formaten
- **Schnittstellen und Integration:** Interoperabilität von Systemen und Integration in bestehende Systemlandschaft
- **Skalierbarkeit:** Technische Erweiterung, Zugriff und Leistung von Systemen in der Organisation
- **Standardisierung:** Einheitliche Formate und Systeme in der Organisation

Herausforderungen

- **Gewachsene IT-Infrastruktur:** Geringe Flexibilität und Legacy-Systeme mit verschiedenen Tools
- **Skalierbarkeit und Leistung:** Hohe Anforderungen an Rechenleistung und Speicher
- **Einbettung von KI:** Begrenzte Integration und Interoperabilität mit bestehenden Workflows und Systemen

Erfolgsfaktoren

Herausforderungen

- **Ressourcen:** Hohe Investitionskosten für Infrastruktur
- **Technologische Komplexität:** Komplexität der IT-Systeme und Schnittstellen in der Organisation
- **Lokalisierung von Infrastruktur:** Bestimmte Daten unterliegen länder- oder regionalspezifischen Vorschriften, die einen nahtlosen Austausch erschweren
- **Cyber-Security:** Zunehmende Sicherheitsbedrohungen gefährden die Stabilität der Infrastruktur angesichts verschärfter Vorschriften zum Datenschutz
- **Datensouveränität:** Komplexe rechtliche und regulatorische Bedingungen erschweren den Austausch von Daten

3.7 Technische Schnittstellen und Integration

Schnittstellen und technische Integration beziehen sich auf die Verbindungen zwischen verschiedenen Software- und Hardwaresystemen, die es ermöglichen, Daten auszutauschen und Prozesse zu koordinieren. Eine nahtlose Integration in bestehende Systeme ist entscheidend für den Erfolg von KI-Anwendungen in Organisationen. Gut definierte Schnittstellen sind erforderlich, um sicherzustellen, dass verschiedene Technologien harmonisch zusammenarbeiten und Informationen effizient fließen können. Durch die Verbindung von KI mit bestehenden Produktions- oder Qualitätsmanagementsystemen können Unternehmen Daten aus verschiedenen Quellen nutzen, bspw. Maschinen-Sensorik oder Kundenfeedback-Systeme. Darüber hinaus verbessert eine gut integrierte Infrastruktur die Zusammenarbeit zwischen Abteilungen. Bspw. können Produktions-, Qualitäts- und

Entwicklungsteams aufgrund gemeinsamer Datenzugriffe besser kommunizieren und kooperieren, was zu einer schnelleren Problemlösung führt.

Erfolgsfaktoren

Herausforderungen

- **Definition Schnittstellen:** Klare Definition der Schnittstellen zwischen den verschiedenen KI-Systemen
 - **Vernetzung:** Vernetzung verwandter KI-Anwendungen und Agenten, wobei die Ergebnisse für die weitere Verarbeitung als Daten zur Verfügung stehen
 - **Durchgängigkeit:** Verknüpfung der KI-Systeme in Qualitätsmanagementfunktionen mit anderen Funktionen und Domänen in der Organisation
 - **Integration:** Integration der KI-Systeme in Qualitätsmanagementsysteme, Umgebungen und Prozesse
- **Fehlende Einheitlichkeit:** Unklare oder fehlende Definition einheitlicher und ganzheitlicher Schnittstellen
 - **Designkonzept-Silos:** Fehlende Designkonzepte für KI-Agenten, die später miteinander integriert werden sollen
 - **Daten-Silos:** Fehlende Datenkonzepte für KI-Systeme, die Daten für andere Systeme aufbereiten

3.8 Einsatzpotenziale, Anwendungsfälle, Tools und Methoden

Der Einsatz von KI im Qualitätsmanagement eröffnet neue Chancen, von effizienteren Analysen bis hin zur vorausschauenden Qualitätssicherung. Doch der Erfolg hängt nicht allein von den verfügbaren Tools oder Methoden ab, sondern auch von ihrer zielgerichteten Anwendung und der Überwindung zentraler Herausforderungen wie Datenverfügbarkeit, Prozessintegration und Nutzerakzeptanz. Der erfolgreiche Einsatz von KI erfordert eine Kombination aus klar definierten Zielen und Anforderungen an Prozesse, Methoden und Tools. Konkrete Anwendungsfälle schaffen Transparenz über die Anforderungen für den zielgerichteten Einsatz von KI.

Erfolgsfaktoren

- **Klare Problemdefinition:** Eindeutige und messbare Definition des anvisierten Mehrwerts für den Einsatz von KI, bspw. in den Bereichen Fehlererkennung und Prozessoptimierung
- **Passende Tool-Auswahl:** Auswahl von KI-Tools, die zur Datenlage, zum Anwendungsfall und zur Nutzergruppe passen
- **Tool-Usability:** Förderung der Akzeptanz durch intuitive Bedienung und gute Visualisierung
- **Integration in bestehende Systeme:** Identifizierung potenzieller Verbesserungen und Ergänzungen zu bestehenden QM-Methoden wie SPC, FMEA oder 8D-Report
- **Pilotierung und iteratives Vorgehen:** Etablierung kleiner und kontrollierter Tests mit klaren Kennzahlen zur Bewertung des Nutzens
- **Datenqualität und -verfügbarkeit:** Strukturierte, saubere und ausreichend große Datenmengen als Basis für eine effektive und effiziente Nutzung von KI
- **Erfolgreiche Anwendungsfälle:** Förderung der Sichtbarkeit und Austausch von Best Practices und Erfolgsgeschichten

Herausforderungen

- **Unklare Zielsetzung:** Einführung von KI ohne konkretes Ziel und vorstellbaren Nutzen
- **Komplexität der KI-Modelle:** Überkomplexität der KI-Modelle für die Anwendung oder das vorhandene Fachwissen
- **Tool-Overload:** Überfluss an Tools ohne klare Abgrenzung oder Integration
- **Fehlende, fehlerhafte oder unvollständige Datenbasis:** Fehlende oder unbrauchbare historische Daten, bspw. unstrukturiert oder manuell erfasst. Fragmentierte Daten, mangelnde Verfügbarkeit oder unzureichende Datenqualität
- **Skalierungsaufwand:** Funktionsfähige Piloten lassen sich nur eingeschränkt oder mit hohem Aufwand in den stabilen Produktivbetrieb überführen
- **Fehlende Standardisierung:** Eingeschränkte Vergleichbarkeit und Reproduzierbarkeit aufgrund unterschiedlicher Tools und Methoden
- **Fehlende Berücksichtigung regulatorischer Anforderungen:** Eingeschränkte Möglichkeit der Validierung, Auditierbarkeit und Nachvollziehbarkeit von KI-Modellen

4 KI-Kompetenzen im QM

Der Einsatz von KI im Qualitätsmanagement verändert nicht nur Aufgaben und Prozesse, sondern auch die notwendigen Fähigkeiten und Kompetenzen der Mitarbeiter:innen.

Ziel dieses Kapitels ist es, den Mitarbeiter:innen in Qualitätsfunktionen einen Überblick über relevante **Kompetenzanforderungen** zur Verfügung zu stellen, damit sie in ihren aktuellen Rollen durch die Nutzung von KI effektiver und effizienter arbeiten können. Ebenso dienen die Kompetenzanforderungen der Weiterentwicklung und Vorbereitung auf zukünftige Tätigkeits- und Verantwortungsumfänge.

KI-Kompetenz bezeichnet gemäß dem EU AI Act die Fähigkeiten, die Kenntnisse und das Verständnis, die es Anbietern, Betreibern und Betroffenen unter Berücksichtigung ihrer jeweiligen Rechte und Pflichten ermöglichen, KI-Systeme **sachkundig einzusetzen** sowie sich der **Chancen und Risiken** von KI und möglicher Schäden, die sie verursachen kann, bewusst zu werden. Die KI-Kompetenz sollte Anbieter, Betreiber und betroffene Personen mit den notwendigen Konzepten ausstatten, um fundierte Entscheidungen über KI-Systeme zu treffen. Diese Konzepte können in Bezug auf den jeweiligen Kontext unterschiedlich sein. Sie beinhalten drei Bereiche zum Verständnis der korrekten Anwendung von KI-Systemen:

- die **Entwicklungsphase** des KI-Systems,
- die bei der **Verwendung** anzuwendenden Maßnahmen und
- die angemessene **Nutzung der Ausgaben** des KI-Systems

Zudem umfassen die Konzepte das nötige Wissen, um zu verstehen, wie sich mit Hilfe von KI getroffene Entscheidungen auf die davon eventuell betroffenen Personen auswirken können. Der EU AI Act definiert die Rahmenbedingungen für KI-Kompetenzen. Unternehmen müssen sicherstellen, dass ihre Mitarbeiter:innen, je nach Rolle im Umgang mit KI, über die entsprechenden Kompetenzen verfügen. Die Kompetenzanforderungen berücksichtigen die Vorgaben des EU AI Act und der ISO42001.

4.1 Haupt-Kompetenzen für KI im Qualitätsmanagement

Die wesentlichen Kompetenzen im Umgang mit KI-Systemen im Qualitätsmanagement sind folgende **vier Haupt-Kompetenzen**:

1. Grundverständnis von KI/ML-Konzepten (KI-Modelle, Datenmodelle, Modellbewertung, Validierung)

Die Mitarbeiter:innen kennen typische KI-Anwendungen im Qualitätsmanagement – darunter Predictive Quality, Anomalieerkennung, Computer Vision und prozessbezogene Optimierungsansätze – und sind in der Lage, diese hinsichtlich ihres Nutzens sowie der Anwendungsvoraussetzungen (z. B. hohe Datenqualität) beurteilen zu können.

Sie können den Unterschied zwischen klassischen regelbasierten Systemen und maschinellem Lernen erklären und deren Einsatzgrenzen einordnen.

Darüber hinaus verstehen die Mitarbeiter:innen grundlegende ML-Konzepte wie Supervised und Unsupervised Learning, können mit den Trainings- und Testdaten umgehen sowie geeignete Modelltypen wie Klassifikation, Regression oder Anomalieerkennung auswählen.

2. Grundverständnis von Datenanalyse- und Statistikenkenntnissen (Big Data, Visualisierung, Datenqualität)

Die Mitarbeiter:innen verstehen die Notwendigkeit guter und ausreichender Datenmengen (Vollständigkeit, Genauigkeit und Konsistenz) als Voraussetzung für erfolgreiche KI-Ergebnisse und können die Konsequenzen einschätzen, falls verzerrte Daten für Analysen genutzt werden.

Sie begreifen, dass unzureichende Datenqualität zu falschen Entscheidungen führen kann. Besonders im Qualitätsmanagement tritt das Problem auf, dass Fehler selten und oft nicht reproduzierbar sind, so dass Daten von Fehlteilen nicht in sehr großer Zahl vorkommen.

Es ist den Mitarbeiter:innen bewusst, dass verzerrte Daten zu verzerrten KI-Modellen und damit zu schlechten Vorhersagen („Garbage-in/Garbage-out“) führen können.

3. Umgang mit generativer KI (LLMs) (inkl. Prompting)

Die Mitarbeiter:innen sind in der Lage, für den praktischen Einsatz generativer KI (LLMs) entsprechende Tools, wie die strukturierte Erstellung und Überarbeitung von Dokumentationen (8D-Reports oder FMEA-Beschreibungen), effiziente Recherche von Best Practices und normativen Anforderungen sowie die Optimierung schriftlicher Kommunikation in Form von E-Mails oder Berichten, gezielt zwecks Arbeitserleichterung einzusetzen.

Sie können Anweisungen in Form von grundlegenden Prompt-Engineering-Techniken klar und präzise formulieren, um verlässliche und arbeitsrelevante Ergebnisse zu erhalten.

Die Mitarbeiter:innen sind in der Lage, erzeugte Inhalte kritisch zu validieren und auf mögliche Halluzinationen zu prüfen, um Fehlinformationen zu vermeiden und eine sachlich korrekte Nutzung sicherzustellen.

4. Bewusstsein für KI-Chancen, Risiken, KI-Ethik, KI-Rechtsgrundlagen und mögliche Schäden

Die Mitarbeiter:innen sind in der Lage, die Ergebnisse von KI-Systemen zu interpretieren, indem sie Wahrscheinlichkeiten, Zuverlässigkeiten und modellbasierte Vorhersagen verstehen und in den jeweiligen Anwendungskontext einordnen. Dabei hinterfragen sie kritisch die KI-Ergebnisse, gleichen sie mit fachlichem Expertenwissen ab und prüfen ihre Plausibilität. Ein grundlegendes Verständnis für Erklärbarkeit ist hierbei essenziell, um nachvollziehen zu können, warum ein Modell zu einer bestimmten Entscheidung gelangt ist – insbesondere im Hinblick auf Auditierbarkeit und Compliance-Anforderungen.

Die Mitarbeiter:innen kennen die ethischen, rechtlichen und organisatorischen Rahmenbedingungen und setzen diese verantwortungsvoll um. Dazu zählen:

1. der bewusste Umgang mit personenbezogenen Daten im Sinne der DSGVO,
2. die Sensibilität für mögliche Bias-Effekte in KI-Modellen – etwa durch verzerrte Daten, die zu unfairen Bewertungen führen können –, und
3. das Verständnis, dass KI Entscheidungen unterstützt, jedoch nicht die menschliche Verantwortung ersetzt

Darüber hinaus sind die Mitarbeiter:innen in der Lage, Transparenz sicherzustellen, indem die Nutzung von KI-Systemen nachvollziehbar dokumentiert wird, um regulatorische und auditbezogene Anforderungen erfüllen zu können.

Anmerkung: Aufgrund einer immer einfacher werdenden Toollandschaft können mit KI auch komplexere Problemstellungen gelöst werden. Beispiele:

- Programmcode kann über ein LLM generiert werden und vollständige Applikationen bilden
- Für die Automatisierung von Prozessabläufen sind oftmals keine tiefen Programmierkenntnisse mehr nötig, da immer mehr No-Code oder Low-Code-Tools (die ohne Programmcode oder mit sehr wenig davon auskommen) direkt zum Erstellen von Applikationen genutzt werden
- Agentic-Workflows, die eigenständig agieren und Entscheidungen treffen

An dieser Stelle ist besondere Vorsicht geboten und die damit einhergehenden Risiken müssen den Mitarbeiter:innen bewusst sein.

4.2 Rollen im Qualitätsmanagement und relevante KI-Kompetenzen

Die vier Haupt-Kompetenzen sind grundsätzlich für sehr viele der Arbeitstätigkeiten und somit für die Beschäftigten im Qualitätsmanagement

relevant. Ergänzend und um die Identifikation und Anwendung für die Mitarbeiter:innen zu erleichtern und zu konkretisieren, hilft es, die notwendigen KI-Kompetenzen mit den **typischen Rollen im Qualitätsmanagement** zu verknüpfen:

4.2.1 Klassische Rollen im Qualitätsmanagement

In diesem Band werden **sieben „klassische“ Rollen** im Qualitätsmanagement der Automobilindustrie beschrieben, die den Großteil der Qualitätstätigkeiten abdecken und in denen sich Mitarbeiter:innen in Qualitätsfunktionen wiederfinden können. Zu diesen Rollen sind hier beispielhafte, KI-spezifische Kompetenzanforderungen definiert, die dazu beitragen können, die bisherigen Tätigkeiten mit Hilfe von KI effektiver und effizienter zu erledigen.

Sieben klassische Rollenprofile:

- **Mitarbeiter:innen in der Lieferantenqualität** (SQE, SDE)
- **Mitarbeiter:innen in der Entwicklungsqualität** (Produktentwicklung)
- **Mitarbeiter:innen in der Produktionsqualität** (Prozessentwicklung, Industrialisierung, Serienproduktion)
- **Mitarbeiter:innen in der Kundenqualität** (Reklamationen, Gewährleistung, Feldbeobachtung)
- **Mitarbeiter:innen im Qualitätsmanagementsystem** (QMS-Verantwortliche)
- **Qualitäts-Auditor:in** (System, Prozess, Produkt) und **Qualitäts-Assessor:in** (ASPICE)
- **Qualitätsführungskraft** (Leitungsfunktion für Personal, Budget, Strategie)

Auf die typischen Arbeitsinhalte dieser Rollen wird an dieser Stelle nicht im Besonderen eingegangen, da diese grundlegend bekannt und u. a. über andere VDA-Regelwerke mit beschrieben sind (z. B. im Rahmen der Reifegrad- oder Audit-Richtlinien).

Bei den Rollen handelt es sich um generische Beschreibungen, nicht um konkrete Positions-/Funktionsbeschreibungen. Der genaue Tätigkeits- und Verantwortungsumfang einzelner Mitarbeiter:innen kann dabei teils mehrere Rollen beinhalten.

4.2.2 Neue Rollen im Qualitätsmanagement

Aufgrund des steigenden Einsatzes von KI in Qualitätstätigkeiten werden in diesem Band zusätzlich **vier „neue“ Rollen** im Qualitätsmanagement der Automobilindustrie identifiziert, die sich daraus ergeben (können).

Zu diesen Rollen sind beispielhafte, KI-spezifische Kompetenzanforderungen definiert, die dazu beitragen können, die neuen Rollen umfänglich auszuüben. In der Praxis werden Mitarbeiter:innen im Qualitätsmanagement sich in ihrer Rolle zunehmend mit KI-Nutzung beschäftigen (Ausfüllen der „klassischen“ Rolle mit entsprechender KI-Ergänzung) und sich dann teilweise zur Ausübung „neuer“ Rollen hinentwickeln. Je intensiver KI in Unternehmen und damit auch in der Qualität genutzt wird, desto mehr Rollen und Unterthemen sind denkbar. Deren Betrachtung wird in diesem Band nicht vorgenommen.

Bei den Rollen handelt es sich um generische Beschreibungen, nicht um konkrete Positions-/Funktionsbeschreibungen. Der genaue Tätigkeits- und Verantwortungsumfang einzelner Mitarbeiter:innen kann dabei teils mehrere Rollen beinhalten.

- **AI Q-Data Engineer** (Fokus: Datenerfassung und -aufbereitung)

Ein:e Quality Data Engineer baut und betreibt die Dateninfrastruktur für QM-Anwendungen. Sie/Er entwickelt automatisierte Datenpipelines, integriert heterogene Quellen (Sensoren, MES, ERP), sichert Datenqualität und implementiert Monitoring-Systeme. Sie/Er arbeitet mit Cloud-Technologien, etabliert DevOps-Praktiken und schafft skalierbare Architekturen. Ziel: Zuverlässige, performante Daten für Analyst:innen und Data Scientists bereitstellen.

- **AI Q-Data Analyst** (Fokus: Datenanalyse und -visualisierung)

Ein:e Quality Data Analyst analysiert qualitätsrelevante Daten aus Produktion und Prüfprozessen, um Muster, Abweichungen und Verbesserungspotenziale zu identifizieren. Sie/Er wendet

statistische Methoden (SPC, Cpk) an, nutzt KI-Tools zur Muster- und Anomalieerkennung, erstellt Dashboards und kommuniziert Erkenntnisse an Stakeholder. Ziel: Datengetriebene Qualitätsverbesserung und Fehlerprävention betreiben.

- **AI Q-Data Scientist** (Fokus: Entwicklung von Datenanalyse-Methoden)

Ein:e Quality Data Scientist entwickelt und trainiert Machine-Learning-Modelle für QM-spezifische Anwendungen wie Fehlervorhersage, Anomalieerkennung oder optische Qualitätsprüfung. Sie/Er führt Feature Engineering durch, evaluiert Modellperformance, stellt Interpretierbarkeit sicher und arbeitet mit Data Engineers am Deployment zusammen. Ziel: Innovative KI-Lösungen für Qualitätsverbesserung und Predictive Quality entwickeln.

- **AI Q-Data Manager** (Fokus: KI-Strategie, Daten-Governance und -Compliance)

Ein:e Data Manager:in entwickelt und steuert die Datenstrategie im QM-Bereich, etabliert Data Governance und definiert Standards. Sie/Er baut interdisziplinäre Teams auf, wählt Technologien aus, managt Budgets und orchestriert die Zusammenarbeit zwischen IT, QM und Fachbereichen. Ziel: Datengetriebene Transformation vorantreiben, Business Value schaffen und Compliance sicherstellen.

4.2.3 Rollenspezifische KI-Kompetenzen

Die **KI-Kompetenzanforderungen** umfassen die o. g. vier Haupt-Anforderungen (identisch für alle „klassischen“ Rollen), die dann für die jeweiligen Rollen in konkreten Kompetenzanforderungen detailliert sind. Dies hilft dabei, eine Zuordnung zur Verfügung zu stellen: von typischen Aufgaben und Tätigkeiten in den einzelnen Rollen zu konkreten Kompetenzen und beispielhafter KI-Nutzung. Den Mitarbeiter:innen in Qualitätsfunktionen erleichtert das die **Identifikation mit den eigenen Aufgaben** und den sich daraus ergebenden KI-Kompetenzen. Wenn komplexere Applikationen wie beispielsweise größere KI-gesteuerte Prozessautomatisierungen aufgebaut werden, sind zusätzlich noch KI-Ingenieur:innen und Software-Ingenieur:innen nötig. Diese benötigen

jedoch keine tieferen Kenntnisse im Qualitätsmanagement und sind daher hier nicht beschrieben.

Im Anhang 1 sind die o. g. Rollen und deren empfohlene KI-Kompetenzen mit beispielhafter KI-Nutzung beschrieben.

Selbstverständlich gibt es für die einzelnen Kompetenzanforderungen unterschiedliche Ausprägungen, abhängig von der genauen Aufgabenanforderung, der unternehmensspezifischen Ausprägung von Stellen- und Arbeitsinhalten und der bisherigen und zukünftig notwendigen Kompetenz individueller Mitarbeiter:innen. Grundsätzlich lässt sich sagen, dass in den „klassischen“ Qualitätsrollen ein **KI-Kompetenzlevel** von „**Basis**“ (= Mitarbeiter:in versteht und wendet an) bis „**Fortgeschritten**“ (= Mitarbeiter:in meistert den Umgang) meist ausreichend ist. Bei den „neuen“ Rollen geht das relevante KI-Kompetenzlevel dann hin bis zu „**Expertin/Experte**“ (= Mitarbeiter:in gestaltet und schult). Diese Ausprägung ist jedoch, wie bereits erwähnt, stark vom individuellen Stellen- und Funktionsprofil der Mitarbeiter:innen abhängig.

5 KI-Systeme im QM freigeben

Dieser Abschnitt beschreibt das Vorgehen, wie mögliche Risiken eines KI-Systems, das für einen bestimmten Einsatzzweck entwickelt wurde, ermittelt und als Basis für eine Freigabe genutzt werden können. Diese Risiken sollten spätestens vor Inbetriebnahme des KI-Systems ermittelt und bewertet werden, es wird aber empfohlen, bereits bei der Planung und Entwicklung des Systems die Kriterien aus dieser Methode anzuwenden.

Die Ermittlung der Risiken geschieht in zwei Schritten:

- Im ersten Schritt erfolgt eine Projektrisikobewertung für den Umfang der KI-bezogenen Aufgabe im Hinblick auf die Auswirkung einer Fehlfunktion auf das Unternehmen. Das Ergebnis dieser Aktivität, die Projektrisikoklasse, definiert den Umfang der bewertungsrelevanten Anforderungen an das jeweilige KI-System.
- Im nächsten Schritt sind die Anforderungen zu bewerten, die je nach Risikoklasse des Projekts empfohlen werden. Hierbei ist zu bewerten, ob das KI-System die bewertungsrelevanten Anforderungen erfüllt.

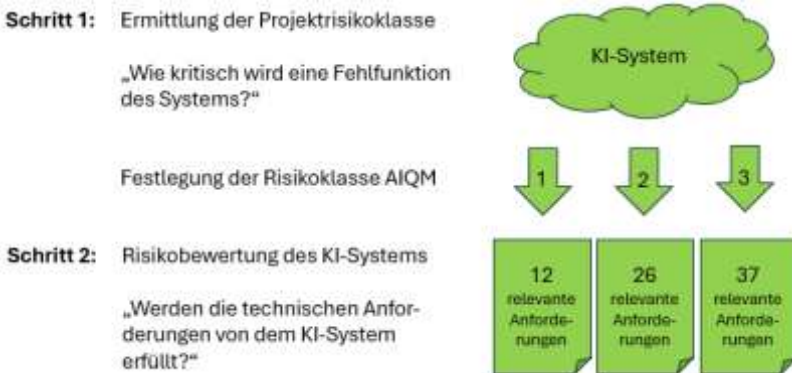


Abbildung 5-1: Skizze zur Bewertung der Risiken des KI-Systems

Ziel ist es, alle relevanten Anforderungen auf Erfüllung oder Nichterfüllung zu bewerten und diese Bewertung dokumentiert zu begründen. Sind Anforderungen in der Checkliste nicht zutreffend (n. a.), so ist der Grund für das Nichtzutreffen zu dokumentieren.

Eine negative Wertung ergibt sich durch die Nichterfüllung der entsprechenden bewertungsrelevanten Anforderung. Dies stellt nicht automatisch eine Ablehnung für den Einsatz des KI-Systems dar. Vielmehr wird jedoch eine Abschätzung der verbleibenden Risiken und eine Umsetzung entsprechender Maßnahmen (z. B. Risikoakzeptanz, Umsetzung von Risikominierungsmaßnahmen) erwartet.

Das eigentliche Freigabeverfahren auf Basis dieser bewerteten Anforderungen sowie gegebenenfalls notwendiger Zusatzmaßnahmen ist von dem anwendenden Unternehmen selbst festzulegen. Es wird empfohlen, je nach Projektrisikoklasse die relevanten Stakeholder zu ermitteln und in den Genehmigungsumlauf des KI-Systems einzubeziehen.

Auch wenn aufgrund einer niedrigen Projektrisikoklasse lediglich ein Teil der bewertungsrelevanten Anforderungen zutreffend und somit zu bewerten wäre, wird empfohlen, auch alle anderen Anforderungen für höhere Projektrisikoklassen mit zu betrachten.

Hinweis: Bewusst wurden die bewertungsrelevanten Kriterien auf eine technische Ebene zur Entwicklung von KI-Systemen zurückgeführt, die für einen bestimmten Geschäftszweck entwickelt und somit trainiert werden. Eine generelle Bewertung der Eignung von KI-Tools, d. h. Werkzeugen für eine automatisierte Entwicklung von KI-Systemen, erfolgt mit dieser Methode nicht.

5.1 Schritt 1: Ermittlung der Projektrisikoklasse

Die Projektrisikoklasse AIQM bestimmt sich aus verschiedenen Risikokriterien mit Bewertungsmaßstäben, die in sieben Risikogruppen unterteilt sind. Die hierbei ermittelte Projektrisikoklasse spiegelt die Komplexität sowie das unternehmerische Risiko für die Nutzung des KI-Systems wider.

In jeder der Risikogruppen ist zu werten, ob ein hohes Risiko der Klasse AIQM-3, ein mittleres der Klasse AIQM-2 oder ein niedriges der Klasse AIQM-1 vorliegt. Die Projektrisikobewertung ergibt sich aus der höchsten Risikostufe aus allen sieben Risikogruppen.

Beispiel: Ist eine Risikogruppe mit AIQM-3, zwei weitere mit AIQM-2 und alle anderen mit AIQM-1 bewertet, so ist die Projektrisikoklasse des KI-Systems mit AIQM-3 zu werten.

Nr.	Risiko- gruppe	Risikokriterium	Bewertung	Risiko- klasse
1.	KI-Regu- latorik	<p>EU: Hochrisiko-KI-System gemäß EU AI Act, aber mit etablierten Kontrollmechanismen oder spezifische Homologationsanforderungen</p> <p>Andere Märkte: Hohe regulatorische Anforderungen der jeweiligen Zielmärkte</p>	<p>EU: KI-System ist gemäß EU AI Act ein Hochrisiko-KI-System</p>	AIQM-3
		<p>Moderate regulatorische Anforderungen und/oder GPAI</p>	<p>EU: KI-System unterliegt Transparenzpflichten gemäß EU AI Act Artikel 50</p>	AIQM-2
		<p>Geringe regulatorische Anforderungen</p>	<p>EU: KI-System ist gemäß EU AI Act kein Hochrisiko-KI-System und unterliegt keinen besonderen Auflagen (siehe Artikel 95)</p>	AIQM-1
2.	Daten- schutzrecht	<p>Personenbezogene und sensible Daten vorhanden oder</p>	<p>Hohes Risiko besteht bei der Verwendung</p>	AIQM-3

		Datenrechte Dritter könnten verletzt werden	personenbezogener Daten und bei Daten mit hohem Schutzbedarf, da Informationen aus dem Unternehmen abfließen oder unbeabsichtigte Ergebnisse erzielt werden, oder bei Daten Dritter, deren Datenrechte verletzt werden könnten	
		Pseudonymisierte Daten	Geringes Risiko durch technische und organisatorische Maßnahmen	AIQM-2
		Keine personenbezogenen oder sensiblen Daten verarbeitet	Es werden ausschließlich anonymisierte oder synthetische Daten genutzt	AIQM-1
3.	Bias & Fairness	Fairness wird verletzt oder ein unerwünschter Bias liegt vor	Bias wird höchstwahrscheinlich auftreten	AIQM-3
		Fairness kann verletzt werden oder ein unerwünschter Bias könnte vorliegen	Bias kann auftreten, aber es gibt Prozesse zur Erkennung und Korrektur	AIQM-2
		Keine relevanten Hinweise auf Bias oder	Trainingsdaten und Modell besitzen	AIQM-1

		Fairnessverletzung vorhanden	einen akzeptablen Bias	
4.	Transparenz	Weitgehend Black Box, kaum Erklärbarkeit	Wichtige Entscheidungen sind überwiegend nicht nachvollziehbar	AIQM-3
		Teilweise erklärbar, Black-Box-Anteile vorhanden	Einige Entscheidungen sind nicht nachvollziehbar	AIQM-2
		Modell ist weitgehend erklärbar und dokumentiert	Alle Entscheidungen sind grundsätzlich nachvollziehbar	AIQM-1
5.	Finanzielles Risiko	Hohe wirtschaftliche Verluste durch fehlerhafte Ergebnisse	Fehlerhafte Ergebnisse oder der Ausfall von Geschäftsprozessen (z. B. Absatz, Wartung) führen zu erheblichen Kosten oder Fehlallokationen	AIQM-3
		Moderate wirtschaftliche Auswirkungen bei Fehlfunktion	Fehler führen zu erhöhtem Aufwand, aber keine kritischen Verluste	AIQM-2
		Geringe wirtschaftliche Auswirkungen	Fehlerhafte Entscheidungen führen zu Mehraufwand, aber sind durch Prozesse	AIQM-1

			abgedeckt oder wirtschaftlich vertretbar	
6.	Reputationsrisiko	Öffentlichkeitswirksamer Vorfall mit Vertrauensverlust	KI-System verursacht Skandal oder Diskriminierung, was zu massiven Imageschäden führt	AIQM-3
		Moderate Reputationsrisiken	Bei Vorfällen werden betroffene Stakeholder einbezogen, begrenzter Reputationsverlust mit der Möglichkeit, ihn durch offene Kommunikation und Problemlösung einzugrenzen	AIQM-2
		Kein relevantes oder geringes Reputationsrisiko	KI-Anwendung wird nur für interne Zwecke eingesetzt. Es ist keine Auswirkung außerhalb des Unternehmens zu erwarten	AIQM-1
7.	Produkt-eigenschaften	Bewertung der Funktionalen Sicherheit (FuSi, ISO 26262) oder der Gebrauchssicherheit (ISO 21448) des	Hoch = FuSi ASIL D/C, ISO 21448 Typ 3/4 oder Besonderes Merkmal oder Komponenten mit	AIQM-3

	betreffenden Produkts sowie der Besonderen Merkmale (IATF 16949) und der Cyber-Security (CS)	aktiver Kommunikation (Einfluss CS-Requirements)	
		Mittel = FuSi ASIL B/A, ISO 21448 Typ 2 und kein Besonderes Merkmal und geringe CS-Relevanz	AIQM-2
		Niedrig = FuSi ASIL QM, ISO 21448 Typ 1 und kein Besonderes Merkmal und keine CS-Relevanz	AIQM-1

5.2 Schritt 2: Risikobewertung des KI-Systems

Die relevanten Anforderungen sind Prozessphasen bei der Entwicklung bis zur Inbetriebnahme eines KI-Systems zugeordnet. Dies dient hier der besseren Übersicht und Lesbarkeit. Die Phasen bzw. bewertungsrelevanten Anforderungen können ebenso Projekt- und/oder Entwicklungsphasen der im anwendenden Unternehmen festgelegten KI-Entwicklungsprozesse zugeordnet werden.

5.2.1 Übersicht der Leitfragen zu den bewertungsrelevanten Anforderungen in Prozessphasen

1	Anwendungsgebiet
1.1	Sind das Anwendungsgebiet und ein realistisches Ziel für die KI-Anwendung definiert?
1.2	Sind besondere Anforderungen an die Erklärbarkeit des erwarteten KI-Verhaltens geklärt?
1.3	Sind alle relevanten Rollen definiert und notwendige Kompetenzen sichergestellt, um einen reibungslosen Projekt-/Aktivitätsstart zu ermöglichen?
1.4	Ist die Anwendung des KI-Systems mit dem Kunden abgestimmt?
1.5	Sind relevante regulatorische Anforderungen, interne und externe Standards und vertragliche Vereinbarungen bekannt und werden diese eingehalten?

2	Datenverständnis
2.1	Ist festgelegt, welche Daten benötigt werden, um ein KI-System für den gewünschten Geschäftszweck zu entwickeln?

3	Datenerfassung
3.1	Ist die Datenaufzeichnung ausreichend dokumentiert, um reproduziert zu werden?
3.2	Erfüllen die Daten die Anforderung der Aufgabe?
3.3	Ist die Datenversionierung sichergestellt?

3.4	Sind relevante Stakeholder in die Datenaufzeichnung eingebunden?
-----	--

4	Datenaufbereitung
---	-------------------

4.1	Sind die bereitgestellten Daten vollständig, korrekt und konsistent, um eine vertrauenswürdige Modellbildung, Test und Validierung zu ermöglichen?
-----	--

4.2	Wurde sichergestellt, dass die Daten fair und repräsentativ sind sowie relevanter Bias auf ein akzeptables Maß reduziert sind?
-----	--

4.3	Sind die Schritte der Datenaufbereitung reproduzierbar, nachvollziehbar und technisch korrekt umgesetzt?
-----	--

4.4	Sind für die Datenverarbeitung sicherheitskritische Aspekte berücksichtigt?
-----	---

5	KI-Modellierung
---	-----------------

5.1	Ist das Training deterministisch konfiguriert und reproduzierbar?
-----	---

5.2	Ist sichergestellt, dass eine Modellkalibrierung vorliegt, die die Unsicherheiten korrekt widerspiegelt?
-----	--

5.3	Ist das Training vollständig dokumentiert, inkl. Verfahren, Leistungsgrenzen und Einschränkungen?
-----	---

6	Evaluierung/Test
---	------------------

6.1	Ist sichergestellt, dass das Modellkonzept Erklärbarkeit und Interpretierbarkeit für relevante Stakeholder ermöglicht?
-----	--

6.2	Sind Metriken für Reproduzierbarkeit, System Fairness, Überwachung, Gesetzeskonformität und DSGVO-Konformität der KI-Tests sichergestellt?
-----	--

6.3	Sind Robustheits-, Edge-Case- und Sensitivitätstests systematisch, messbar und dokumentiert?
-----	--

6.4	Stellen Tests, KPIs und Protokolle funktionale Äquivalenz, Konsistenz und Compliance sicher?
-----	--

6.5	Sind irreführende Antworten auf ein akzeptables Maß reduziert?
7 Einsatzvorbereitung	
7.1	Sind IT-Sicherheits- und Cybersicherheits-Belange berücksichtigt?
7.2	Ist ein Betriebskonzept inklusive Eskalations- und Notfallkonzepte ausgearbeitet und abgestimmt?
7.3	Ist ein sicherer Ausrollprozess geplant?
8 Applikationsintegration	
8.1	Ist ein Zugriffs- und Identitätsmanagementplan für das KI-System, die KI-Anwendung sowie relevante Datensätze festgelegt und umgesetzt?
8.2	Ist die Dokumentation für Daten, Algorithmus, Hard- und Software erstellt und ist die Versionierung sichergestellt?
9 Leistungsverifikation	
9.1	Wird der geplante Geschäftszweck von dem KI-System erfüllt?
9.2	Wie robust reagiert das KI-System unter extremen Belastungen? Werden Systemgrenzen definiert und überwacht?
9.3	Ist die Leistungsfähigkeit des KI-Systems unter normalen Bedingungen und Berücksichtigung verschiedener Testmethoden überprüft?
9.4	Sind Validierungsergebnisse bei jeder KI-System-Version vollständig, nachvollziehbar und audittierbar dokumentiert?
10 Produktivstart	
10.1	Ist die Verantwortlichkeit für das KI-System an den Kunden übergeben worden?

11	Kontinuierliche Verbesserung
11.1	Ist ein Data-Monitoring aufgesetzt?
11.2	Ist ein Model-Monitoring aufgesetzt?
11.3	Ist ein Performance-Monitoring für die Laufzeitumgebung im Produktivbetrieb aufgesetzt?
11.4	Ist ein Änderungsmanagementsystem aufgesetzt?
11.5	Ist eine Überwachungs- und Retraining-/Update-Prozedur vorhanden?

5.2.2 Bewertungsrelevante Anforderungen an KI-Systeme

Prozessphase 1 Anwendungsgebiet				
1.1	Sind das Anwendungsgebiet und ein realistisches Ziel für die KI-Anwendung definiert?	Anwendbar für Risikoklasse		
		AIQM-1	AIQM-2	AIQM-3
		Ja	Ja	Ja
Bewertungsrelevante Anforderungen		Beispiele zur Umsetzung		
<p>Das Anwendungsgebiet und der Einsatz der KI-Anwendung sind festgelegt und mit der Strategie (Unternehmen, IT ...) abgeglichen. Messbare Ziele und Anwendungsfälle für das KI-System wurden definiert:</p> <ul style="list-style-type: none"> - mathematisch oder - anhand von Datenbeispielen oder - semantisch/verbal (z. B. Szenarienkatalog, gezielt eingeschränkter operativer Gestaltungsbereich) <p>Die Skalierbarkeit muss nach Möglichkeit definiert werden (d. h., die Anwendbarkeit auf verschiedene Linien, Anlagen, IT-Systeme). Robustheitsanforderungen sowie IT-Sicherheitsanforderungen sind zu beschreiben, sofern erforderlich.</p> <p>Wo aufgrund interner Vorgaben erforderlich, ist eine Return-of-Investment (ROI)-Analyse oder eine Benchmarking-Analyse zu erstellen, ebenso wie erforderliche Kostenverrechnungsmodelle (z. B. Lizenzen, Tokens).</p> <p>Sofern Anforderungen zu der End-of-Life-Planung des KI-Systems bestehen (z. B. Langzeitarchivierung aller Daten und Modelle), sind diese in das Anwendungsgebiet mit zu integrieren.</p> <p>Generative KI: Generative KI erfordert klare Vorgaben für die Inhaltsgenerierung. Die Anwendungsfallbeschränkungen sind entscheidend. Außerdem muss definiert werden, welche risikoreichen Inhalte (z. B. Gewalt, Hass) herausgefiltert werden sollen.</p>		<ul style="list-style-type: none"> • Projektbeschreibung • Contracting • Spezifikation • Zielerreichungsbogen 		

Prozessphase 1 Anwendungsgebiet				
1.2	Sind besondere Anforderungen an die Erklärbarkeit des erwarteten KI-Verhaltens geklärt?	Anwendbar für Risikoklasse		
		AIQM-1	AIQM-2	AIQM-3
		Nein	Ja	Ja
Bewertungsrelevante Anforderungen		Beispiele zur Umsetzung		
<p>Die gewünschte Funktion eines erwarteten KI-Verhaltens (Erklärbarkeit) ist beschrieben, dokumentiert und nachvollziehbar. Die Beziehung zwischen Inputs und Outputs kann in einem der Problemstellung angemessenen Detaillierungsgrad beschrieben werden. Die Grenzen des Systems und vorhersehbarer Fehlgebrauch sind zu berücksichtigen. Die Einflussnahme der Ergebnisse des KI-Systems auf weitere Prozesse oder Entscheidungen ist mit einzubeziehen. Eine gegebenenfalls notwendige Verifikation der KI-Systemergebnisse ist zu planen (z. B. Human-in-the-loop).</p> <p>Hinweis: Gemäß Art. 50 des EU AI Act ist Transparenz hinsichtlich Erklärbarkeit und Nachvollziehbarkeit erforderlich.</p> <p>Generative KI: Erklärbarkeit ist beim Einsatz generativer KI eine Herausforderung, aber unerlässlich für das Verständnis der Ergebnisse und die Vertrauensbildung.</p>		<ul style="list-style-type: none"> • Projektbeschreibung • Contracting • Spezifikation • Zielerreichungsbogen 		

Prozessphase 1 Anwendungsgebiet				
1.3	Sind alle relevanten Rollen definiert und notwendige Kompetenzen sichergestellt, um einen reibungslosen Projekt-/Aktivitätsstart zu ermöglichen?	Anwendbar für Risikoklasse		
		AIQM-1	AIQM-2	AIQM-3
		Ja	Ja	Ja
Bewertungsrelevante Anforderungen		Beispiele zur Umsetzung		
<p>Ein:e Projektmanager:in/Hauptverantwortliche:r und die entsprechenden Expert:innen für Software, KI, Systeme und Domänen sind Teil des Projekts / der Aktivität und werden mit den erforderlichen Kapazitäten dafür eingesetzt. Die erforderlichen Kompetenzen der Beteiligten sind sicherzustellen (u. a. Anforderungen zu AI-Literacy gemäß EU AI Act).</p> <p>Zusätzlich benötigte Rollen können die oder der Open-Source-Officer, die Rechtsabteilung, die Cyber-Sicherheitsabteilung, die oder der Datenschutzbeauftragte oder ein Product-Compliance-Mitglied sein.</p> <p>Anmerkung: Weitere Rollenbeschreibungen und Kompetenzen finden sich in Kapitel 4 dieses Bandes.</p>		<ul style="list-style-type: none"> • Projektbeschreibung • Contracting • Organigramm • Schulungsmatrizen 		

Prozessphase 1 Anwendungsgebiet				
1.4	Ist die Anwendung des KI-Systems mit dem Kunden abgestimmt?	Anwendbar für Risikoklasse		
		AIQM-1	AIQM-2	AIQM-3
		Ja	Ja	Ja
Bewertungsrelevante Anforderungen		Beispiele zur Umsetzung		
<p>Der Einsatz des KI-Systems wird frühzeitig mit dem Kunden/Stakeholder geklärt und abgestimmt, sofern benötigt.</p> <p>Die Überwachung im Nutzungszeitraum und gegebenenfalls das Aktualisierungskonzept des KI-Systems ist mit dem Kunden/Stakeholder abgestimmt.</p> <p>EU AI Act Art. 50: Für alle KI-Systeme, die zur direkten Interaktion mit Personen bestimmt sind, besteht Informationspflicht. Bei Hochrisiko-KI-Systemen besteht in jedem Fall Informationspflicht.</p>		<ul style="list-style-type: none"> • Projektbeschreibung • Contracting • Vertrag • Spezifikation 		

Prozessphase 1 Anwendungsgebiet				
1.5	Sind relevante regulatorische Anforderungen, interne und externe Standards und vertragliche Vereinbarungen bekannt und werden diese eingehalten?	Anwendbar für Risikoklasse		
		AIQM-1	AIQM-2	AIQM-3
		Ja	Ja	Ja
Bewertungsrelevante Anforderungen		Beispiele zur Umsetzung		
<p>Daten und Algorithmen müssen Eigentum des Unternehmens sein, vom Unternehmen lizenziert oder für die kommerzielle Nutzung frei zugänglich sein.</p> <p>Es muss geklärt sein, wer im Rahmen von Verträgen (z. B. mit Lieferanten) für das Verhalten der KI haftet.</p> <p>Zu klären ist auch, welche zusätzlichen Anforderungen, z. B. Sicherheitsstandards und Produkthaftung, zu berücksichtigen sind.</p> <p>Mindestens zu prüfen sind:</p> <ul style="list-style-type: none"> • die Anforderungen des EU AI Act und/oder anderer regulatorischer Vorschriften in den Zielmärkten • Regelungen zur Exportkontrolle • DSGVO-/Regionale Datenschutzanforderungen. Der Schutz personenbezogener Daten ist in allen Phasen sichergestellt. Die Umsetzung von Löschkonzepten und -rechten ist berücksichtigt (Recht auf Löschung, Aufbewahrungsrichtlinien) • Anforderungen an Datenrecht, z. B. EU Data Act oder Chinese Data Security Law • Software- und Datensatzlizenzierung • Open-Source-Softwaremanagement 		<ul style="list-style-type: none"> • Projektbeschreibung • Contracting • Vertrag • Spezifikation 		

-
- Patentrecherche: Es werden keine Patente verletzt
 - Bei der Zusammenarbeit mit Lieferanten und Kunden muss die Haftung geklärt und Vertragsbestandteil sein
 - Berücksichtigung vorhandener und eingehender Daten hinsichtlich ihrer rechtlichen Verwertbarkeit, insbesondere bezüglich Urheberrechten und DSGVO

Wo anwendbar, sind ethische Grundsätze und Fairness-Kriterien zu berücksichtigen.

Prozessphase 2 Datenverständnis				
2.1	Ist festgelegt, welche Daten benötigt werden, um ein KI-System für den gewünschten Geschäftszweck zu entwickeln?	Anwendbar für Risikoklasse		
		AIQM-1	AIQM-2	AIQM-3
		Ja	Ja	Ja
Bewertungsrelevante Anforderungen		Beispiele zur Umsetzung		
<p>Es ist festzulegen, welche Daten für das Training, den Test, die Validierung und im Betrieb des KI-Systems benötigt werden. Insbesondere ist auf folgende Eigenschaften zu achten:</p> <ul style="list-style-type: none"> • Art, Umfang und Variationen der Daten (z. B. Aktualität) • Datenformate • Erwartete Datentransferraten • Datenschemata • Attribuierung/Indizierung • Metadaten • Herkunft der Daten: Datenquellen (zur Nutzung vorhandener Daten oder zur Erhebung neuer Daten) sind technisch zugänglich und dokumentiert. Bei Drittquellen sind Lizenzbedingungen und Nutzungsrechte dokumentiert • Rohdaten werden unverändert und versioniert in einem unveränderlichen Storage gespeichert, um vollständige Rückverfolgbarkeit sicherzustellen • Anforderungen an die Archivierung festlegen • Datenkompression vs. Verlust von Daten 		<ul style="list-style-type: none"> • Spezifikation • Übersicht über die Datenformate und Inhalte 		

Prozessphase 3 Datenerfassung				
3.1	Ist die Datenaufzeichnung ausreichend dokumentiert, um reproduziert zu werden?	Anwendbar für Risikoklasse		
		AIQM-1	AIQM-2	AIQM-3
		Nein	Ja	Ja
Bewertungsrelevante Anforderungen		Beispiele zur Umsetzung		
<p>Die Methodik und Systematik zur Erhebung der Daten ist vollständig beschrieben und nachvollziehbar dokumentiert.</p> <p>Die Verfahren zur Datenaufzeichnung sowie zur Annotation der erhobenen Daten sind spezifiziert.</p> <p>Die Reproduzierbarkeit der Datenaufnahme ist gewährleistet.</p>		<ul style="list-style-type: none"> • Spezifikation • Arbeitsanweisung • Übersicht über die Datenformate, Quellen und Inhalte 		

Prozessphase 3 Datenerfassung				
3.2	Erfüllen die Daten die Anforderung der Aufgabe?	Anwendbar für Risikoklasse		
		AIQM-1	AIQM-2	AIQM-3
		Ja	Ja	Ja
Bewertungsrelevante Anforderungen		Beispiele zur Umsetzung		
<p>Die erhobenen Daten sind geeignet, den in der Prozessphase „Anwendungsgebiet“ definierten Zweck des KI-Systems zu erfüllen, insbesondere für Training, Test/Verifizierung und Performance Validation.</p> <p>Falls zusätzliche Daten benötigt werden: Die Generierung synthetischer Daten ist unter bestimmten Umständen möglich:</p> <ul style="list-style-type: none"> • Simulieren der Daten z. B. defekter Teile mit Hilfe verfügbarer Daten und Fachkenntnissen, z. B. physikalischer Modelle. Fachexpert:innen dokumentieren die Gründe für die Verwendung synthetischer Daten und 		<ul style="list-style-type: none"> • Übersicht über die Datenformate, Quellen und Inhalte • Risikobewertung • Prüfergebnis der Daten 		

<p>die Unterschiede zu realen Daten (siehe hierzu auch VDA 5.3 Kapitel 6).</p> <ul style="list-style-type: none"> • Synthetische Daten dürfen nicht von dem KI-System erzeugt werden, das trainiert bzw. getestet werden soll (Gefahr von Bias). • Die synthetischen Daten müssen vor Verwendung geprüft werden. 	
--	--

Prozessphase 3 Datenerfassung				
3.3	Wurde die Datenversionierung sichergestellt?	Anwendbar für Risikoklasse		
		AIQM-1	AIQM-2	AIQM-3
		Nein	Nein	Ja
Bewertungsrelevante Anforderungen		Beispiele zur Umsetzung		
Jegliche Änderungen an den Datenbeständen oder den Rohdaten werden durch ein etabliertes Versionierungssystem transparent nachvollzogen und dokumentiert.		<ul style="list-style-type: none"> • Arbeitsanweisung • Datenbank oder Datenfelder für Versionsdaten 		

Prozessphase 3 Datenerfassung				
3.4	Wurden relevante Stakeholder in die Datenaufzeichnung eingebunden?	Anwendbar für Risikoklasse		
		AIQM-1	AIQM-2	AIQM-3
		Nein	Ja	Ja
Bewertungsrelevante Anforderungen		Beispiele zur Umsetzung		
Die relevanten Stakeholder (z. B. Datenschutzbeauftragte, Kunden, interne Fachbereiche) wurden über die Datenaufnahme informiert und einbezogen.		<ul style="list-style-type: none"> • Contracting • Vertrag • Organigramm • Nachweis zur Information, z. B. E-Mail 		

Prozessphase 4 Datenaufbereitung				
4.1	Sind die bereitgestellten Daten vollständig, korrekt und konsistent, um eine vertrauenswürdige Modellbildung, Test und Validierung zu ermöglichen?	Anwendbar für Risikoklasse		
		AIQM-1	AIQM-2	AIQM-3
		Ja	Ja	Ja
Bewertungsrelevante Anforderungen		Beispiele zur Umsetzung		
<p>Nach Erfassung der geplanten Daten erfolgt deren Überprüfung nach folgenden Kriterien:</p> <ul style="list-style-type: none"> • Vollständigkeit (ausreichende Datenmenge vorhanden, Lücken sind identifiziert) • Eindeutigkeit (jeder Datenpunkt / jeder Datensatz ist identifizierbar und zuordenbar, Duplikate sind erkannt) • Konsistenz (Wertebereiche, Formate) • Qualitätsgrenzen für Daten definiert und dokumentiert 		<ul style="list-style-type: none"> • Prüfprotokoll mit Ergebnis • Festlegung der Qualitätsgrenzen der Daten • Automatisierte Prüfungen bei Datenerhebung oder Weiterverarbeitung 		

Prozessphase 4 Datenaufbereitung				
4.2	Wurde sichergestellt, dass die Daten fair und repräsentativ sind sowie relevanter Bias auf ein akzeptables Maß reduziert sind?	Anwendbar für Risikoklasse		
		AIQM-1	AIQM-2	AIQM-3
		Nein	Nein	Ja
Bewertungsrelevante Anforderungen		Beispiele zur Umsetzung		
<p>Eine Bias-Analyse wurde durchgeführt und dokumentiert.</p> <p>Datensets repräsentieren die Zielpopulation (Verteilungen dokumentiert).</p> <p>Maßnahmen zur Bias-Reduktion sind definiert und umgesetzt.</p> <p>Fairness-Metriken sind definiert und erfüllt.</p>		<ul style="list-style-type: none"> • Analyseprotokoll mit Ergebnis • Dokumentation zu Maßnahmen mit Verantwortlichkeiten, Termin und Status • Wirksamkeit eingeführter Maßnahmen • Festlegung der Fairnessmetriken • Automatisierte Prüfungen bei Datenerhebung oder Weiterverarbeitung 		

Prozessphase 4 Datenaufbereitung				
4.3	Sind die Schritte der Datenaufbereitung reproduzierbar, nachvollziehbar und technisch korrekt umgesetzt?	Anwendbar für Risikoklasse		
		AIQM-1	AIQM-2	AIQM-3
		Nein	Ja	Ja
Bewertungsrelevante Anforderungen		Beispiele zur Umsetzung		
<p>Alle Änderungen am Datensatz sind nachvollziehbar protokolliert.</p> <p>Ausreißer-Analysen sind vorhanden und bewertet.</p> <p>Feature-Auswahl und -Generierung sind begründet und dokumentiert.</p>		<ul style="list-style-type: none"> • Dokumentation von Änderungen • Analyseergebnis von Ausreißern • Dokumentation zu Feature-Auswahl und Generierung 		

Prozessphase 4 Datenaufbereitung				
4.4	Sind für die Datenverarbeitung sicherheitskritische Aspekte berücksichtigt?	Anwendbar für Risikoklasse		
		AIQM-1	AIQM-2	AIQM-3
		Nein	Nein	Ja
Bewertungsrelevante Anforderungen		Beispiele zur Umsetzung		
<p>Sicherheitskritische Daten mit Bezug auf besondere Merkmale (siehe hierzu IATF16949) sind mit speziellen Prüfmaßnahmen abgesichert.</p>		<ul style="list-style-type: none"> • Liste der besonderen Merkmale gem. IATF16949 • Dokumentation der Maßnahmen mit Verantwortlichkeit, Termin und Status • Wirksamkeit eingeführter Maßnahmen 		

Prozessphase 5 KI-Modellierung				
5.1	Ist das Training deterministisch konfiguriert und reproduzierbar?	Anwendbar für Risikoklasse		
		AIQM-1	AIQM-2	AIQM-3
		Ja	Ja	Ja
Bewertungsrelevante Anforderungen		Beispiele zur Umsetzung		
<p>Es sind deterministische Abläufe implementiert, z. B. Seed-Management. Trainingsprozesse sind auf unterschiedlichen Entwicklungsumgebungen reproduzierbar validiert. Reproduzierbarkeit und Nachvollziehbarkeit sind sichergestellt (z. B. Abspeichern der Hyperparameter, toolgestützte Dokumentation).</p> <p>Generative KI: Diese Anforderungen sind bei generativer KI nicht umsetzbar. In diesem Fall sind zusätzliche Maßnahmen, z. B. zur Überwachung und Test, vorzusehen.</p>		<ul style="list-style-type: none"> • Ergebnis von Tests • Dokumentation von Hyperparametern 		

Prozessphase 5 KI-Modellierung				
5.2	Wurde sichergestellt, dass eine Modellkalibrierung vorliegt, die die Unsicherheiten korrekt widerspiegelt?	Anwendbar für Risikoklasse		
		AIQM-1	AIQM-2	AIQM-3
		Nein	Ja	Ja
Bewertungsrelevante Anforderungen		Beispiele zur Umsetzung		
<p>Geeignete Kalibriermethoden sind angewendet und dokumentiert.</p> <p>Evaluierung der Vorhersageunsicherheiten ist durchgeführt.</p> <p>Kalibrierung ist regelmäßig überprüft und aktualisiert.</p> <p>Generative KI: Diese Anforderungen sind bei generativer KI nicht umsetzbar. In diesem Fall sind zusätzliche Maßnahmen, z. B. zur Überwachung und Test, vorzusehen.</p>		<ul style="list-style-type: none"> • Kalibriermethoden wie z. B. Platt Scaling, Isotonic Regression, Temperature or Matrix Scaling, Label Smoothing • Bewertung der Vorhersageunsicherheit • Dokumentation der Kalibrierung 		

Prozessphase 5 KI-Modellierung				
5.3	Ist das Training vollständig dokumentiert, inkl. Verfahren, Leistungsgrenzen und Einschränkungen?	Anwendbar für Risikoklasse		
		AIQM-1	AIQM-2	AIQM-3
		Nein	Nein	Ja
Bewertungsrelevante Anforderungen		Beispiele zur Umsetzung		
<p>Geeignete Trainings- und Testverfahren sind vorhanden und deren Ergebnisse dokumentiert.</p> <p>Performance-Metriken und deren Grenzen sind dokumentiert.</p> <p>Bekannt Limitierungen und Risiken sind explizit beschrieben.</p> <p>Generative KI: Diese Anforderungen sind bei generativer KI nicht umsetzbar. In diesem Fall sind zusätzliche Maßnahmen, z. B. zur Überwachung und Test, vorzusehen.</p>		<ul style="list-style-type: none"> • Verfahrensanweisungen • Performance-Metriken • Risikobewertung • Beschreibung des Anwendungsgebiets • Spezifikation 		

Prozessphase 6 Evaluierung/Test				
6.1	Ist sichergestellt, dass das Modellkonzept Erklärbarkeit und Interpretierbarkeit für relevante Stakeholder ermöglicht?	Anwendbar für Risikoklasse		
		AIQM-1	AIQM-2	AIQM-3
		Nein	Nein	Ja
Bewertungsrelevante Anforderungen		Beispiele zur Umsetzung		
<p>Anforderungen an Erklärbarkeit sind geprüft und erfüllt.</p> <p>Interpretierbarkeitmethoden (z. B. SHAP, LIME) sind spezifiziert und dokumentiert.</p> <p>Kritische Features und deren Einfluss sind nachvollziehbar dargestellt.</p>		<ul style="list-style-type: none"> • Liste der Anforderungen an Erklärbarkeit (z. B. Safety, EU AI Act) • Liste kritischer Features 		

Prozessphase 6 Evaluierung/Test				
6.2	Sind Metriken für Reproduzierbarkeit, System Fairness, Überwachung, Gesetzeskonformität und DSGVO-Konformität der KI-Tests sichergestellt?	Anwendbar für Risikoklasse		
		AIQM-1	AIQM-2	AIQM-3
		Nein	Ja	Ja
Bewertungsrelevante Anforderungen		Beispiele zur Umsetzung		
<p>Klare, quantifizierbare Leistungsmetriken für überwachte Verfahren und unüberwachte Verfahren sind festgelegt und auf Benchmark-Datensätzen validiert.</p> <p>Detaillierte Bias-Analyse des Modells erfolgt an klar definierten Gruppen, deren Datenabdeckung und Stichprobengrößen geprüft werden.</p> <p>Fairness-Metriken wurden erfasst und ausgewertet.</p> <p>Der Auswertungsprozess erfolgt standardisiert mit Visualisierung der Gruppen-Disparitäten und Peer-Review.</p>		<ul style="list-style-type: none"> • Metriken für überwachte Verfahren wie z. B. F1, Precision, Recall, Accuracy • Metriken für unüberwachte Verfahren wie z. B. Silhouette, Davies-Bouldin • Ergebnis der Bias-Analyse des Modells • Ergebnis der Fairness-Metriken wie z. B. Fehlerraten-Differenzen, Disparate-Impact/Disparitäts-Index, Demographic-Parity-Gap, Equalized-Odds (TPR-FPR-Gaps) und Calibration-Gap. • Ergebnis der Auswertung • Teilnehmer des Peer-Reviews 		

Prozessphase 6 Evaluierung/Test				
6.3	Sind Robustheits-, Edge-Case- und Sensitivitätstests systematisch, messbar und dokumentiert?	Anwendbar für Risikoklasse		
		AIQM-1	AIQM-2	AIQM-3
		Nein	Ja	Ja
Bewertungsrelevante Anforderungen		Beispiele zur Umsetzung		
<p>Standardisierte Sensitivitätsanalysen sind mit messbaren Kennzahlen/KPI systematisch eingeführt.</p> <p>Edge-Case-Tests unter extremen Eingaben sind durchgeführt und ausgewertet.</p>		<ul style="list-style-type: none"> • Festlegung von Kennzahlen und Akzeptanzkriterien • Protokolle von Tests und Analysen 		

Prozessphase 6 Evaluierung/Test				
6.4	Stellen Tests, Kennzahlen/KPI und Protokolle die funktionale Äquivalenz, Konsistenz und Compliance sicher?	Anwendbar für Risikoklasse		
		AIQM-1	AIQM-2	AIQM-3
		Nein	Nein	Ja
Bewertungsrelevante Anforderungen		Beispiele zur Umsetzung		
<p>Tests belegen die funktionale Äquivalenz des KI-Modells anhand klar definierter Input-/Output-Sätze, sofern verfügbar bzw. machbar. Konsistenz-Kennzahlen/KPI werden kontinuierlich überwacht: Abweichung, Fehlerquote, Antwortzeiten liegen innerhalb festgelegter Toleranzen.</p>		<ul style="list-style-type: none"> • Festlegung von Kennzahlen und Akzeptanzkriterien • Kennzahlverfolgung • Protokolle von Tests und Analysen 		

Prozessphase 6 Evaluierung/Test				
6.5	Sind irreführende Antworten in Systemen mit generativer KI auf ein akzeptables Maß reduziert?	Anwendbar für Risikoklasse		
		AIQM-1	AIQM-2	AIQM-3
		Nein	Nein	Ja
Bewertungsrelevante Anforderungen		Beispiele zur Umsetzung		
Für generative KI sind spezielle Tests zu Kreativität, Konsistenz und Plausibilität (Benchmarking) integriert. Es erfolgt eine zusätzliche Prüfung erzeugter Inhalte auf stereotype oder unbeabsichtigte Vorurteile (Bias).		<ul style="list-style-type: none"> • Festlegung von Kennzahlen und Akzeptanzkriterien • Kennzahlverfolgung • Protokolle von Tests und Analysen 		

Prozessphase 7 Einsatzvorbereitung				
7.1	Sind IT-Sicherheits- und Cybersicherheits-Belange berücksichtigt?	Anwendbar für Risikoklasse		
		AIQM-1	AIQM-2	AIQM-3
		Ja	Ja	Ja
Bewertungsrelevante Anforderungen		Beispiele zur Umsetzung		
<p>Sofern notwendig und relevant, sind Maßnahmen vorhanden, um die Integrität, Robustheit und allgemeine Cybersicherheit des KI-Systems bzw. der KI-Anwendung gegen potenzielle Angriffe während des gesamten Lebenszyklus zu gewährleisten.</p> <p>Gegebenenfalls sind Cybersicherheitsbeauftragte einbezogen, falls in dem Anwendungsfall ein Risiko für KI-Sicherheit besteht.</p> <p>Mögliche Risiken können sein:</p> <ul style="list-style-type: none"> • Eingabeaufforderungsinjektion – Direkte Injektionen überschreiben Systemeingabeaufforderungen, während indirekte Eingaben aus externen Quellen manipulieren • Data Poisoning – Bei Data-Poisoning-Angriffen werden gezielt manipulierte Daten eingeschleust, um die 		<ul style="list-style-type: none"> • Risikoanalyse zur Cybersicherheit 		

Trainingsdaten zu verunreinigen und so den Modelltrainingsprozess zu beeinträchtigen

- Umgehung – Bei Umgehungsangriffen werden kleine Störungen im KI/ML-Modell-Input gefunden, die zu erheblichen Änderungen des Outputs führen
 - Modellextraktion – Bei Modellextraktionsangriffen werden KI/ML-Algorithmen extrahiert oder kopiert. Böswillige Akteure könnten Modellextraktionsangriffe nutzen, um das Modell zu stehlen
 - Inferenz – Versuch festzustellen, ob die Informationen eines bestimmten Datensatzes, z. B. einer Person, Teil der Trainingsdaten eines trainierten ML-Modells waren oder nicht
 - Offenlegung sensibler Informationen – Die Einbeziehung sensibler Informationen in die Trainingsdaten kann zur Offenlegung von Informationen führen
 - Modell-Denial-of-Service – Angriff mit ressourcenintensivem Betrieb
 - Sicherheitslücken in der Lieferkette – Die Verwendung von Datensätzen von Drittanbietern, vortrainierten Modellen und Plugins erhöht die Sicherheitslücken
 - Übermäßige Abhängigkeit – Systeme oder Personen, die sich übermäßig und ohne Aufsicht auf LLMs verlassen, können aufgrund falscher oder unangemessener Inhalte, die von LLMs generiert werden, Fehlinformationen, Missverständnissen, rechtlichen Problemen und Sicherheitslücken ausgesetzt sein
 - Prompt Injection
-

Prozessphase 7 Einsatzvorbereitung				
7.2	Ist ein Betriebskonzept inklusive Eskalations- und Notfallkonzepten ausgearbeitet und abgestimmt?	Anwendbar für Risikoklasse		
		AIQM-1	AIQM-2	AIQM-3
		Ja	Ja	Ja
Bewertungsrelevante Anforderungen		Beispiele zur Umsetzung		
<p>Ein Betriebskonzept für die geplante Nutzung des KI-Systems ist aufzustellen.</p> <p>Dieses sollte die folgenden Punkte berücksichtigen:</p> <ul style="list-style-type: none"> • Definieren und Implementieren der Eskalations- und Notfallprozesse für Vorfälle während des Betriebs, abhängig vom Sicherheitsrisiko • Ein Notfallplan soll berücksichtigt werden, wie andere Ausfälle in Unternehmen • Support- und Wartungsplan festlegen • Definierte Verantwortliche (RASIC), definierte Vertragsgestaltung mit Lieferanten und Dienstleistern • Aktualisierungsstrategie • Definiertes Überwachungskonzept • Definierter Kanal für Feedback und Informationen zu Systemeigenschaften • Definierte Rollback-Konditionen • Implementierte Strategie für die Wartung nach der Freigabe des KI-Systems • Definition der Übergabe und der Handshakes an Dienstleister • Dokumentation des Zustands im Moment der Freigabe als Referenz für spätere Updates • Verfügbarkeit der Lizenzen für das Produkt 		<ul style="list-style-type: none"> • Vereinbartes Betriebskonzept • Wartungspläne • Notfallpläne • Verantwortlichkeiten • Befugnisse 		

Prozessphase 7 Einsatzvorbereitung				
7.3	Ist ein sicherer Ausrollprozess geplant?	Anwendbar für Risikoklasse		
		AIQM-1	AIQM-2	AIQM-3
		Nein	Ja	Ja
Bewertungsrelevante Anforderungen		Beispiele zur Umsetzung		
<p>Ein Ausrollprozess für das KI-System ist zu erstellen.</p> <p>Dieses sollte die folgenden Punkte berücksichtigen:</p> <ul style="list-style-type: none"> • IT-Infrastruktur des zukünftigen Produktivsystems (u. a. Absicherung Cloud vor Datenabfluss/Datenverlust) • Schulungskonzept 		<ul style="list-style-type: none"> • Ausrollplanung • Verantwortlichkeiten • Befugnisse • Planung IT-Infrastruktur • Schulungen 		

Prozessphase 8 Applikationsintegration				
8.1	Ist ein Zugriffs- und Identitätsmanagementplan für das KI-System, die KI-Anwendung sowie relevante Datensätze festgelegt und umgesetzt?	Anwendbar für Risikoklasse		
		AIQM-1	AIQM-2	AIQM-3
		Nein	Nein	Ja
Bewertungsrelevante Anforderungen		Beispiele zur Umsetzung		
<p>Die Zugriffsrechte und das Identitätsmanagement für das KI-System sind in einem Zugriffs- und Identitätsmanagementplan definiert und umgesetzt. Dieser berücksichtigt insbesondere:</p> <ul style="list-style-type: none"> • Rollen • KI-Anwendung • Datensätze • Dokumentation 		<ul style="list-style-type: none"> • Übersicht über Rollen und Personen • Verantwortlichkeiten • Befugnisse • Beispiele für Zugriffsrechte und Identitätsmanagement: Wer hat Zugriff auf Trainingsdaten, welche Vertraulichkeitsstufe wird berührt, wer modifiziert Hyperparameter, führt Trainings durch usw.? Insbesondere sind Anforderungen der DSGVO sicherzustellen. 		

Prozessphase 8 Applikationsintegration				
8.2	Ist die Dokumentation für Daten, Algorithmus, Hard- und Software erstellt und ist die Versionierung sichergestellt?	Anwendbar für Risikoklasse		
		AIQM-1	AIQM-2	AIQM-3
		Nein	Nein	Ja
Bewertungsrelevante Anforderungen		Beispiele zur Umsetzung		
<p>Die Dokumentation der Daten, Algorithmen, Hard- und Software, einschließlich der Versionen der verwendeten Codes und Pakete für die finale KI-Lösung im Betrieb, ist abgeschlossen. Der Umfang der erforderlichen Dokumentation hängt von der Risikobewertung und der Möglichkeit der Wiederherstellung der</p>		<p>Dokumentation des KI-Systems.</p> <p>Die wichtigsten Aspekte sind:</p> <ul style="list-style-type: none"> • Datenherkunft und Merkmale der Datensätze sollten dokumentiert werden (strukturierte 		

<p>Lösung ab, basierend auf den für die Veröffentlichung verwendeten Test- und Trainingsdatensätzen. Die Dokumentation muss vollständig sein.</p> <p>Es ist sicherzustellen, dass die aktuelle operative Version für jeden Zeitraum dokumentiert wird.</p>	<p>Datenblätter mit detaillierten Informationen zur Datenverarbeitung, zum Datenerfasser und zur Datenerhebungsmethode).</p> <ul style="list-style-type: none">• Eigenschaften des KI-Systems sollten dokumentiert werden• Architektur oder Modellgraph (Anzahl der Schichten, Parameter, Konnektivität, Input-Output-Dimensionen)• Erwartete Eingabedaten (Struktur der in das KI-System eingegebenen Daten)• Erwartete Ausgabedaten (Struktur der KI-Systemausgabe)• Parametergenauigkeit (z. B. 8/16/32 Bit)• Hardwareanforderungen• Trainingsmethode (z. B. online/offline/...)• Systemarchitektur• Informationsfluss
--	---

Prozessphase 9 Leistungsverifikation				
9.1	Wird das geplante Anwendungsgebiet von dem KI-System erfüllt?	Anwendbar für Risikoklasse		
		AIQM-1	AIQM-2	AIQM-3
		Ja	Ja	Ja
Bewertungsrelevante Anforderungen		Beispiele zur Umsetzung		
<p>Es ist zu ermitteln, ob und wie das KI-System die Funktionen und Ziele erfüllt, die für das KI-System geplant wurden (siehe hierzu auch Frage 1.1 zur Festlegung der Funktionen und Ziele).</p> <p>Diese Verifizierung erfolgt beispielsweise auf Basis geeigneter Datensätze, die unabhängig von den Test- und Verifizierungsdaten sind. Die erzielten Ergebnisse des KI-Systems werden mit den geplanten Ergebnissen gemäß dem Anwendungsgebiet verglichen und bewertet.</p>		<ul style="list-style-type: none"> • Testergebnisse • Protokolle • Bewertungen von Abweichungen 		

Prozessphase 9 Leistungsverifikation				
9.2	Wie robust reagiert das KI-System unter extremen Belastungen? Werden Systemgrenzen definiert und überwacht?	Anwendbar für Risikoklasse		
		AIQM-1	AIQM-2	AIQM-3
		Nein	Ja	Ja
Bewertungsrelevante Anforderungen		Beispiele zur Umsetzung		
<p>Belastungstests sind dokumentiert. Die folgenden Tests sind empfohlen:</p> <ul style="list-style-type: none"> • Belastung der Infrastruktur • Modelleistung unter extremen Bedingungen prüfen • Model mit Grenzfällen und Edge-Cases konfrontieren • Adversariale Beispiele, die speziell darauf ausgelegt sind, das Model zu verwirren 		<ul style="list-style-type: none"> • Testergebnisse • Bewertung der Robustheit 		

Prozessphase 9 Leistungsverifikation				
9.3	Ist die Leistungsfähigkeit des KI-Systems unter normalen Bedingungen und Berücksichtigung verschiedener Testmethoden überprüft?	Anwendbar für Risikoklasse		
		AIQM-1	AIQM-2	AIQM-3
		Nein	Ja	Ja
Bewertungsrelevante Anforderungen		Beispiele zur Umsetzung		
Testpläne sind erstellt und umgesetzt: <ul style="list-style-type: none"> • supervised testing • unsupervised testing • Vergleich mit Referenzsystemen Abweichungen wurden adressiert		<ul style="list-style-type: none"> • Testpläne mit Annahmekriterien • Testergebnisse • Maßnahmen mit Verantwortlichkeiten, Terminen und Status 		

Prozessphase 9 Leistungsverifikation				
9.4	Sind Verifizierungsergebnisse bei jeder KI-System-Version vollständig, nachvollziehbar und auditierbar dokumentiert?	Anwendbar für Risikoklasse		
		AIQM-1	AIQM-2	AIQM-3
		Nein	Nein	Ja
Bewertungsrelevante Anforderungen		Beispiele zur Umsetzung		
Es liegt eine vollständige, prüfbare Dokumentation aller Verifizierungsergebnisse, Metriken, Erklärungs- und Fairness-Aktivitäten vor.		<ul style="list-style-type: none"> • Dokumentation der Verifizierungsergebnisse z. B. in einer Model Card 		

Prozessphase 10 Produktivstart				
10.1	Ist die Verantwortlichkeit für das KI-System an den Kunden übergeben worden?	Anwendbar für Risikoklasse		
		AIQM-1	AIQM-2	AIQM-3
		Ja	Ja	Ja
Bewertungsrelevante Anforderungen		Beispiele zur Umsetzung		
<p>Informieren Sie den Kunden über die Nutzung von KI, je nach Ausrichtung und rechtlichen Anforderungen. Stellen Sie sicher, dass der Kundenübergabeprozess den Transparenzanforderungen des EU AI Act entspricht und den Kunden klare und verständliche Informationen über die Funktionalität, Einschränkungen und potenziellen Risiken des KI-Systems bietet. Dies umfasst die Festlegung abgestimmter Verfahren für den Kundensupport und die Behandlung potenzieller Probleme oder Bedenken.</p> <p>Bei Wiederverwendung des KI-Systems oder der KI-Ergebnisse müssen Anforderungen der Kunden im neuen Kontext und ihre Nutzung der KI-Ergebnisse durch die KI-Systemspezifikation abgedeckt sein.</p>		<ul style="list-style-type: none"> • Dokumentation des KI-Systems • Nachweis der Verantwortungsübergabe z. B. durch Vertrag, Übergabeprotokoll 		

Prozessphase 11 Kontinuierliche Verbesserung				
11.1	Ist ein Data-Monitoring aufgesetzt?	Anwendbar für Risikoklasse		
		AIQM-1	AIQM-2	AIQM-3
		Nein	Ja	Ja
Bewertungsrelevante Anforderungen		Beispiele zur Umsetzung		
<p>Ein Data-Monitoring Process existiert für die Ergebnisse/Outputs des KI-Systems und ist aktiv.</p> <p>Metriken zur Datenqualität, Zielwerte sowie Mechanismen zur Umsetzung von Korrekturmaßnahmen bei Abweichungen sind definiert.</p>		<ul style="list-style-type: none"> • Prozessbeschreibung • Festlegung von Kennzahlen/KPI und Zielen 		

Prozessphase 11 Kontinuierliche Verbesserung				
11.2	Ist ein Model-Monitoring aufgesetzt?	Anwendbar für Risikoklasse		
		AIQM-1	AIQM-2	AIQM-3
		Nein	Nein	Ja
Bewertungsrelevante Anforderungen		Beispiele zur Umsetzung		
<p>Ein Model-Monitoring-Prozess existiert für das KI-System und ist aktiv.</p> <p>Metriken zur Modelqualität, Zielwerte sowie Mechanismen zur Umsetzung von Korrekturmaßnahmen bei Abweichungen sind definiert.</p>		<ul style="list-style-type: none"> • Prozessbeschreibung • Festlegung von Kennzahlen/KPI und Zielen 		

Prozessphase 11 Kontinuierliche Verbesserung				
11.3	Ist ein Performance-Monitoring für die Laufzeitumgebung im Produktivbetrieb aufgesetzt?	Anwendbar für Risikoklasse		
		AIQM-1	AIQM-2	AIQM-3
		Nein	Ja	Ja
Bewertungsrelevante Anforderungen		Beispiele zur Umsetzung		
<p>Ein Performance-Monitoring-Prozess für die Laufzeitumgebung im Produktivbetrieb existiert und ist aktiv.</p> <p>Metriken zur Performance, Zielwerte sowie Mechanismen zur Umsetzung von Korrekturmaßnahmen bei Abweichungen der Laufzeitumgebung sind definiert.</p>		<ul style="list-style-type: none"> • Prozessbeschreibung • Festlegung von Kennzahlen/KPI und Zielen 		

Prozessphase 11 Kontinuierliche Verbesserung				
11.4	Ist ein Änderungsmanagementsystem aufgesetzt?	Anwendbar für Risikoklasse		
		AIQM-1	AIQM-2	AIQM-3
		Nein	Ja	Ja
Bewertungsrelevante Anforderungen		Beispiele zur Umsetzung		
<p>Änderungen am operativen KI-System geschehen gemäß einem definierten Änderungsprozess und einem Rollout-Plan. Ein Rollback auf die vorherige Version ist jederzeit möglich.</p>		<ul style="list-style-type: none"> • Prozessbeschreibung • Festlegung des Vorgehens bei Rollback • Rollen und Befugnisse 		

Prozessphase 11 Kontinuierliche Verbesserung				
11.5	Ist eine Überwachungs- und Retraining-/ Update-Prozedur vorhanden?	Anwendbar für Risikoklasse		
		AIQM-1	AIQM-2	AIQM-3
		Nein	Ja	Ja
Bewertungsrelevante Anforderungen		Beispiele zur Umsetzung		
<p>Es ist zu definieren bzw. abzuschätzen, wie oft sich der Prozess ändert und wie oft das Modell aktualisiert oder getestet werden sollte.</p> <p>Es ist zu überprüfen, ob die Eingabedaten noch in der definierten Datenverteilung vorliegen.</p> <p>Aktualisieren bedeutet, dass das Modell neu trainiert oder durch ein neues Modell ersetzt werden muss.</p> <p>In der Regel ist ein Modellupdate erforderlich, wenn sich die Datenbasis und/oder das System (aufgrund neuer Erkenntnisse oder neuer Funktionen) geändert hat, z. B.:</p> <ul style="list-style-type: none"> - Änderung der Rechenumgebung/des Messsystems - Aktualisierung der Maschinenfirmware - Einführung neuer Technologien - Änderung der KPIs - Notwendigkeit von Sicherheits- und Datenschutzverbesserungen aufgrund neuer Kunden-/Regulierungsanforderungen oder Erkenntnisse zur Cybersicherheit <p>Generative KI: Beim Einsatz generativer KI müssen sich Modelle an sich entwickelnde Daten anpassen, was Umschulungsstrategien und Degradationsüberwachung erfordert.</p>		<ul style="list-style-type: none"> • Prozessbeschreibung • Festlegung des Vorgehens bei Rollback • Rollen und Befugnisse 		

6 Handlungsempfehlungen und Anwendungsbeispiele

Das Ziel des Kapitels ist es, Handlungsempfehlungen zu definieren, wie KI-Anwendungen betrieben, angepasst und deren Ergebnisse fachlich interpretiert und bewertet werden können.

Dabei wird KI nicht als isolierte Technologie betrachtet, sondern als integraler Bestandteil eines modernen, datenbasierten QM-Systems – von der Wissensbereitstellung über die Prozessanalyse bis hin zur automatisierten Bewertung.

Um die Nutzung und Validierung dieser Anwendungsbeispiele nachvollziehbar zu machen, werden diese jeweils anhand der unten aufgeführten Kategorien beschrieben.

Die **Handlungsempfehlungen für die Anwendungsbeispiele** sind in folgender Struktur aufgeteilt:

- **Beschreibung:** Worum geht es bei dem Anwendungsfall? Kurz und konkret: Ziel, Funktionsweise und Ergebnis des KI-Systems.
- **Rahmenbedingungen:** Welche Voraussetzungen müssen erfüllt sein, damit der Anwendungsfall sinnvoll und sicher funktioniert (z. B. Inhalte, Quellen, Verantwortlichkeiten, Infrastruktur)? Welche Maßnahmen reduzieren das Risiko des Anwendungsbeispiels?
- **Mehrwert:** Der konkrete Nutzen für Qualität, Zeit, Kosten, Sicherheit oder Zusammenarbeit. Was verbessert sich gegenüber dem bisherigen Vorgehen?
- **Herausforderungen:** Typische Stolpersteine, Risiken und Schwierigkeiten, die auftreten können (z. B. fachlich, technisch, organisatorisch, rechtlich).
- **Vorgehensweise:** Wie setzt man den Anwendungsfall um und betreibt ihn (z. B. Aufbau, Konfiguration, Test/Validierung, Pilot, Rollout, Pflege)?
- **Beispiel:** Eine greifbare, reale Nutzungssituation mit konkretem Ablauf.
- **Umgang mit Änderungen (auf Basis des Beispiels):** Wie werden Änderungen schnell und kontrolliert vorgenommen, ohne eine neue

Gesamtfreigabe zu benötigen (z. B. Änderungsklassifizierung, Leitplanken, Prozessbeschreibung für Änderungen ...)?

- **Interpretation und Bewertung des KI-Ergebnisses (auf Basis des Beispiels):** Wie wird das KI-Ergebnis fachlich geprüft und anhand klarer Regeln bewertet, damit es nachvollziehbar, sachlich passend und methodisch korrekt verwendet werden kann (z. B. durch Transparenz der KI erstellte Inhalte, Validierung mittels Referenzbeispielen, Feedback-Mechanismen ...)?

Die nachfolgend beschriebenen KI-Verfahren dienen ausschließlich der Orientierung und der systematischen Einordnung möglicher Lösungsansätze. Die Darstellung erhebt keinen Anspruch auf Vollständigkeit und bildet nicht zwingend den jeweils aktuellen Stand der Technik ab. Welche KI-Verfahren im konkreten Anwendungsfall geeignet sind, ist stets projektspezifisch zu bewerten und von Fall zu Fall zu entscheiden (u. a. abhängig von Zielsetzung, Datenlage, Risiken, regulatorischen Anforderungen, Verfügbarkeit und Reifegrad der Technologie).

Folgende KI-Verfahren werden betrachtet:

- **Sprach- und textbasierte KI (Natural Language Processing [NLP], Large Language Models [LLM], Retrieval Augmented Generation [RAG]):**
Diese Systeme ermöglichen die semantische Verarbeitung, Strukturierung und Generierung von Texten. Sie verstehen natürlichsprachliche Eingaben, klassifizieren Inhalte, erkennen Zusammenhänge und liefern kontextbezogene Antworten. Technologien wie Chatbots fallen in diese Kategorie und können beispielsweise zur qualifizierten Beantwortung von Fragen zu Regelwerken eingesetzt werden.
- **Multimodale KI:**
Multimodale KI-Lösungen verarbeiten unterschiedliche Eingabeformen – etwa gesprochene Sprache, Bilder, Audio- oder Textdaten in gemeinsamen Modellen. Ein Beispiel hierfür ist das „Speech Mining“, bei dem gesprochene Beschreibungen von Tätigkeiten transkribiert und mit visuellen oder prozessualen Daten verknüpft werden, um Arbeitsanweisungen zu generieren.

- **Assistive KI:**
Dies bezeichnet eine unterstützende Form Künstlicher Intelligenz, die Nutzer:innen interaktiv durch Aufgaben führt, fehlende Informationen erfragt und kontextbezogene Vorschläge liefert. Assistive KI-Systeme gehen über klassische Analysefunktionen hinaus, indem sie neue Inhalte oder Handlungsempfehlungen erzeugen. Durch Multi-Agenten-Architekturen können dabei komplexe Aufgaben in spezialisierte Teilschritte zerlegt werden.
- **Governance- und regelorientierte KI:**
Diese Technologie kombiniert KI-Verfahren mit vordefinierten Bewertungs- und Entscheidungslogiken. Ergebnisse entstehen hierbei oft nach festen, nachvollziehbaren Regeln (deterministisch). Sie bildet das Rückgrat für die strukturierte Anwendung von Qualitätsmethoden, wobei der Fokus auf Erklärbarkeit, Nachvollziehbarkeit und der Einhaltung regulatorischer Anforderungen liegt.
- **Agentische KI (Agentic AI):**
Diese Systeme erweitern generative Modelle um die Fähigkeit zur Handlungsplanung und Werkzeugnutzung. Sie können autonom komplexe Aufgabenketten im QM-Prozess steuern, indem sie nicht nur Inhalte erstellen, sondern über definierte Schnittstellen (Tools) aktiv Aktionen ausführen – wie beispielsweise das selbstständige Zusammentragen von Reklamationsdaten aus verschiedenen Systemen zur Vorbereitung einer Entscheidungsfindung.
- **Computer Vision (CV):**
CV beschreibt ein KI-Verfahren zur automatisierten Auswertung von Bild- und Videodaten. Im Qualitätsmanagement wird CV vor allem eingesetzt, um Merkmale, Abweichungen und Fehlerbilder objektiv, reproduzierbar und in hoher Taktung zu erkennen. Entweder als Unterstützung der Werker-Prüfung oder als vollautomatisches Prüfverfahren.
- **Daten- und prozessanalytische KI:**
Daten aus Prozessen und anderen Quellen, z. B. Spezifikationen, Sensormessungen und Qualitätsprüfungen, können zur Entwicklung von Datenmodellen für die Qualitätssicherung in Produktionsprozessen verwendet werden. Die in den Daten erkannten Muster ermöglichen die Detektion von Anomalien, die Vorhersage von Qualitätsmerkmalen und Ursachenanalysen. Auf diese Weise

können vorausschauende Maßnahmen zur Verbesserung von Produktionsprozessen abgeleitet werden.

6.1 KI-gestützte optische Qualitätskontrolle

Beschreibung

Automatisierte Machine-Vision-Systeme werden zur optischen Qualitätsprüfung in der Produktion eingesetzt. Mithilfe von Kamerasystemen werden Bilddaten von Bauteilen oder Baugruppen erfasst und anschließend ausgewertet, um Fehler, Abweichungen oder Unregelmäßigkeiten zu erkennen. Die Auswertung der Bilddaten kann sowohl durch klassische Bildverarbeitungsverfahren als auch durch datengetriebene Verfahren der Künstlichen Intelligenz erfolgen. Ziel dieser Systeme ist es, eine robuste, reproduzierbare und schnelle Qualitätskontrolle zu ermöglichen, die menschliche Inspektionen unterstützt oder ersetzt und gleichzeitig die Prozesssicherheit erhöht.

Bei **regelbasierten Verfahren** werden Bilddaten anhand deterministischer, vorab definierter Regeln ausgewertet. Dabei werden beispielsweise Merkmale wie Helligkeit, Kontrast, Form oder Größe analysiert, um Objekte oder Strukturen im Bild zu identifizieren. Diese Methoden eignen sich besonders für einfache und klar definierbare Prüfaufgaben. Bei komplexeren Bildinhalten oder stark variierenden Bedingungen, etwa durch unterschiedliche Beleuchtung, Position oder Oberflächenbeschaffenheit, stoßen regelbasierte Verfahren jedoch häufig an ihre Grenzen.

Neben klassischen Bildverarbeitungsverfahren kommen zunehmend **KI-basierte Verfahren** zum Einsatz. Dazu gehören folgende Methoden:

Objektklassifikation: Ein Bild oder Bildausschnitt wird einer bekannten Kategorie zugeordnet, beispielsweise „Schweißnaht“ oder „Schraube“. Das System erkennt somit, welche Objektklasse im Prüfausschnitt vorhanden ist, bestimmt jedoch nicht die genaue Position des Objekts im Bild.

Objekterkennung: Hier wird nicht nur festgestellt, welche Objekte vorhanden sind, sondern auch ihre Position im Bild bestimmt. Die Lage

der Objekte wird typischerweise durch Begrenzungsrahmen (Bounding Boxes) dargestellt.

Segmentierung: Dabei wird jeder Pixel im Bild einer Klasse oder einem Objektbereich zugeordnet. Das Ergebnis ist eine segmentierte Darstellung, häufig in Form einer farbigen Maske, die zeigt, welche Bildbereiche zu welchen Objektklassen gehören.

Anomalie-Erkennung: Das System erlernt anhand von Beispielen den Normalzustand eines Produkts oder Prozesses. Anschließend können Abweichungen erkannt werden, die auf ungewöhnliche oder fehlerhafte Zustände hinweisen. Dieses Verfahren ist besonders geeignet, wenn seltene Fehler erkannt werden sollen und nur wenige oder keine Beispielbilder fehlerhafter Produkte verfügbar sind.

Darüber hinaus existieren weitere spezialisierte Verfahren.

In Abhängigkeit von den zu prüfenden Daten, den Randbedingungen der Bildaufnahme sowie den Anforderungen an die Prüfung können einzelne der oben beschriebenen Ansätze mehr oder weniger geeignet sein. Die Auswahl und Bewertung der Verfahren ist daher im jeweiligen Anwendungsfall vorzunehmen.

Rahmenbedingungen

Optische Qualitätskontrollen werden in einem Produktionsumfeld eingesetzt, in dem hohe Stückzahlen, kurze Taktzeiten und stabile Prozesse entscheidend sind. Bauteile können unterschiedliche Varianten, Oberflächen, Materialien und Fertigungstoleranzen aufweisen, was die Bildauswertung anspruchsvoll macht. Die aufgetretenen Fehlerbilder können sehr vielfältig sein, beispielsweise Kratzer, Maßabweichungen oder Montagefehler. Daher muss das Prüfsystem ausreichend flexibel konfigurierbar sein. Produktionslinien bieten häufig nur begrenzten Bauraum, sodass Kamera- und Beleuchtungskonzepte exakt auf die Einbausituation abgestimmt werden müssen. Gleichzeitig können äußere Einflüsse wie Beleuchtungsschwankungen, Verschmutzungen oder Vibrationen die Bildqualität beeinflussen und müssen bei der Systemauslegung entsprechend berücksichtigt und technisch kompensiert werden.

Darüber hinaus ist eine Integration der Lösung in die bestehende Anlagen- und IT-Landschaft erforderlich. Dazu gehören beispielsweise Produktionssteuerungssysteme wie SPS, Manufacturing-Execution-Systeme (MES), Rückverfolgbarkeits-Lösungen sowie gegebenenfalls Edge- oder Cloud-Infrastrukturen für Datenverarbeitung und -speicherung. Werden Bilddaten gespeichert oder für das Training von KI-Modellen verwendet, sind zudem Anforderungen an Datenschutz und Datensicherheit zu berücksichtigen.

Auch die Anwenderperspektive spielt eine wichtige Rolle. Operatoren und Qualitätsmitarbeitende benötigen verständliche Bedien- und Visualisierungskonzepte, um Prüfergebnisse nachvollziehen und das System bei Bedarf anpassen zu können. Darüber hinaus ist die Verfügbarkeit von Fachexpertise in den Bereichen Bildverarbeitung, Qualität und Produktionsprozesse eine wichtige Voraussetzung für den erfolgreichen Betrieb und die kontinuierliche Weiterentwicklung solcher Systeme.

Insgesamt bewegen sich die Rahmenbedingungen optischer Qualitätskontrolle zwischen technischen Restriktionen, prozessbedingten Anforderungen und dem Ziel, eine robuste, stabile und skalierbare Inspektionslösung zu realisieren.

Mehrwert

- **Reproduzierbare und objektive Qualitätsprüfung**, unabhängig von Tagesform oder Erfahrung einzelner Mitarbeitender.
- **Fehlerreduktion durch frühzeitige Erkennung** von Abweichungen im Prozess.
- **Höhere Prozesssicherheit** und geringere Ausschuss- sowie Nacharbeitskosten.

Herausforderungen

- **Hoher Aufwand für Beleuchtung & Optik**, da Beleuchtungsbedingungen einen wesentlichen Einfluss auf die Bildqualität und die Fehlererkennung haben.
- **Variantenvielfalt**, die flexible Prüfstrategien und robustes Training erfordert.
- **Qualität der Trainingsdaten** bei KI-basierten Systemen

- **Lückenlose Dokumentation** von Prüfergebnissen und Bilddaten für Audits und Rückverfolgbarkeit.
 - **Stabile Prüfung** auch bei hohen Taktzeiten und komplexen Inspektionen.
 - **Entlastung der Mitarbeitenden** von monotonen, ermüdenden Sichtprüfungen.
 - **Schnelle Anpassung von Inspektionen** bei Produkt- oder Prozessänderungen (insbesondere mit KI-Unterstützung).
- (Fehlerbilder, Gutteile, Grenzfälle).
- **Änderungen im Produkt oder Produktionsprozess**, die die Vision-Parameter beeinflussen können.
 - **Wartung im laufenden Betrieb** (Reinigung der Optik, Kalibrierung, Verschleiß).
 - **Integration in die Automatisierungsumgebung**, einschließlich Echtzeitfähigkeit.
 - **Akzeptanz im Shopfloor**, insbesondere wenn KI-Systeme als schwer nachvollziehbar (Blackbox) wahrgenommen werden.
 - **Kosten für Skalierung**, besonders bei mehreren Linien oder Standorten.
 - **Datenmanagement und Speicherung großer Bilddatenmengen**, insbesondere bei langfristiger Archivierung für Rückverfolgbarkeit oder KI-Training und -Validierung.
 - **Validierung und Absicherung von KI-basierten Prüfsystemen**, insbesondere hinsichtlich Robustheit, Nachvollziehbarkeit und Freigabeprozessen.

Vorgehensweise

Für den Eignungsnachweis ist VDA-Band 5.3 zu berücksichtigen.

Hinweis: Es wird empfohlen, die in Kapitel 6 beschriebenen „Eignungsnachweise von attributiven Prüfprozessen“ anzuwenden. Ein KI-Output (z. B. ein Konfidenzwert bzw. „Confidence Score“) stellt keinen Messwert dar und eignet sich daher nicht als quantitatives Merkmal für den Fähigkeitsnachweis.

QM-Rollen

Mitarbeiter:innen in der Produktionsqualität

6.1.1 Beispiel

Bei der Prüfung einer Strukturstreben-Variante soll die KI ein Bild auswerten, die Strukturstrebe im Bild lokalisieren und anschließend ausgeben, welche Variante der Strebe verbaut ist. Die KI übernimmt damit die Bildanalyse und Variantenklassifikation, der Mensch bewertet und nutzt das Ergebnis zur weiteren Qualitätssicherung.

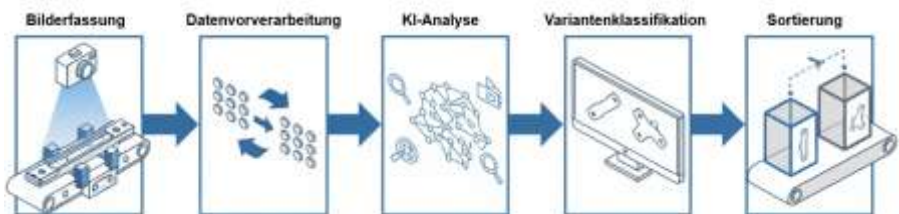


Abbildung 6-1: Schematische Darstellung der fünfstufigen Prozesskette einer KI-gestützten optischen Qualitätskontrolle

Umgang mit Änderungen	Interpretation und Bewertung des KI-Ergebnisses
<ul style="list-style-type: none"> • Kleinere Änderungen (Klasse A): Bei Änderungen des Modells, der Gewichte oder Auswerte-Logik ist dieses zu Versionieren und die Änderungen zu dokumentieren. Bei Auswirkungen auf das Prüfergebnis ist ein erneuter Fähigkeitsnachweis durchzuführen. • Fachlich relevante Änderungen (Klasse B) gezielt testen: Wesentliche Änderungen am zu prüfenden Objekt, an Sensorik oder Fremdeinflüssen sollten im Einzelfall betrachtet werden und können zu der Notwendigkeit einer erneuten Bewertung des gewählten Prüfverfahrens und der benötigten Sicherheitsmaßnahmen führen. • Grundlegende Änderungen (Klasse C) als neue Systemversion behandeln: Wenn mit einer neuen Methode oder mit einer etablierten Methode neue Datenarten geprüft werden. Sowie bei generellen 	<ul style="list-style-type: none"> • Risiko: KI erkennt und klassifiziert ein anderes Bauteil statt der Strebe → Vor der Variantenentscheidung sollte geprüft werden, ob die erkannte Position der Strebe plausibel ist, z. B. durch einen regelbasierten Positionsabgleich mit einem Referenzmerkmal wie einer Schraube oder einem definierten Befestigungspunkt im Bild. Nur wenn die KI-Erkennung räumlich zur erwarteten Lage der Strebe passt, wird die Klassifikation weiter berücksichtigt. • Risiko: Abweichende Kameraperspektive (Blickwinkel/Zoom) im Vergleich zum Trainings-/Validierungsset → Bei der Interpretation der KI-Ergebnisse ist zu prüfen, ob alle relevanten Referenzmerkmale (z. B. Schrauben, Kanten, Bohrungen) an den erwarteten Positionen im Bild sichtbar sind. Wenn Referenzmerkmale stark verschoben, nur teilweise oder gar nicht zu erkennen

Erstumssetzungen, bspw. neuer Lieferant oder Erstanwendung in der Organisation.

- sind, sollte das KI-Ergebnis als unsicher gewertet und manuell überprüft werden.
- **Risiko: Klassifikation wird durch Licht, Verschmutzung oder Verdeckung „halluziniert“**
→ Zur Absicherung der KI-Aussage kann die Größe, Form und/oder Farbintensität (z. B. RGB-Werte) der erkannten Strebe mit bekannten Erwartungswerten aus validierten Referenzdaten verglichen werden. Weichen diese Merkmale deutlich vom erwarteten Bereich ab (z. B. zu dunkler Bereich wegen Schatten, untypische Farbwerte), sollte das Ergebnis als kritisch eingestuft und einer zusätzlichen Kontrolle unterzogen werden.
→ Statt nur eine Region zu betrachten, sollte die KI mehrere voneinander unabhängige Bereiche der Strebe klassifizieren (z. B. Kopfbereich, Mittelstück, Anschlussbereich). Stimmen die Klassifikationsergebnisse dieser Teilbereiche überein, erhöht dies die Glaubwürdigkeit. Bei widersprüchlichen

	<p>Ergebnissen (z. B. ein Bereich Variante A, ein anderer Variante B) ist das Gesamtergebnis als unsicher zu bewerten.</p> <p>→ Die Bewertung des KI-Ergebnisses sollte auf mehreren redundanten Merkmalen beruhen (z. B. Kontur, Lochbild, Lage zu Schraube, Farb-/Oberflächenmerkmal). Passen diese Merkmale zueinander (Positionsabgleich und Formkongruenz), spricht das für ein plausibles Ergebnis. Stimmen sie nicht überein, ist eine manuelle Nachprüfung oder eine Wiederholung der Bildaufnahme sinnvoll.</p>
--	--

6.2 Regelorientierte KI-Agenten zur Unterstützung des 8D-Prozesses

Beschreibung

Dieses Anwendungsbeispiel beschreibt den Einsatz eines KI-basierten Agentensystems zur Unterstützung bei der Erstellung und Qualitätssicherung von 8D-Berichten. Das System kombiniert generative KI (für Entwürfe und Formulierungen) mit regelorientierter Prozesssteuerung (für Vollständigkeit, Konsistenz, Nachweisführung und Freigaben).

Im Kern unterstützen spezialisierte Agenten u. a. bei:

- der Informationsaufnahme (z. B. aus Reklamationsdaten, Sperr-/Sortierprotokollen, CAQ/DMS),
- der Erstellung strukturierter Entwürfe (insbesondere Problembeschreibung D2, optional auch D3/D4 ...),
- der Recherche ähnlicher historischer Fälle und Verknüpfung von Referenzen,
- dem Erkennen von Informationslücken/Widersprüchen und dem Stellen gezielter Rückfragen.

Ein übergeordneter Guard-Agent, der als automatisierter Aufseher zu verstehen ist, prüft die Ergebnisse anhand definierter Regeln (z. B. Pflichtinhalte je Disziplin, Konsistenz zwischen D2–D4, Evidenz/Nachweisführung, Risikoindikatoren, Eskalations- und Freigabekriterien) und steuert den Prozess über Freigabepunkte („Gates“). Zusätzlich werden fest programmierte Workflow-Regeln eingeführt (z. B. Pflichtfelder, Statuslogik, Sperren bei fehlenden Nachweisen, automatische Eskalation bei Risikoindikatoren, 4-Augen-Freigabe), die unabhängig von Prompts wirken und Umgehungen verhindern.

Ziel ist, wiederkehrende Dokumentationsaufwände zu reduzieren und gleichzeitig die Qualität, Standardisierung, Auditierbarkeit und Freigabefähigkeit der 8D-Dokumentation zu erhöhen. Das Vorgehen ist so ausgelegt, dass die KI unterstützt, jedoch keine fachliche Entscheidung ersetzt. Verantwortung und Freigabe liegen weiterhin bei Personen mit entsprechenden Berechtigungsrollen.

Rahmenbedingungen

- **Prozess & Standards:** Definierter 8D-Standard im Unternehmen (Templates, Pflichtinhalte je Disziplin, Rollen/Freigaben, Eskalationswege), angebunden an QM-Vorgaben/Verfahrensweisungen.
- **Daten- und Wissensbasis:** Vorab definierte, testbare und versionierte Regeln für Vollständigkeit, Konsistenz, Evidenz/Nachweisführung, Risikoindikatoren, Eskalation und Freigabe.

- Rollen & Governance: Klare Benutzerrollen, Zugriffsrechte, Verantwortlichkeiten (inkl. Freigabekompetenzen).
- Kontinuierlicher Feedback-/Verbesserungsprozess (z. B. Nutzerbewertungen oder Lessons Learned).
- Datenschutz & Informationsschutz: Regelungen für personenbezogene Daten, Vertraulichkeitsstufen, Maskierung/Anonymisierung sowie Vorgaben zur externen Kommunikation.

Mehrwert

- Prozessbeschleunigung durch schnellere Entwürfe und reduzierte manuelle Dokumentation.
- Höhere Daten- und Ergebnisqualität durch strukturierte, vollständige und konsistente Inhalte je Disziplin.
- Einheitliche Formulierungen, nachvollziehbare Quellen- und Änderungs-dokumentation helfen bei der Standardisierung und Auditierbarkeit.
- Harte Prozessregeln verhindern Umgehungen.
- Offene Punkte werden systematisch nachgefordert.
- Wissenssicherung durch Wiederverwendung von Strukturbausteinen, Lessons Learned und

Herausforderungen

- Regeln müssen eindeutig, widerspruchsfrei, testbar und versioniert sein.
- Balance zwischen zu strengen Gates, die blockieren, und zu weichen Gates, die Fehler zulassen.
- Strikte Quellen- und Evidenzpflicht sowie geeignete Guard-Mechanismen sind notwendig, um Halluzinationen und Scheingenauigkeit vorzubeugen.
- KI unterstützt, entscheidet aber nicht und birgt das Risiko von Übervertrauen.
- Datenqualität und Datenverfügbarkeit.

Vergleichbarkeit ähnlicher Fälle.

Vorgehensweise

- Use Case & Scope festlegen: Welche Disziplinen werden unterstützt, welche Outputs sind verbindlich und welche Entscheidungen bleiben beim Menschen?
- Daten-/Wissensbasis aufbauen: Datenquellen anbinden, Berechtigungen klären, Referenzierung (IDs/Links) für Evidenz sicherstellen.
- Agentenrollen definieren: Beispielsweise Erfassungs-Agent (Input/Ähnlichkeitssuche), Formulierungs-Agent (D2-Entwurf), Analyse-Agent (optional D4-Hypothesen), Guard-Agent (Prüfung/Steuerung).
- Regelwerk aufsetzen & versionieren: Vollständigkeit, Konsistenz, Evidenz, Risikoindikatoren, Eskalation und Freigabe definieren sowie entsprechende Eigner und den Change-Prozess festlegen.
- Harte Prozessgates implementieren: Pflichtfelder, Statuslogik, Sperren ohne Evidenz, automatische Eskalationen.
- Validierungsmechanismen: Plausibilitätschecks, Widerspruchserkennung, „Ist/Ist nicht“-Logik, Quellenpflicht für kritische Aussagen.
- Pilot & Rollout: Testfälle/Regressionstests, Bewertung von Regeltreffern/Fehlalarmen, Schulung, schrittweise Ausweitung.

QM-Rollen

Mitarbeiter:innen in der Lieferantenqualität, in der Kundenqualität, in der Produktionsqualität

6.2.1 Beispiel

Ein Kunde reklamiert einen sporadischen Funktionsausfall. Die zuständige QM-Mitarbeiterin oder der zuständige QM-Mitarbeiter startet einen

8D-Bericht und gibt zunächst wenige Eckdaten ein (Baureihe, Bauteil, Funktion, kundenseitige Fehlerbeschreibung).

- Der Erfassungs-Agent sucht automatisch nach ähnlichen Fällen in freigegebenen historischen 8D-Berichten sowie Reklamationen und liefert Treffer inkl. Referenzen. Die QM-Mitarbeiterin oder der QM-Mitarbeiter markiert, welche Fälle tatsächlich relevant sind.
- Der Formulierungs-Agent erstellt auf Basis der vorhandenen Daten einen strukturierten D2-Entwurf (z. B. mit Auftretensbedingungen, Häufigkeit, Abgrenzung, Symptomen/Warmmeldungen). Fehlende Informationen erkennt er und stellt gezielte Rückfragen an die QM-Mitarbeiterin oder den QM-Mitarbeiter (z. B. Zeitraum/Produktionsstand, Varianten, Randbedingungen).
- Der Guard-Agent prüft den Entwurf auf Einhaltung von Regeln (z. B. Pflichtinhalte, Konsistenz, Quellen-/Evidenzhinweise) und setzt ggf. einen Gate-Status (z. B. „Offen“, „Ergänzung notwendig“ oder „Freigabefähig“).
- Harte Workflow-Regeln verhindern das Weitergehen im Prozess, wenn Pflichtfelder/Evidenz fehlen (z. B. Sperre bestimmter Statuswechsel oder 4-Augen-Review).

Alle KI-Vorschläge bleiben editierbar. Änderungen werden transparent protokolliert inkl. Angabe, ob KI-generiert oder durch Menschen angepasst. Freigabe erfolgt ausschließlich durch berechtigte Personen.



Abbildung 6-2: Schematische Darstellung des regelorientierten KI-Agenten-Workflows im 8D-Prozess

<p>Umgang mit Änderungen</p>	<p>Interpretation und Bewertung des KI-Ergebnisses</p>
-------------------------------------	---

- **Änderungsklassen definieren:** Unterteilung der Anpassungen am KI-System in Klassen – bezogen auf den oben beschriebenen Ablauf mit Erfassungs-Agent, Formulierungs-Agent, Guard-Agent und harten Workflow-Regeln:
 - A = rein sprachlich/kosmetisch (ändert weder Trefferliste und D2-Inhalte noch Gate-Status).
 - B = fachlich relevant innerhalb des gleichen Rahmens (verbessert z. B. Suche/Regeln/Prompts, aber keine neue Rolle der KI, kein neues Risikoprofil).
 - C = grundlegend verhaltens- oder risikoprofiländernd (z. B. KI bewertet Ursachen/Maßnahmen oder ändert Leitplanken/Eskalationslogik). Klasse C löst eine Neufreigabe aus.
 - **Kleinere Änderungen (Klasse A):** Änderungen mit rein sprachlichem oder kosmetischem Charakter (z. B. klarere Formulierungen, zusätzliche Beispiele, Tippfehlerkorrekturen), die das fachliche Verhalten des Systems nicht verändern.
- Beispiel: Umformulierungen der Textbausteine, die der Formulierungs-Agent für den

- **KI-generierte Inhalte sind als Entwürfe/Vorschläge zu betrachten** (keine automatische Übernahme in den 8D-Bericht).
- **Systematische Nutzung von Erfahrungswissen:** Der Agent schlägt ähnliche Fälle vor. Die Anwenderin oder der Anwender wählt aktiv aus, welche Fälle tatsächlich vergleichbar/relevant sind (z. B. bzgl. Änderungsstand, Variante, Linie/Anlage, Rand- und Umgebungsbedingungen). Diese Auswahl (inkl. Begründung/Markierung) wird im System dokumentiert, sodass nachvollziehbar ist, dass vorhandene Erfahrung bewusst einbezogen wurde.
- **Aussagen mit Relevanz für Freigabe oder Risiko müssen referenzierbar sein** (Quelle/Evidenz/Verweis auf Fall, Dokument, Messwert, Ticket etc.).
- **Treffer aus ähnlichen Fällen kritisch prüfen** (z. B. Änderungsstände,

D2-Entwurf nutzt, ohne die inhaltliche Struktur/Prüflogik zu ändern.

Beispiel: Anpassung von Hilfetexten/Labels in der Eingabemaske (Baureihe, Bauteil, Funktion, Fehlerbeschreibung).

Können nach einem kurzen fachlichen Vier-Augen-Review und Dokumentation ohne neue formale Freigabe in Betrieb gehen.

Dienen primär der besseren Verständlichkeit und Nutzbarkeit, nicht der inhaltlichen Erweiterung.

- **Fachlich relevante Änderungen (Klasse B):** Änderungen, die das Unterstützungsverhalten fachlich erweitern oder schärfen, ohne den grundlegenden Einsatzzweck bzw. das Risikoprofil des Systems zu verändern (z. B. zusätzliche Datenquellen für ähnliche Fälle, neue Prüfreden innerhalb des bestehenden Rahmens).

Beispiele im Kontext:
Der Erfassungs-Agent nutzt zusätzliche freigegebene Datenquellen (z. B. weitere interne Reklamationsdatenbank) zur Ähnlichkeitssuche, bleibt aber

Varianten, Linien, Anlagen oder Umgebungsbedingungen) und nur nach plausibilisierter Vergleichbarkeit übernehmen.

- **Transparente Kennzeichnung und Bestätigung von KI-Vorschlägen:** Alle Textteile, die von der KI vorgeschlagen wurden, sind im System eindeutig erkennbar. Sie müssen aktiv übernommen oder angepasst werden, bevor sie Teil des 8D-Berichts werden. Die Historie zeigt, welche KI-Vorschläge übernommen, geändert oder verworfen wurden.
- **Validierung mit Testfällen:** Vor und während des Einsatzes werden bekannte Reklamationsfälle als Test genutzt. KI-unterstützte Problembeschreibungen werden mit bewährten Referenz-8Ds verglichen. Expert:innen bewerten Vollständigkeit, Klarheit und methodische Stimmigkeit. Abweichungen werden dokumentiert und zur Verbesserung (z. B.

bei „Treffer liefern + Mensch markiert relevant“.

Der Formulierungs-Agent stellt präzisere Rückfragen (z. B. Produktionsstand/Zeitraum/Varianten/Randbedingungen) oder erzeugt einen konsistenteren D2-Entwurf, ohne neue 8D-Schritte zu „entscheiden“.

Der Guard-Agent erhält zusätzliche Regeln (z. B. stärkere Konsistenzprüfung, klarere Evidenz-/Quellenhinweise), setzt aber weiterhin nur Gate-Status und fordert Ergänzungen an (keine automatische Freigabe).

Müssen über einen definierten Change-Prozess beantragt, fachlich bewertet und mit repräsentativen 8D-Testfällen geprüft werden (inkl. Regressionstest, z. B. mit historischen Fällen „sporadischer Funktionsausfall“).

Nach dokumentierter Prüfung und Freigabe können sie in die produktive Anwendung überführt werden.

- **Grundlegende Änderungen (Klasse C):**

Änderungen, die das Verhalten oder das Risikoprofil des Systems grundlegend verändern

Regeln/Prompts/Testkatalog) genutzt.

(z. B. neuer Anwendungsfall, bei dem die KI Ursachen oder Maßnahmen bewertet oder vorschlägt, veränderte Leitplanken, neue Eskalationslogiken).

Beispiele im Kontext:

KI bewertet Ursachen oder Maßnahmen (D4/D5) oder priorisiert Maßnahmen verbindlich, statt nur zu unterstützen.

Änderung der Leitplanken, z. B. Aufweichung von „Freigabe ausschließlich durch Menschen“ oder „KI-Vorschläge sind gekennzeichnet“.

Guard-/Workflow-Logik wird so geändert, dass Statuswechsel auch ohne Evidenz/4-Augen-Review möglich werden oder umgekehrt eine neue Eskalation (z. B. automatische Meldung/Stop) eingeführt wird.

Werden wie eine neue Systemversion behandelt und erfordern eine vollständige Neubewertung (Risiko, Freigabekriterien, Tests, Dokumentation).

Leitplanken wie „KI entscheidet nicht über Ursachen/Maßnahmen“,

„KI-Vorschläge sind gekennzeichnet“ und „Freigaben erfolgen ausschließlich durch Menschen“ dürfen nur mit besonderer Begründung und erweiterter Freigabe geändert werden.

- **Hinweis:** Der Änderungsprozess (wie Änderungen klassifiziert, geprüft, getestet, dokumentiert und in die Anwendung überführt werden) sollte in einem eigenen Prozess beschrieben werden – inkl. Protokollierung, ob Inhalte KI-generiert oder durch Menschen angepasst wurden, und mit klaren Rollen/Verantwortlichkeiten für die Freigabe.

6.3 KI-gestütztes Audit

Beschreibung

Eine KI analysiert die Auditfeststellungen eines internen Prozessaudits, welches beispielsweise nach VDA Band 6.3 durchgeführt wurde. Anhand einer vorhandenen Wissensdatenbank kann die KI mögliche Abstellmaßnahmen für die jeweiligen Abweichungen vorschlagen. Durch Regel- und Governance-orientierte KI wird die Analyse um feste Bewertungslogiken und definierte Entscheidungsregeln ergänzt. Dadurch werden Feststellungen konsistent interpretiert, Regelverstöße klar zugeordnet und Maßnahmenvorschläge nachvollziehbar dokumentiert. Audit-Trails und Regelreferenzen sichern die Nachvollziehbarkeit, während Governance-Mechanismen die Pflege von Regeln, Rollen und Wissensständen unterstützen.

So entsteht ein strukturiertes, reproduzierbares und erklärbares KI-gestütztes Auditverfahren.

Rahmenbedingungen

- Konsistente Wissensdatenbank sowie klar definierte Auditregeln erforderlich (z. B. VDA 6.3, interne Auditstandards). Die Wissensbasis umfasst insbesondere das Prozessverständnis des Unternehmens, frühere Auditberichte sowie ein zentrales Unternehmens-FAQ. Zusätzlich benötigt die KI strukturierte und möglichst einheitliche Auditdokumente, um Feststellungen zuverlässig auszuwerten.
- Definition von Verantwortlichkeiten, Nutzergruppen und mehrstufigem Zugriff basierend auf Fachkenntnis und Feedbackqualität.
- Die Kompetenz zur Bewertung der Maßnahmenvorschläge der KI sollte weiterhin erhalten bleiben.
- Integration in das bestehende Auditmanagementsystem (z. B. zur Wirksamkeitskontrolle).
- Integration von User-Feedback über Daumen hoch/runter oder aktives Text-Feedback.
- Änderungsmanagement für Inhalte und Datenpflege (z. B. Aktualisierung von FAQs).
- Definition fester Kategorien und Prioritäten. Jede Empfehlung muss eine Regelreferenz und Belege aus der Historie aufweisen.

Mehrwert

- Schnellere konsistente Ableitung von Maßnahmen auf Basis der Erfahrung der Organisation.

Herausforderungen

- Qualität vorhandener Auditberichte ist entscheidend.
- Uneinheitliche Auditberichte können die Dokumentation erschweren.

- | | |
|---|--|
| <ul style="list-style-type: none"> • Durch Regel- und Governance-orientierte Entscheidungslogiken werden Bewertungen einheitlicher, nachvollziehbarer und reproduzierbar. • Identifikationen von Maßnahmen, die sich in vergleichbaren Fällen als besonders wirksam erwiesen haben. • Automatisierte Priorisierung möglich. • Automatischer Wissensaustausch sowie -aufbau auch mit anderen Standorten. | <ul style="list-style-type: none"> • Nachvollziehbarkeit der gewählten Abstellmaßnahmen muss sichergestellt werden. • Die Pflege der Wissensdatenbank und der zugrunde liegenden Regeln erfordert klare Verantwortlichkeiten und regelmäßige Aktualisierung. |
|---|--|

Vorgehensweise

- Festlegung und Sammlung relevanter Daten.
- Aufbau einer Wissensdatenbank mit den relevanten Daten einschließlich Regelwerken, Maßnahmenhistorie und Unternehmens-FAQ.
- Entwicklung der entsprechenden KI-Funktionen wie Textanalyse von Auditberichten und Vorschlagsgenerierung zur regelbasierten Ableitung von Maßnahmen.
- Durchführung eines Pilotprojekts mit ausgewähltem Nutzerkreis, um Nutzerfeedback zu sammeln und das Modell anzupassen.
- Vergleich zwischen den Maßnahmenvorschlägen der KI und der Auditorenbewertung, um Konsistenz und Nachvollziehbarkeit sicherzustellen.
- Integration in das Audit-Tool. Bereitstellung der Maßnahmenvorschläge direkt in der Auditsoftware.

- Implementierung eines kontinuierlichen Änderungs- und Verbesserungsprozesses zur Pflege der Wissensbasis und der Entscheidungslogiken.

QM-Rollen

Qualitäts-Auditor:in/Qualitäts-Assessor:in

6.3.1 Beispiel

Im Unternehmen werden interne Prozessaudits nach VDA 6.3 durchgeführt. Jede Feststellung wird dabei von der Auditorin oder vom Auditor direkt der entsprechenden Frage (z. B. P5, P6, P7) zugeordnet.

Da die Zuordnung zur VDA-6.3-Frage bereits im Audit erfolgt, nutzt die KI diese Information als Eingangswert.

Die KI führt anschließend folgende Aufgaben aus:

- Analyse des Feststellungstextes
- Klassifikation der Abweichung anhand fest vorgegebener Regeln (z. B. systemisch, dokumentationsbezogen)
- Abgleich mit historischen Auditfällen aus der Wissensdatenbank
- Vorschlag geeigneter Maßnahmen inklusive Regelreferenzen
- Bereitstellung von nachvollziehbaren Begründungen (Audit-Trail)

Ergebnisdarstellung:

Auditfeststellung: Produktionslenkung „Arbeitsanweisung für Rüstvorgang nicht aktuell und nicht am Arbeitsplatz verfügbar.“

KI-Analyse

- Klassifikation: Dokumentationsabweichung
- Historische Fälle erkannt: Ähnliche Abweichungen in 2023/03 (Audit 2311) und 2024/01 (Audit 8512)
- Vorschlag der KI:
 - Aktualisierung und Freigabe der Arbeitsanweisung

- Sicherstellung der Verfügbarkeit am Arbeitsplatz
- Kontrolle im nächsten Review-Zyklus
- Begründung: „Maßnahme entspricht etablierten Regeln zur Dokumentenlenkung; erfolgreiche Wirksamkeit in früheren Fällen nachgewiesen“



Abbildung 6-3: Sechsstufiger Workflow der KI-gestützten Auditfeststellungsbearbeitung

Umgang mit Änderungen	Interpretation und Bewertung des KI-Ergebnisses
<ul style="list-style-type: none"> • Änderungsklassen definieren: Unterteilung der Anpassungen am KI-System in Klassen (z. B. A: rein sprachlich/kosmetisch, B: fachlich relevant, aber im gleichen Rahmen, C: grundlegend verhaltensändernd). Klasse C würde eine neue Freigabe auslösen. • Kleinere Änderungen (Klasse A): Bei kleineren Änderungen (z. B. klarere Formulierungen, zusätzliche Beispiele, Tippfehlerkorrekturen) reicht ein kurzer fachlicher Vier-Augen-Review. Nach Dokumentation können diese Anpassungen ohne neue Freigabe direkt angewendet werden. • Fachlich relevante Änderungen (Klasse B) gezielt testen: Erweiterungen, die das Verhalten präzisieren, aber die Rolle der KI 	<ul style="list-style-type: none"> • Arbeiten auf Basis klarer Regeln und Wissensbasis: Die KI stützt sich auf definierte Regelwerke (z. B. VDA 6.3, interne Auditstandards), feste Kategorien (z. B. systemisch, dokumentationsbezogen) und eine geprüfte Wissensdatenbank mit historischen Auditfällen und FAQs. Jede Empfehlung enthält eine Regelreferenz und Beispiele aus der Historie. Damit ist nachvollziehbar, auf welchen Regeln und Erfahrungen ein Vorschlag basiert. • Strukturierte, standardisierte Eingabedaten nutzen: Feststellungen sind bereits im Audit der jeweiligen VDA-Frage zugeordnet (z. B. P5, P6, P7) und liegen in einem strukturierten Format vor. Die KI nutzt diese Struktur, klassifiziert die Abweichung nach vorgegebenen Kriterien und sucht gezielt nach passenden Fällen. Das reduziert Interpretationsspielraum und sorgt für

nicht verändern – z. B. neue Kategorien, verfeinerte Klassifikationsregeln, zusätzliche Standardmaßnahmen, neue Regelreferenzen –, werden mit einem kleinen Satz von Test-Auditfällen überprüft. Auditoren beurteilen, ob Klassifikation, Vorschläge und Begründungen weiterhin sinnvoll und konsistent sind.

- **Grundlegende Änderungen (Klasse C) als neue Systemversion behandeln:**

Anpassungen, die das Risikoprofil des Systems verändern (z. B. automatische Bewertung von Auditfragen) erfordern eine neue Freigabe.

- **Festlegen von Leitplanken:** Diese Leitplanken sollten nicht leichtfertig geändert werden. Typische Leitplanken wären z. B.: Die KI gibt nur Vorschläge, keine finalen Bewertungen; jede Empfehlung enthält eine

reproduzierbare Ergebnisse bei gleicher Datengrundlage.

- **Nachvollziehbare Vorschläge mit Quelle:** Zu jeder Empfehlung liefert die KI eine Begründung (z. B. „Dokumentationsabweichung – Maßnahme entspricht Regel zur Dokumentenlenkung; in Fällen X und Y als wirksam bewertet“).
- **Menschliche Prüfung und Freigabe der Maßnahmen:** Der Auditor sieht die KI-Vorschläge inkl. Klassifikation, Regelreferenzen und Historie. Er entscheidet, welche Maßnahmen übernommen, angepasst oder verworfen werden.
- **Feedbackmechanismen als Qualitätskontrolle:** Auditoren können Vorschläge mit „Daumen hoch/runter“ bewerten oder Textfeedback geben („Vorschlag zu allgemein“, „sehr passend“). Dieses Feedback wird genutzt, um Regeln und Wissensbasis zielgerichtet zu verbessern.

<p>Regelreferenz und Hinweise auf historische Fälle; Maßnahmen werden immer von Auditoren entschieden und freigegeben. Bleiben diese Leitplanken unverändert, so benötigt die Anwendung keine neue Freigabe.</p> <ul style="list-style-type: none"> • Hinweis: Der Änderungsprozess, wie Änderungen klassifiziert, geprüft, getestet, dokumentiert und in die Anwendung gehen, sollte in einem Prozess beschrieben werden. 	<ul style="list-style-type: none"> • Validierung mit Referenzaudits und Kennzahlen: In Pilotphasen und regelmäßig danach werden KI-gestützte Maßnahmenvorschläge mit bestehenden, manuell erstellten Auditbewertungen verglichen. Fachleute prüfen, ob die KI konsistent klassifiziert, passende Maßnahmen vorschlägt und die Regelreferenzen korrekt sind. Ergänzend können Kennzahlen ausgewertet werden (z. B. Anteil der Vorschläge, die Auditoren übernehmen, Anzahl der Nachbesserungen nach externen Reviews).
--	---

6.4 KI-gestützte FMEA

Beschreibung

KI-gestützte Lösung zur begleiteten unterstützten Erstellung von FMEA nach VDA-Richtlinien und unternehmensspezifischen Vorgaben.

Das agentenbasierte System unterstützt Moderator:innen und Teams durch wissensbasierte Struktur- und Formulierungsvorschläge und sorgt für konsistente Dokumentation. Durch kontextbasierte Unterstützung hilft die Lösung, Informationen zu erfassen und die Risiken umfassend zu beschreiben und zu minimieren. Zudem liefert sie Vorschläge für Maßnahmen und Formulierungen basierend auf historischen FMEA-Datenquellen.

Rahmenbedingungen

- Verfügbarkeit einer geprüften FMEA-Wissensbasis.
- Verfügbarkeit eines standardisierten FMEA-Datenmodells.
- Entwicklung eines methodischen Modells zur Sicherstellung der methodischen Richtigkeit der FMEA-Analyse.
- Integration von Feedbackmöglichkeiten für Nutzer:innen zur Verbesserung der Lösung.
- Definition von Zugriffsrechten und Verantwortlichkeiten im System
- Implementierung von Logik-Regeln (z. B. Fehlerfolge passt zur Ursache).
- Nur regelkonforme Einträge werden akzeptiert.

Mehrwert

- Qualitativ hochwertigere Risikoanalysen und Minimierungsstrategien, entsprechend VDA-Standards.
- Unterstützung der Anwender:innen durch strukturierte Eingabeprozesse.
- Prozessbeschleunigung durch Nutzung von vorhandenem Wissen der Risikoanalysemethodik zur Unterstützung des Teams.
- Verbesserung der Datenqualität durch konsistentere und präzisere Risikoableitungen und -minimierungen.

Herausforderungen

- Sicherstellung der Richtigkeit der durch den Nutzer gelieferten Daten zum Aufbau des KI-Systems.
- Balancierung zwischen Automatisierung und notwendiger Expertenprüfung.
- Nutzerakzeptanz und Schulung im Umgang mit der KI-Lösung.
- Datenschutz und Sicherheit bei sensiblen Qualitätsdaten.

- Reduktion des manuellen Aufwands bei der initialen Beschreibung struktureller Funktions- und Fehlerableitung.

Vorgehensweise

- Einsatz einer KI-Lösung, die auf einer strukturierten Wissensdatenbank (historische FMEAs, Richtlinien, Unternehmensstandards) basiert und als Assistenzsystem zur Unterstützung von FMEAs dient.
- Aufbau einer strukturierten Wissensbasis mit historischen FMEAs.
- Implementierung von vorformulierten Prompts zur Unterstützung der Moderator:innen und Teams bei der Erstellung der FMEA.
- Pilotierung mit ausgewählten Expert:innen, FMEA-Moderator:innen und Entscheidungsträger:innen zur Sammlung von Feedback und iterativen Verbesserungen der Lösung.
- Schrittweise Ausweitung des Nutzerkreises abhängig von der bestätigten Antwortqualität und Nutzerzufriedenheit.
- Skalierung des Einsatzes nach erfolgreicher Pilotphase und ständiger Anpassung der Wissensbasis sowie der Nutzerführung.
- Implementierung eines kontinuierlichen Änderungs- und Verbesserungsprozesses zur Pflege der Wissensbasis.

Rollen

Mitarbeiter:innen in der Entwicklungsqualität und in der Produktionsqualität

6.4.1 Beispiel

Ein Entwicklungsteam arbeitet an einem neuen Steuergerät für ein Fahrzeug. Bevor das Produkt in Serie geht, soll geklärt werden, wo mögliche Fehler auftreten können und wie man sie vermeidet. Dafür nutzt das Team eine FMEA, also eine strukturierte Risikoanalyse: Man

beschreibt, was das Bauteil tun soll, überlegt, was schiefgehen kann, welche Auswirkungen das hat, warum das passieren könnte und welche Maßnahmen nötig sind, um Fehler zu verhindern oder rechtzeitig zu erkennen.

Statt alles komplett neu aufzubauen, nutzt die FMEA-Moderatorin oder der FMEA-Moderator ein KI-gestütztes Assistenzsystem. Sie/Er gibt kurz ein, worum es geht – etwa „Steuergerät für Fahrdynamikregelung in Fahrzeugplattform Z“. Die KI greift auf eine Wissensbasis aus früheren FMEAs, Richtlinien und internen Standards zu und macht passende Vorschläge: typische Funktionen des Steuergeräts, mögliche Fehler, wahrscheinliche Ursachen, typische Auswirkungen und erprobte Gegenmaßnahmen. Das Team wählt aus, passt an und ergänzt, was für den konkreten Fall relevant ist.

So könnte die KI zum Beispiel vorschlagen, dass ein Fehler „Kommunikationsabbruch zum Sensor“ auftreten kann mit der Auswirkung „Fahrdynamikregelung nicht verfügbar, Warnmeldung für den Fahrer“ und Ursachen wie „lose Steckverbindung“ oder „Leitungsbruch“. Als Maßnahmen schlägt sie etwa robustere Steckverbinder oder eine 100%-Prüfung im End-of-Line-Test vor. Das Team prüft diese Vorschläge, entscheidet, was übernommen wird, und legt fest, wer welche Maßnahme bis wann umsetzt.

Währenddessen kontrolliert die KI im Hintergrund, ob die FMEA logisch und umfassend ist: Ob zu wichtigen Funktionen Fehler erfasst sind, ob kritische Fehler Ursachen und Maßnahmen haben und ob die Kombinationen von Fehlern, Ursachen und Wirkungen plausibel sind. Sie weist die Moderatorin oder den Moderator darauf hin, wenn etwas fehlt oder nicht zusammenpasst. Am Ende entsteht eine FMEA, die schneller erstellt wurde, auf erprobtem Wissen basiert und methodisch stimmig ist – aber alle fachlichen Entscheidungen liegen weiterhin beim Team, die KI unterstützt nur mit Vorschlägen und Prüfungen.

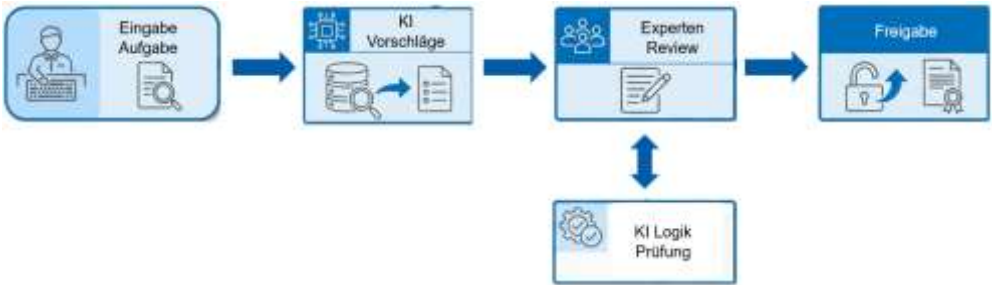


Abbildung 6-4: KI als Assistenzsystem in der FMEA

Umgang mit Änderungen	Interpretation und Bewertung des KI-Ergebnisses
<ul style="list-style-type: none"> • Änderungsklassen definieren: Unterteilung der Anpassungen am KI-System in Klassen (z. B. A: rein sprachlich/kosmetisch, B: fachlich relevant, aber im gleichen Rahmen, C: grundlegend verhaltensändernd). Klasse C würde eine neue Freigabe auslösen. • Kleinere Änderungen (Klasse A): Bei kleineren Änderungen (z. B. klarere Formulierungen, zusätzliche Beispiele, Tippfehlerkorrekturen) reicht ein kurzer fachlicher Vier-Augen-Review. Nach Dokumentation können diese Anpassungen ohne neue 	<ul style="list-style-type: none"> • Dokumentation der Regelprüfungen: Das System protokolliert, welche inhaltlichen und logischen Regeln geprüft wurden (z. B. „Zu jeder Funktion gibt es mindestens einen potenziellen Fehler“, „Jede Ursache ist mit einem Fehler verknüpft“, „Jeder Fehler mit hoher Bedeutung hat mindestens eine präventive oder entdeckende Maßnahme“). • Nachweis im Audit: Welche Regeln sind definiert? Welche Instanz der FMEA wurde mit welchem Ergebnis geprüft? • Moderator:innen-Freigaben: Jede FMEA besitzt eine eindeutige Freigabehistorie (z. B. Wer hat wann welche

Freigabe direkt angewendet werden.

- **Fachlich relevante Änderungen (Klasse B) gezielt testen:** Änderungen, die das Unterstützungsverhalten erweitern, aber die Rolle der KI nicht verändern (z. B. neue Standardbausteine für typische Fehler/Ursachen), werden mit definierten Testfällen und Fachreviews geprüft und dokumentiert.
- **Grundlegende Änderungen (Klasse C) als neue Systemversion behandeln:** Anpassungen, die das Risikoprofil des Systems verändern (z. B.: neuer Anwendungsfall: Die KI soll eigenständig Bewertungen vorschlagen/festlegen) erfordern eine neue Freigabe.
- **Festlegen von Leitplanken:** Diese Leitplanken sollten nicht leichtfertig geändert werden. Typische Leitplanken wären z. B.: Die KI macht nur Vorschläge, entscheidet nicht; sie legt keine Bewertungen endgültig fest;

Version freigegeben? Mit welchem Freigabestatus?).

- **Kennzeichnung von KI-Beiträgen:** Das System markiert, welche Einträge (Fehler, Ursachen, Wirkungen, Maßnahmen) auf KI-Vorschläge zurückgehen; gleichzeitig ist dokumentiert, dass ein Mensch (Moderator:in/Expertin/Experte) diese Vorschläge genehmigt oder angepasst hat.
- **Validierung der Ergebnisse durch Vergleich mit Referenz-FMEAs:** Bestehende, als qualitativ gut akzeptierte FMEAs werden als Referenz genutzt (die KI generiert für einen bekannten Anwendungsfall Vorschläge und diese werden durch die Expert:innen und Moderator:innen verglichen. Deckt die KI die gleichen oder mehr relevante Fehler/Ursachen/Wirkungen ab? Sind die vorgeschlagenen Maßnahmen vergleichbar sinnvoll?).
- **Validierung der Ergebnisse durch unabhängige Expertenreviews:** Unabhängige Expert:innen (nicht am Projekt beteiligt)

jeder FMEA-Eintrag wird von eine:r Moderator:in / eine:r Expert:in geprüft; nur regelkonforme Einträge werden akzeptiert; Freigaben der FMEA erfolgen ausschließlich durch berechnete Personen. Bleiben diese Leitplanken unverändert, so benötigt die Anwendung keine neue Freigabe.

- **Hinweis:** Der Änderungsprozess, wie Änderungen klassifiziert, geprüft, getestet, dokumentiert und in die Anwendung gehen, sollte in einem Prozess beschrieben werden.

prüfen stichprobenartig FMEAs, die mit KI-Unterstützung erstellt wurden (Bewertung der Vollständigkeit [sind wesentliche Risiken abgedeckt?]; Bewertung der Plausibilität [passen Ursachen und Maßnahmen?]) inklusive Reviewprotokolle als Nachweis.

- **Mögliche Kennzahlen:** Anzahl der nachträglichen FMEA-Änderungen durch Reklamationen/Feldrückläufer; Zeitaufwand für FMEA-Erstellung (vor/nach KI-Einsatz); Vollständigkeitsindikatoren (z. B. Durchschnitt der Logik- und Konsistenzfehler vor/nach Einführung). → Die Verbesserung dieser Kennzahlen über die Zeit unterstützt den Nachweis der Prozesssicherheit.
- **Tests der Methoden-Regeln:** Kann man künstliche FMEAs mit bewusst eingebauten Fehlern erzeugen und prüfen, ob das System sie erkennt? Werden Verletzungen der VDA-Methodik zuverlässig gemeldet? Dokumentation der Testfälle und Testergebnisse als Prozess-Qualifizierungsnachweis für das Tool.

6.5 Prädiktive Prozesslenkung

Prädiktive Prozesslenkung (z. B. Predictive Quality) ermöglicht eine proaktive Qualitätssicherung in der Produktion. Statistische Methoden wie SPC, ML-Methoden oder eine Kombination verschiedener Ansätze können frühzeitige Korrekturmaßnahmen ermöglichen. Die daraus resultierenden Datenmodelle können für die Ursachenanalyse verwendet werden, beispielsweise mit Hilfe klassischer Tools wie Ishikawa, zuvor entwickelten Kausalmodellen und/oder erklärbaren KI-Methoden (XAI).

Beschreibung

Der Einsatz von prädiktiver Prozesslenkung kann auf unterschiedliche Weise umgesetzt werden. Ein zentraler Aspekt bei der Konzeption ist die Zeitspanne zwischen einer Vorhersage und der Einleitung von Korrektur- oder Verbesserungsmaßnahmen auf Grundlage der Erkenntnisse. Ziel ist es, die Prozessstabilität und -fähigkeit zu verbessern und gleichzeitig den Umfang der erforderlichen Prüfungen zu reduzieren. Der Mehrwert prädiktiver Prozesslenkung wird bei aufwendigen oder zerstörenden Prüfungen und langen Zeitintervallen zwischen der Vorhersage eines Qualitätsmerkmals oder Erkennung der Abweichung und der Möglichkeit, Gegenmaßnahmen zu ergreifen, maximiert.

Rahmenbedingungen

- **Definition Anwendungsfall:** Ziele, Metriken (z. B. Ausschussquote, Genauigkeit, ROI), Geltungsbereich, Umfang, Anforderungen, Team und Fachbereiche.
- **Expertenwissen und Schulungen:** Einbeziehung von Expert:innen zur Bewertung der Daten- und Prognosequalität und zur Überprüfung von Annahmen, der Plausibilität der Modellergebnisse und der abgeleiteten Erkenntnisse. Schulung der relevanten Stakeholder hinsichtlich der Entwicklung, Nutzung und Wartung der Lösung.
- **Datenverfügbarkeit und -qualität:** Vollständigkeit, Korrektheit, Konsistenz, Aktualität und Relevanz der Daten (z. B. Spezifikationen, Sensormessungen und Qualitätsprüfungen). Die

Daten müssen in ausreichender Menge für die Modellierung verfügbar sein.

- **Automatisierte Erfassung von Messdaten:** Die Grundlage für eine proaktive Steuerung ist die korrekte Erfassung von Sensordaten, die präzise Vorhersagen und entsprechende Maßnahmen ermöglicht.
- **Datenintegration:** Integration und Aggregation von Daten aus unterschiedlichen Quellen für die Analyse und Modellierung.
- **Standardisierung und Datenmanagement:** Einheitlichkeit der Formate zu Erfassung, Austausch und Management von Daten. Kontinuierliche Pflege von Daten zur Aufrechterhaltung und/oder Verbesserung der Datenqualität.
- **Schwellwert-Logik:** Kombination von KI-Vorhersage mit vordefinierten Schwellwerten.
- **Aufsicht und Prüfung:** Automatische Eingriffe erfolgen nur nach Überschreitung der Schwelle und menschlicher Zustimmung (Begründungspflicht).
- **Validierung:** Implementierung von Pilotprojekten, um die Effektivität der Lösungen in kontrollierten Umgebungen zu testen.
- **Governance und Datenschutz:** Festlegung der Verantwortlichkeiten und Einhaltung bestehender Standards und Regularien.
- **Infrastruktur und technische Schnittstellen:** Geeignete Hardware und Software für die Lösungsentwicklung und die Sicherstellung der technischen Integration in bestehende Systeme.

Mehrwert	Herausforderungen
<ul style="list-style-type: none"> • Verbesserung der Produkt- und Prozessqualität: Vorausschauende Einleitung von 	<ul style="list-style-type: none"> • Datenqualität: Prozessveränderungen beeinflussen die Datenqualität, z. B. Kalibrierung und Verschleiß Sensoren.

Verbesserungsmaßnahmen auf Grundlage von Daten.

- **Verbessertes Prozessverständnis:** Mögliche Erkenntnisse über Zusammenhänge innerhalb des Prozesses und optimale Parametereinstellungen.
- **Ressourceneffizienz:** Reduzierung der Prüfumfänge (bei Stichproben-Prüfung), der erforderlichen Ressourcen und der damit verbundenen Kosten.
- **Unausgewogene Datensätze:** Datensätze können hinsichtlich der Ausgewogenheit zwischen zu klassifizierenden oder vorherzusagenden Zuständen verzerrt sein, z. B. bei Defekten, die sich auf die Unsicherheit der Modelle auswirken könnte.
- **Produkt- und Prozesskomplexität:** Komplexe, nichtlineare Zusammenhänge zwischen Prozessparametern und Qualitätsmerkmalen.
- **Nachvollziehbarkeit und Akzeptanz:** Geringe Akzeptanz bei Fachabteilungen aufgrund fehlender Erklärbarkeit der Modelle („Blackbox“-Effekt).
- **Technische Integration und Schnittstellen:** Integration in bestehende IT- und Produktionssysteme (z. B. MES) und Prozesse.

Vorgehensweise

- **Definition Anwendungsfall:** Ziele, Metriken (z. B. Ausschussquote, Genauigkeit, ROI), Geltungsbereich, Umfang, Anforderungen, Team und Fachbereiche.
- **Ist-Aufnahme:** Analyse der betroffenen Prozesse, Prozessparameter und Qualitätsmerkmale sowie initiale Analyse und Bewertung der vorhandenen Daten.
- **Datenvorbereitung und -modellierung:** gegebenenfalls Anpassung oder Erweiterung der Datenanforderungen, Erfassung

relevanter Daten und anschließendes Training des prädiktiven Modells in einer Testumgebung.

- **Erprobung und Validierung:** Implementierung der Lösung in der Produktion und Ableitung von Verbesserungsmaßnahmen. Iterative Verbesserung des trainierten Modells mit Vergleich der definierten Metriken.

QM-Rollen

Mitarbeiter:innen in der Produktionsqualität

6.5.1 Beispiel

Ein interdisziplinäres Team (z. B. Produktionsleitung, Produktionsqualität, Prozessentwicklung, Technologieexpert:innen, IT, Data Scientist) arbeitet daran, Nacharbeiten und zerstörende Prüfungen in Schweißprozessen zu minimieren, z. B. beim Widerstandspunktschweißen (Karosseriebau) oder beim Laserstrahlschweißen (z. B. Getriebeteile oder Batteriepacks). Durch die Vorhersage kritischer Punkte soll die Anzahl manueller Prüfungen an Schweißverbindungen reduziert, die Ausschussquote gesenkt und Anlagenstillstände vermieden werden. Für diese Anwendung werden Daten aus Qualitätsprüfungen (z. B. Ultraschallprüfungen), messbare Umgebungsdaten (z. B. Betriebsmittel, Materialcharge, Temperatur) und relevante steuerbare Prozessparameter der Anlage (z. B. je nach Technologie Schweißzeit, Schweißgeschwindigkeit, Schweißstrom, Laserleistung) benötigt.

Die erforderlichen Daten werden z. B. aus Qualitätsprüfungen, Sensordaten und/oder MES-Systemen gesammelt und können zur Entwicklung von Testanwendungen separat gespeichert werden. Während der Datenaufbereitung werden erste Analysen der Datenverteilungen und Ausreißer durchgeführt. Die Qualität der Daten wird bewertet (z. B. Vollständigkeit, Interpretierbarkeit und Zuordenbarkeit der Qualitätsdaten). Mit Hilfe von Expert:innen werden die Daten bereinigt und für die Modellierung aufbereitet (je nach Datenformat kann dies einen aufwendigen Schritt darstellen).

Bei einer Klassifikationsanalyse werden die Daten gekennzeichnet (i. O. und n. i. O.), Schwellenwerte und Metriken (z. B. Prädiktionsintervalle und Klassifikationsmetriken) für die KI-Modelle definiert und bei Bedarf zusätzliche Features aus den vorhandenen Daten entwickelt, um den bestehenden Prozess und Ziele des Anwendungsfalls besser abzubilden. Verschiedene Algorithmen werden ausgewählt (z. B. neuronale Netzwerke, Random Forests), um die Modelle zu trainieren und die Vorhersagegenauigkeit (Unsicherheit) zu bewerten.

Die resultierenden Modelle werden durch Testen verschiedener Hyperparameter (Modellparameter) optimiert und anhand zusätzlicher historischer Daten getestet. Mit Hilfe von Sensitivitätsanalysen oder XAI-Methoden können die Einflüsse verschiedener Parameter im Modell auf die Vorhersageergebnisse (z. B. Parametergrenzen für eine Vorhersage als i. O. oder n. i. O.) ermittelt und mit dem Team diskutiert werden.

Nach erfolgreicher iterativer Validierung in einer Testumgebung werden Maßnahmen zur automatisierten Korrektur der identifizierten kritischen Prozessparameter abgeleitet. Dies kann z. B. die automatische Anpassung der Parameter der Schweißquelle oder automatische Fehlalarme zum frühzeitigen Ausschluss des fehlerhaften Teils umfassen. Die Integration in die Produktion wird in Zusammenarbeit mit dem Team in die bestehende IT-Architektur pilotiert und kontinuierlich auf Wirksamkeit und Verbesserungsmöglichkeiten hin überwacht.



Abbildung 6-5: KI-gestützter Regelkreis zur prädiktiven Prozesslenkung

Umgang mit Änderungen	Interpretation und Bewertung des KI-Ergebnisses
<ul style="list-style-type: none"> • Festlegen von Leitplanken: Diese Leitplanken können anwendungsspezifische Schwellenwerte wie Vorhersageintervalle, Klassifizierungsmetriken oder Toleranzfelder für die statistische Variabilität der Eingabevariablen sein. Eine Über- oder Unterschreitung dieser Schwellenwerte kann eine Änderung nach den Änderungsklassen A, B oder C auslösen. 	<ul style="list-style-type: none"> • Überwachung der Erkennungs- und Falschalarmlraten (Prüfung False-Positives). • Überwachung der Über- oder Unterschreitung definierter Leitplanken und Metriken. • Für jede Vorhersage können XAI-Methoden verwendet werden, um den Einfluss eines bestimmten Parameters

- **Änderungsklassen definieren:** A: Retraining, Aktualisierung des Modells, B: Anpassung der Features und/oder der Schwellenwerte; C: Anpassung Modellarchitektur oder Umfang der Anwendung.
- **Änderungsklasse A (Retraining, Aktualisierung des Modells):** Retraining des Modells mit neuen Daten. Nach Dokumentation können diese Anpassungen ohne neue Freigabe direkt angewendet werden.
- **Änderungsklasse B (Anpassung der Features und/oder der Schwellenwerte):** Hinzufügen oder Änderungen von Features für das Modell (z. B. Luftfeuchtigkeit). Nach der Dokumentation, dem Testen der Verbesserung in einer Testumgebung und einem Fachreview können diese Anpassungen freigegeben werden.
- **Änderungsklasse C (Anpassung Modellarchitektur oder Umfang der Anwendung):** Änderung der Modellklasse, falls nicht vorher in der Entwicklung bereits getestet (z. B. von neuronalen

auf die Vorhersage anzuzeigen.

- Plausibilitäts- und Konsistenzprüfung der Ergebnisse. Abweichungen werden von Expert:innen geprüft, um die Klassifizierung in A, B oder C gegebenenfalls anzupassen.
- Dokumentation manueller Stichproben und Abgleiche mit Messergebnissen.

Netzwerken auf Random Forests) oder Änderung des Umfangs (z. B. Vorhersage anderer nachgelagerter Prozessgrößen). Nach der Dokumentation, dem Testen des neuen Modells oder der Anwendung in einer Testumgebung und einem umfänglicheren Fachreview können diese Anpassungen freigegeben werden.

- **Dokumentation:**
Versionierung der Daten und Modelle, Änderungen gemäß Änderungsklassen dokumentieren: Speicherung der durchgeführten Tests und Hyperparameter des Modells, Dokumentation der Änderungen an Features und Schwellenwerten sowie Dokumentation der eingesetzten Algorithmen und Vergleiche.
- **Datenüberwachung:**
Regelmäßig prüfen, ob sich Datenumfang oder Datenqualität deutlich verändert (z. B. fehlende Werte, neue Wertebereiche). Kalibrierung der Sensoren und Überwachung der Datenqualität. Bei auffälligen Änderungen Ursache klären.

6.6 Vorbeugende Instandhaltung

Daten aus Anlagen, Maschinen, Prozessen und anderen Quellen, z. B. technischen Spezifikationen, können zur Entwicklung von Datenmodellen für die Erkennung von Anomalien und die Vorhersage möglicher Störungen und Stillstände in der Produktion verwendet werden. Auf diese Weise können vorausschauende Maßnahmen zur Erhöhung der Anlagen- und Maschinenverfügbarkeit sowie Ursachenanalysen für Gewährleistungsansprüche abgeleitet werden.

Beschreibung

Vorbeugende Instandhaltung ist eine Instandhaltungsstrategie, bei der der Zustand von Anlagen und Maschinen kontinuierlich überwacht und auf der Grundlage von Vorhersagen rechtzeitig Wartungsmaßnahmen durchgeführt werden. Die hierfür verwendeten Datenmodelle basieren auf historischen Daten und können verschiedene Datenquellen und Sensordaten wie Temperatur, Vibration und Druck umfassen. Mit Hilfe von statistischen und KI-basierten Methoden können Daten analysiert werden, um Trends zu erkennen und potenzielle Ausfälle vorherzusagen. Ein Mehrwert ist die Planung von Instandhaltungsmaßnahmen entsprechend dem tatsächlichen Zustand der Anlagen und Maschinen sowie die Reduzierung der damit verbundenen Kosten und Ausfallzeiten.

Rahmenbedingungen

- **Definition Anwendungsfall:** Ziele, Metriken (z. B. Ausschussquote, Genauigkeit, ROI), Geltungsbereich, Umfang, Anforderungen, Team und Fachbereiche.
- **Spezifikationen Anlagenhersteller:** Berücksichtigung der Herstellerspezifikationen und Datenschnittstellen sowie gegebenenfalls Einbeziehung von Herstellern hinsichtlich bestehender Modelle und Daten.
- **Expertenwissen und Schulungen:** Einbeziehung von Expert:innen zur Bewertung der Daten- und Prognosequalität und zur Überprüfung von Annahmen, der Plausibilität der Modellergebnisse

und der abgeleiteten Erkenntnisse. Schulung der relevanten Stakeholder hinsichtlich der Entwicklung, Nutzung und Wartung der Lösung.

- **Datenverfügbarkeit und -qualität:** Vollständigkeit, Korrektheit, Konsistenz, Aktualität und Relevanz der Daten (z. B. Maschinenparameter und Sensormessungen). Die Daten müssen in ausreichender Menge für die Modellierung verfügbar sein.
- **Automatisierte Erfassung von Messdaten:** Die Grundlage für eine Anomalie- oder Ausfallvorhersage ist die korrekte Erfassung von Sensordaten, die präzise Vorhersagen und entsprechende Maßnahmen ermöglicht.
- **Datenintegration:** Integration und Aggregation von Daten aus unterschiedlichen Quellen für die Analyse und Modellierung.
- **Standardisierung und Datenmanagement:** Einheitlichkeit der Formate zu Erfassung, Austausch und Management von Daten. Kontinuierliche Pflege von Daten zur Aufrechterhaltung und/oder Verbesserung der Datenqualität.
- **Schwellwert-Logik:** Kombination von KI-Vorhersagen mit festen Schwellwerten und Bereitstellung verständlicher Begründungen durch die KI (z. B. „Vibration > Grenzwert“).
- **Aufsicht und Prüfung:** Zwingende Prüfung innerhalb definierter Fristen bei Sicherheitsrelevanz.
- **Validierung:** Implementierung von Pilotprojekten, um die Effektivität der Lösungen in kontrollierten Umgebungen zu testen.
- **Governance und Datenschutz:** Festlegung der Verantwortlichkeiten und Einhaltung bestehender Standards und Regularien.
- **Infrastruktur und technische Schnittstellen:** Geeignete Hardware und Software für die Lösungsentwicklung und die Sicherstellung der technischen Integration in bestehende Systeme.

Mehrwert	Herausforderungen
<ul style="list-style-type: none"> • Verbesserung der Produkt- und Prozessqualität: Vorausschauende Einleitung von Verbesserungsmaßnahmen auf Grundlage von Daten. • Verbessertes Maschinen- und Anlagenverhalten: Mögliche Erkenntnisse über Zusammenhänge zwischen Maschinenparametern und Lebensdauer. • Ressourceneffizienz: Erhöhung der Maschinen- und Anlagenverfügbarkeit, Maximierung der Lebensdauer und Maschinenleistung. • Kostenreduzierung: Reduzierung von Ausfallzeiten, Störungen und Optimierung der Wartungsarbeiten. 	<ul style="list-style-type: none"> • Datenqualität: Prozessveränderungen beeinflussen die Datenqualität, z. B. Kalibrierung und Verschleiß Sensoren. • Unausgewogene Datensätze: Datensätze können hinsichtlich der Ausgewogenheit zwischen zu klassifizierenden oder vorherzusagenden Zuständen verzerrt sein, z. B. bei Defekten, die sich auf die Unsicherheit der Modelle auswirken könnten. • Produkt- und Prozesskomplexität: Komplexe, nichtlineare Zusammenhänge zwischen Maschinenparametern und Maschinenzuständen. • Nachvollziehbarkeit und Akzeptanz: Geringe Akzeptanz bei Fachabteilungen aufgrund fehlender Erklärbarkeit der Modelle („Blackbox“-Effekt). • Technische Integration und Schnittstellen: Integration in bestehende IT- und Produktionssysteme (z. B. MES).

Vorgehensweise

- **Definition Anwendungsfall:** Ziele, Metriken (z. B. Falschalarmrate, Frühwarnzeit, ROI), Geltungsbereich, Umfang, Anforderungen, Team und Fachbereiche.
- **Ist-Aufnahme:** Analyse der betroffenen Anlagen, Maschinen und Prozesse sowie initiale Analyse und Bewertung der vorhandenen Daten.
- **Datenvorbereitung und -modellierung:** Aufbereitung und Analyse der Daten, ggf. Datenanforderungen anpassen oder erweitern und entsprechende Daten erheben und anschließende Datenmodellierung und Training des prädiktiven Modells in Testumgebung.
- **Erprobung und Validierung:** Implementierung der Lösung in der Produktion und Ableitung von Verbesserungsmaßnahmen. Iterative Verbesserung des trainierten Modells mit Vergleich der definierten Metriken.

QM-Rollen

Mitarbeiter:innen in der Produktionsqualität

6.6.1 Beispiel

Ein interdisziplinäres Team (z. B. Produktionsleitung, Produktionsqualität, Prozessentwicklung, Technologieexpert:innen, IT) arbeitet daran, die Anlagenstillstände für Reparatur und Wartung zu minimieren, z. B. bei Montagerobotern (Karosseriebau oder Endmontage Fahrzeug) oder Fräsprozessen (z. B. Getriebeteile oder Motorenkomponenten). Durch die Vorhersage kritischer Belastungs- und Leistungspunkte (Anomalieerkennung) sollen Ausfallzeiten reduziert, Wartungsarbeiten optimiert und letztlich die damit verbundenen Kosten gesenkt werden. Beispiele hierfür sind die Vorhersage von Ausfällen oder Positionsabweichungen im Laufe der Zeit bei Gelenkverschleiß (Montageroboter) und vorzeitigem Verschleiß oder Werkzeugbruch (Fräsprozess).

Für diese Anwendungen wird zunächst das Potenzial bewertet (z. B. Erfassung der Anzahl ungeplanter Ausfälle, Ausfallzeiten und damit

verbundenen Kosten). Auf der Grundlage der technischen Spezifikationen des Anlagenherstellers wird das Potenzial ermittelt und die Ziele und Anforderungen definiert. Daten aus Qualitätsprüfungen (z. B. Toleranzmessungen und -abweichungen), Umgebungsdaten (z. B. Vibrationen, Temperatur) und relevante steuerbare Prozessparameter der Anlage (z. B. je nach Technologie Drehmomente, Geschwindigkeiten) sowie technische Spezifikationen des Herstellers werden benötigt.

Die erforderlichen Daten werden aus Qualitätsprüfungen, Sensordaten und Betriebsdaten gesammelt und können zur Entwicklung von Testanwendungen separat gespeichert werden. Während der Datenaufbereitung werden erste Analysen der Datenverteilungen und Ausreißer durchgeführt. Die Qualität der Daten wird bewertet (z. B. Vollständigkeit, Interpretierbarkeit, Relevanz). Mit Hilfe von Expert:innen werden die Daten bereinigt und für die Modellierung aufbereitet (je nach Datenformat kann dies einen aufwendigen Schritt darstellen).

Die Daten werden gekennzeichnet (z. B. Werte in Zeitreihen außerhalb definierter Betriebsgrenzen), Schwellenwerte und Metriken (z. B. Prädiktionsintervalle, Abweichungsmetriken) für die KI-Modelle definiert und bei Bedarf zusätzliche Features aus den vorhandenen Daten entwickelt, um den bestehenden Prozess und Ziele der Anwendungsfälle besser abzubilden. Verschiedene Algorithmen werden ausgewählt (z. B. neuronale Netzwerke, Random Forests), um die Modelle zu trainieren und die Vorhersagegenauigkeit (Unsicherheit) zu bewerten.

Die resultierenden Modelle werden durch Testen verschiedener Hyperparameter (Modellparameter) optimiert und anhand zusätzlicher historischer Daten getestet. Mithilfe von Sensitivitätsanalysen oder XAI-Methoden können die Einflüsse verschiedener Parameter im Modell auf die Vorhersageergebnisse (z. B. Parametergrenzen für eine Vorhersage außerhalb der definierten Schwellenwerte) ermittelt und mit dem Team diskutiert werden.

Nach erfolgreicher iterativer Validierung in einer Testumgebung werden Maßnahmen zur automatisierten Korrektur der identifizierten kritischen Prozessparameter abgeleitet. Dies kann z. B. die automatische Erstellung eines Wartungsauftrags bei Erreichen definierter Verschleißgrenzen oder die Anpassung von Prozessparametern wie Drehmoment oder

Vorschubgeschwindigkeit mit einem Warnsignal umfassen. Die Integration in die Produktion wird in Zusammenarbeit mit dem Team in die bestehende IT-Architektur pilotiert und kontinuierlich auf Wirksamkeit und Verbesserungsmöglichkeiten hin überwacht.



Abbildung 6-6: Prozesskette der KI-gestützten vorbeugenden Instandhaltung

Umgang mit Änderungen	Interpretation und Bewertung des KI-Ergebnisses
<ul style="list-style-type: none"> • Festlegen von Leitplanken: Diese Leitplanken können anwendungsspezifische Schwellenwerte wie Vorhersageintervalle, Betriebsgrenzen oder Toleranzfelder für die statistische Variabilität der Eingabevariablen sein. Eine Über- oder Unterschreitung dieser Schwellenwerte kann eine Änderung nach den 	<ul style="list-style-type: none"> • Überwachung der Erkennungs- und Falschalarmraten (Prüfung False-Positives). • Überwachung der Über- oder Unterschreitung definierter Leitplanken und Metriken. • Für jede Vorhersage können XAI-Methoden verwendet werden, um den Einfluss eines bestimmten

Änderungsklassen A, B oder C auslösen.

- **Änderungsklassen definieren:** A: Retraining, Aktualisierung des Modells, B: Anpassung der Features und/oder der Schwellenwerte; C: Anpassung Modellarchitektur oder Umfang der Anwendung.
- **Änderungsklasse A (Retraining, Aktualisierung des Modells):** Retraining des Modells mit neuen Daten. Nach Dokumentation können diese Anpassungen ohne neue Freigabe direkt angewendet werden.
- **Änderungsklasse B (Anpassung der Features und/oder der Schwellenwerte):** Hinzufügen oder Änderungen von Features für das Modell (z. B. Luftfeuchtigkeit). Nach der Dokumentation, dem Testen der Verbesserung in einer Testumgebung und einem Fachreview können diese Anpassungen freigegeben werden.
- **Änderungsklasse C (Anpassung Modellarchitektur oder Umfang der**

Parameters auf die Vorhersage anzuzeigen.

- Plausibilitäts- und Konsistenzprüfung der Ergebnisse. Abweichungen werden von Expert:innen geprüft, um Klassifizierung in A, B oder C gegebenenfalls anzupassen.
- Pflicht zur menschlichen Bestätigung bei Sicherheitsfällen.
- Dokumentation jeder Wartungsentscheidung.

Anwendung): Änderung der Modellklasse, falls nicht vorher in der Entwicklung bereits getestet (z. B. von neuronalen Netzwerken auf Random Forests) oder Änderung des Umfangs (z. B. Vorhersage anderer nachgelagerter Prozessgrößen). Nach der Dokumentation, dem Testen des neuen Modells oder der Anwendung in einer Testumgebung und einem umfänglicheren Fachreview können diese Anpassungen freigegeben werden.

- **Dokumentation:**
Versionierung der Daten und Modelle, Änderungen gemäß Änderungsklassen dokumentieren:
Speicherung der durchgeführten Tests und Hyperparameter des Modells, Dokumentation der Änderungen an Features und Schwellenwerten sowie Dokumentation der eingesetzten Algorithmen und Vergleiche.

- **Datenüberwachung:**
Regelmäßig prüfen, ob sich Datenumfang oder Datenqualität deutlich verändern (z. B. fehlende Werte, neue Wertebereiche). Kalibrierung der Sensoren und Überwachung der Datenqualität. Bei auffälligen Änderungen Ursache klären.

6.7 Felddatenanalyse

Daten aus der Nutzungsphase und anderen Quellen, z. B. Gewährleistungsansprüche, Kundenfeedback, Telemetriedaten sowie Daten aus Diagnosegeräten, können zur Entwicklung von Datenmodellen für die Überwachung der Produktqualität im Feld verwendet werden. Die in den Daten erkannten Muster ermöglichen die Früherkennung von Fehlertrends, die Analyse von Nutzungsprofilen und Ursachenanalysen bei Reklamationen. Auf diese Weise können Maßnahmen zur Reduzierung von Gewährleistungskosten sowie zur Verbesserung aktueller und zukünftiger Produktgenerationen abgeleitet werden.

Beschreibung

Die Felddatenanalyse ist eine Strategie zur Überwachung der Produktqualität in der Nutzungsphase, bei der Daten aus dem realen Kundenbetrieb kontinuierlich ausgewertet werden. Die hierfür verwendeten Datenmodelle basieren auf einer Kombination aus strukturierten Daten (z. B. Telemetrie, Logs aus Diagnosegeräten) und unstrukturierten Informationen (z. B. Schadensbeschreibungen, Kundenfeedback). Mit Hilfe von KI-basierten Methoden wie Natural Language Processing (NLP) und Anomalie-Erkennung können diese heterogenen Datenquellen

analysiert werden, um Fehlermuster zu identifizieren und Ausfallursachen zu korrelieren.

Rahmenbedingungen

- **Definition Anwendungsfall:** Ziele, Metriken (z. B. Genauigkeit, ROI), Geltungsbereich, Umfang, Anforderungen, Team und Fachbereiche.
- **Definition Entscheidungsgrundlage:** Einsatz transparenter Modelle zur Begründung, Einordnung in Kritikalitätsstufen mittels fester Regeln und Entscheidung durch Fachgremium bei hoher Kritikalität.
- **Datenverfügbarkeit und -qualität:** Große, strukturierte Datenmengen in der Echtzeitübermittlung aus Diagnosegeräten ermöglichen die Anwendung von Datenmodellen zur proaktiven Analyse. Geringere Datenqualität hinsichtlich Konsistenz, Aktualität und Vollständigkeit sowie reaktive Maßnahmen für unstrukturierte Daten, wie Kundenfeedback oder Fehlerbeschreibungen.
- **Datenintegration:** Integration und Aggregation von heterogenen Datenformaten (z. B. unstrukturiertes Kundenfeedback und strukturierte Logs aus Diagnosegeräten) aus unterschiedlichen Quellen für die Ursachenanalyse.
- **Expertenwissen:** Komplexe Wirkzusammenhänge in den erhobenen Felddaten erfordern einen hohen Grad an Fachexpertise, um potenzielle Fehlerursachen abzuleiten, das Nutzungsverhalten zu interpretieren oder Produktoptimierungspotenziale zu identifizieren.
- **Datenschutz und Compliance:** Erfüllung rechtlicher Anforderungen und Wahrung des Vertrauens der Kunden hinsichtlich der Aufbewahrung und Verarbeitung personenbezogener Daten.
- **Infrastruktur und technische Schnittstellen:** Berücksichtigung der Anforderungen hinsichtlich Echtzeitübertragung und hoher Datenverarbeitungskapazitäten an die IT-Infrastruktur.

Mehrwert	Herausforderungen
<ul style="list-style-type: none"> • Proaktive Produktoptimierung: Kontinuierliches Überwachen von Trends und Mustern ermöglicht die Einleitung gezielter Innovationen und Verbesserungsmaßnahmen. • Präventive Fehlerbehandlung: Einleitung proaktiver Maßnahmen, um dem Auftreten von Funktionsfehlern vorzubeugen. • Analyse des Nutzungsverhaltens: Ableitung von Erkenntnissen über das tatsächliche Nutzungsverhalten. • Reduktion von Garantie-/ Servicekosten: Präventive Maßnahmeneinleitung ermöglicht Fehlervorbeugung. • Schnellere Ursachenanalyse: Die Felddatenanalyse unterstützt eine schnellere Identifikation von Fehlerursachen. • Erhöhte Kundenzufriedenheit: Reaktion auf explizite Kundenbedürfnisse. 	<ul style="list-style-type: none"> • Datenqualität: Geringe Qualität unstrukturierter Felddaten, z. B. Kundenfeedback • Datenintegration: Integration heterogener Datenformate aus unterschiedlichen Quellen. • Datenanalyse: Komplexe Wirkzusammenhänge erfordern Expertenwissen (z. B. Unterscheidung von Korrelation und Kausalität). • Aufwand: Der Nutzen der Felddatenanalyse muss größer sein als der dafür notwendige Datenbeschaffungs- und -analyseaufwand.

Vorgehensweise

- **Definition Anwendungsfall:** Ziele, Metriken (z. B. Genauigkeit, ROI), Geltungsbereich, Umfang, Anforderungen, Team und Fachbereiche.
- **Datenbeschaffung:** Bestimmung der Datenquellen, Datenübertragung und -speicherung. Bestimmung der Datenformate bzw. Definition von Datenstandards.
- **Datenvorbereitung und -modellierung:** Definition der Anforderungen an die Datenqualität und Ableitung von Methoden zur Datenqualitätsverbesserung. Ableitung eines Datenmodells zur Darstellung der Wirkzusammenhänge.
- **Datenanalyse und Validierung:** Regelbasierte oder KI-gestützte Analyse der Felddatenmuster und Ableitung bspw. von Fehlermustern, Fehlerursachen oder Mustern im Nutzungsverhalten. Validierung der identifizierten Muster.
- **Ableitung von Maßnahmen:** Ableitung und Einleitung proaktiver oder reaktiver Maßnahmen zur Prävention/Behebung von Fehlern im Produkt in der Nutzungsphase. Ableitung von Potenzialen für die Produktoptimierung.

QM-Rollen

Mitarbeiter:innen in der Kundenqualität

6.7.1 Beispiel

Ein Qualitätsteam wertet wöchentlich Feldausfälle aus, die über verschiedene Kundenportale gemeldet werden (z. B. Garantie-/Kulanzmeldungen, Werkstattbefunde, Rückläuferhinweise). Diese Meldungen werden automatisiert in eine zentrale Felddatenbank übernommen und dort vereinheitlicht.

In der Praxis ist es jedoch aufwendig, aufgrund der Vielzahl an Meldungen frühzeitig zu erkennen, welche Produkt- und Einsatzkombinationen tatsächlich auffällig und welche Schwankungen „normal“ sind, etwa durch saisonale Effekte. So treten Ausfälle an Klimakomponenten im Sommer

häufiger auf als im Winter, ohne dass zwingend ein neuer Qualitätsfehler vorliegt.

Statt jede Auswertung manuell aufzubauen, nutzt das Team ein KI-gestütztes Analysesystem, das einmal pro Woche als Batchlauf startet. Das System bildet automatisch alle relevanten Kombinationen wie z. B. aus Kunde, Produkt, Fahrzeugmodell und Produktionswerk und erzeugt für jede Kombination eine Zeitreihe, z. B. „Anzahl Feldausfälle pro Kalenderwoche“.

Für jede dieser Zeitreihen wird ein Prognose- und Erwartungsmodell erstellt, das sowohl den langfristigen Trend als auch wiederkehrende saisonale Muster berücksichtigt. Dadurch kann das System unterscheiden, ob ein aktueller Anstieg erwartbar ist (z. B. Sommeranstieg bei Klimaanlagen) oder ob er über das zu erwartende Maß hinausgeht. Zusätzlich prüft es die letzten Wochen gegen den berechneten Erwartungsbereich und markiert Ausreißer.

So kann das System beispielsweise erkennen, dass die Kombination „Kunde A – Klimakompressor Variante 3 – Plattform X – Werk 2“ seit mehreren Wochen deutlich über dem erwarteten Verlauf liegt, obwohl die Saisonalität bereits eingerechnet ist. Gleichzeitig zeigt es, dass andere Kombinationen zwar im Sommer ebenfalls ansteigen, aber innerhalb des erwarteten saisonalen Rahmens bleiben. Für die auffällige Kombination berechnet das System einen Abweichungswert und ordnet sie in einem wöchentlichen Ranking weit oben ein.

Das Qualitätsteam nutzt diese Rangliste als Arbeitsvorrat. Für die Top-Fälle erfolgt zunächst eine fachliche Plausibilisierung (z. B. Datenvollständigkeit aus den Portalen, Meldeverzug, Dubletten, geänderte Fehlercodes). Anschließend werden gezielte Vertiefungen angestoßen, etwa nach Produktionszeitraum, Charge, Softwarestand oder Lieferant. Daraus können konkrete Maßnahmen entstehen. Zum Beispiel eine Eingrenzung auf bestimmte Produktionswochen, eine Anpassung von Prüfmerkmalen, eine Ursachenanalyse mit dem Werk oder eine gezielte Rückläuferuntersuchung.

Am Ende steht eine Felddatenanalyse, die regelmäßig und reproduzierbar läuft, saisonale Effekte methodisch berücksichtigt und die Aufmerksamkeit des Teams auf die Kombinationen lenkt, bei denen ein erhöhtes zukünftiges Ausfallrisiko am wahrscheinlichsten ist. Die Bewertung und Entscheidung

über Maßnahmen bleibt dabei vollständig beim Fachteam. Das System unterstützt durch strukturierte Datenaufbereitung, Trendermittlung und Priorisierung.

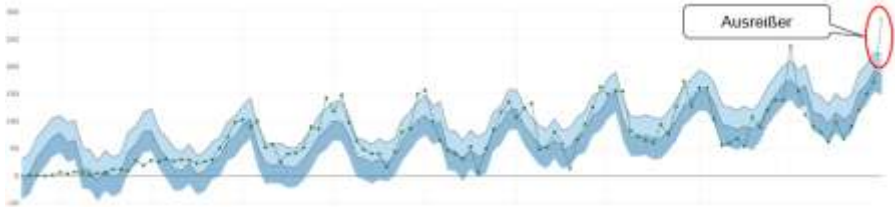


Abbildung 6-7: Saisonale Ausreißer-Erkennung in der Felddatenanalyse – Ausfallzeitreihe einer Klimaanlage mit KI-Erwartungsbereich und markiertem Ausreißer

Umgang mit Änderungen	Interpretation und Bewertung des KI-Ergebnisses
<ul style="list-style-type: none"> • Änderungen an Datenquellen vorab bewerten und versionieren. • Datenüberwachung: Regelmäßig prüfen, ob sich Datenumfang oder Datenqualität deutlich verändert (z. B. fehlende Werte, neue Wertebereiche). Bei auffälligen Änderungen Ursache klären. • Label-Änderungen (z. B. neue Fehlercodes, geänderte Werkstattcodierung) als Change 	<ul style="list-style-type: none"> • Plausibilitäts- und Konsistenzprüfung: Stimmen Trends/Anomalien mit bekannten Ereignissen überein? • Ergebnis als Hinweis verstehen: KI-Ergebnisse sind ein Hinweis, kein Beweis. Vor Entscheidungen zusätzliche Informationen heranziehen (z. B. Werkstattberichte, Teileprüfung, Reklamationen). • Sicherheit durch Gegenprüfung:

<p>erfassen; ReLabeling-Regeln dokumentieren.</p> <ul style="list-style-type: none"> • Rückverfolgung sicherstellen: Jede Analyse muss reproduzierbar sein (Datenstand, Filter, Modellversion, Schwellenwerte). 	<p>Stichproben prüfen und mit bekannten Fällen vergleichen. Falls möglich: Vergleich mit einer einfachen Auswertung (z. B. Häufigkeiten/Trends) zur Absicherung.</p> <ul style="list-style-type: none"> • Klare Entscheidungsklassen verwenden, z. B. „beobachten“, „weiter untersuchen“, „sofort handeln“. Die Entscheidung trifft die verantwortliche Fachstelle, nicht die KI.
--	--

6.8 Review von Entwicklungsarbeitsprodukten

Beschreibung

Während der Produktentwicklung führen die Qualitätsingenieur:innen eine Vielzahl von Reviews von Arbeitsprodukten gegen Qualitätskriterien, Arbeitsanweisungen, Guidelines oder sonstige mitgeltende Unterlagen durch. Diese Aufgabe wird von einem KI-System unterstützt und Abweichungen werden aufgezeigt. Dabei kann die KI ggf. auch nur bestimmte Aspekte prüfen, wenn es andere Aspekte gibt, die nicht genau genug durch die KI geprüft werden können.

Rahmenbedingungen

- Entwicklungsarbeitsprodukte Qualitätskriterien, Arbeitsanweisungen, Guidelines und sonstige mitgeltende Unterlagen liegen in maschinenlesbaren Formaten vor
- Qualitätskriterien, Arbeitsanweisungen, Guidelines und sonstige mitgeltende Unterlagen werden in dem KI-System hinterlegt, um Abweichungen fundiert zu bewerten.

- Großdokumente werden durch automatisiertes Chunking und Parsing in stabile Abschnitte/Klauseln überführt.
- Hybrider Ansatz: Deterministische Textsegmentierung kombiniert mit KI-Inhaltsanalyse.
- Regelbasierte Bewertung: Einstufung des Arbeitsprodukts als freigabefähig erfolgt nach klaren Kriterien anhand von Häufigkeit und Schwere der gefundenen Abweichungen.

Mehrwert

- Abweichungen werden aufgezeigt und ggf. klassifiziert.
- Effizienz- und Zeitgewinne durch KI-gestützten Review.
- Alle gefundenen Abweichungen erscheinen in einer klar gegliederten Übersicht und können somit als Basis für ein Lessons Learned dienen.

Herausforderungen

- Zuverlässige Erkennung kontextueller Abweichungen ist anspruchsvoll; Fehlklassifikationen sind möglich.
- Unstrukturierte Texte, Scans, Tabellen und uneinheitliche Formatierungen verschlechtern Segmentierung und damit die Vergleichsqualität.
- Sehr große Dokumente belasten die semantische Analyse; abschnittsweises Vorgehen (Chunking) verbessert die Ergebnisse.
- Menschliche Validierung/Prüfung notwendig.

Vorgehensweise

- Vergleichsdokumente (z. B. Guidelines, mitgeltende Unterlagen) festlegen.
- Strukturierte Vorverarbeitung: Parsing, Klassifizierung, Gliederung in Abschnitte/Klauseln.

- Erzeugung einer gegliederten Abweichungsübersicht mit Bewertung und Empfehlungen.
- Review des Ergebnisses durch die Fachabteilung/Qualität.

QM-Rollen

Mitarbeiter:innen in der Entwicklungsqualität, Qualitäts-Auditor:in, Qualitäts-Assessor:in

6.8.1 Beispiel

Review der SW-Anforderungen während der Entwicklung eines Steuergeräts gegen Qualitätskriterien.

Ein komplexes Steuergerät hat eine Vielzahl von Anforderungen an die Software. Diese sind gegen Qualitätskriterien (z. B. Rückverfolgbarkeit, Formulierung, Konsistenz, Testbarkeit) zu prüfen. Hierbei unterstützt ein KI-System und meldet Abweichungen zu den im Projekt festgelegten Qualitätskriterien.

Im Vorfeld sind die Anforderungsdokumente in von der KI handhabbare Chunks zu zerlegen, die aber noch die z. B. für die Prüfung der Konsistenz nötigen Zusammenhänge widerspiegeln.

6.8.2 Beispiel

Review der Implementierung einer Steuergeräte-SW gegen Erratas (Liste bekannter Fehler) von Zulieferer-Software.

Moderne Steuergerätesoftware integriert viele Komponenten von Zulieferern (z. B. Autosar-Stack) und benutzt zugelieferte Werkzeuge (z. B. Compiler). Die Hersteller dieser Komponenten veröffentlichen regelmäßig teils umfangreiche Errata-Sammlungen, gegen die die Implementierung der Steuergerätesoftware zu prüfen ist.

Hierzu werden nun die entsprechenden Erratas vorverarbeitet und in das KI-System integriert. Die fertig integrierte SW mit ihren relevanten Quellen wird nun gegen diese implementierten Erratas geprüft und gibt zurück, ob diese ggf. bei der Implementierung nicht berücksichtigt wurden. Dabei ist

auch dies in von der KI verarbeitbare Chunks zu zerlegen, wobei zu beachten ist, dass diese noch den erforderlichen Kontext bereitstellen.

Umgang mit Änderungen	Interpretation und Bewertung des KI-Ergebnisses
<ul style="list-style-type: none"> • Änderungsklassen definieren: Unterteilung der Anpassungen am KI-System in Klassen (z. B. A: rein sprachlich/ kosmetisch, B: fachlich relevant, aber im gleichen Rahmen, C: grundlegend verhaltensändernd). Klasse C löst eine neue Freigabe aus. • Kleinere Änderungen (Klasse A): Neue oder geänderte Qualitätskriterien, Arbeitsanweisungen, Guidelines und sonstige mitgeltende Unterlagen können eingepflegt werden, wenn diese in Umfang und Format nicht wesentlich von den ursprünglichen abweichen. Nach einem Vier-Augen-Review und Dokumentation können diese Anpassungen ohne neue Freigabe aktiviert werden. • Fachlich relevante Änderungen (Klasse B) gezielt testen: 	<ul style="list-style-type: none"> • Messung der Übereinstimmungsquote zwischen KI-Vorschlag und Fach-Review. • Versionierung der Vergleichsberichte mit Freigabevermerk. • Nutzung von Referenzdokumenten mit bekannten Änderungen zur Validierung. • Nachvollziehbarkeit der KI-Vorschläge sicherstellen: Zu jeder gemeldeten Abweichung liefert die KI eine Begründung mit Bezug auf das konkrete Qualitätskriterium (z. B. „Anforderung nicht testbar – fehlende Mess- oder Akzeptanzbedingung“). Damit ist nachvollziehbar, auf welcher Regelgrundlage die Abweichung beruht. • Messung der Übereinstimmungsquote: Die Übereinstimmung

Erweiterungen, die das Verhalten präzisieren, aber die Rolle der KI nicht grundlegend verändern – z. B. neue Qualitätskriterien-Kategorien, verfeinerte Chunking-Regeln, zusätzliche Referenzdokumente – werden mit einem repräsentativen Satz von Test-Anforderungsdokumenten geprüft. Fachleute beurteilen, ob Klassifikation und Abweichungsmeldungen weiterhin sinnvoll und konsistent sind.

- **Grundlegende Änderungen (Klasse C)** als neue Systemversion behandeln: Bei neuen Dokumentformaten (z. B. neue Strukturierungskonventionen, bisher nicht unterstützte Dateiformate) ist ein vollständiger Regressionstest durchzuführen. Anpassungen, die das Risikoprofil des Systems verändern (z. B. automatische Freigabe von Arbeitsprodukten durch die KI), erfordern eine

zwischen KI-Vorschlag und dem Ergebnis des manuellen Fach-Reviews wird gemessen und dokumentiert. Abweichungen zwischen KI-Einschätzung und menschlicher Bewertung werden analysiert und zur kontinuierlichen Verbesserung des Systems genutzt.

- **Versionierung der Vergleichsberichte mit Freigabevermerk:** Jeder erzeugte Abweichungsbericht wird mit Zeitstempel, verwendeter Systemversion und Freigabevermerk der verantwortlichen Qualitätsingenieurin oder des verantwortlichen Qualitätsingenieurs gespeichert. So ist jederzeit nachvollziehbar, welche Version der Qualitätskriterien und des KI-Modells für die Bewertung verwendet wurde.
- **Validierung mittels Referenzdokumenten:** Zur Qualitätssicherung des KI-Systems werden Referenzanforderungsdokumente mit bekannten,

<p>vollständige Neubewertung und Freigabe.</p> <ul style="list-style-type: none"> • Festlegen von Leitplanken: Typische Leitplanken sind z. B.: Die KI gibt nur Vorschläge, keine finalen Freigabeentscheidungen; jede gemeldete Abweichung enthält einen Bezug zum verletzten Qualitätskriterium; die Freigabe des Arbeitsprodukts erfolgt ausschließlich durch die verantwortliche Qualitätsingenieurin oder den verantwortlichen Qualitätsingenieur. Bleiben diese Leitplanken unverändert, benötigt die Anwendung keine neue Freigabe. 	<p>vorab definierten Abweichungen verwendet. Die KI-Ergebnisse werden gegen diese Erwartungswerte geprüft (z. B. Anzahl erkannter Abweichungen, Falsch-Positive, nicht erkannte Abweichungen).</p> <ul style="list-style-type: none"> • Menschliche Prüfung und Freigabe zwingend erforderlich: Die Qualitätsingenieurin oder der Qualitätsingenieur sieht alle KI-Vorschläge inklusive Klassifikation und Kriterien-Bezug. Er entscheidet, welche gemeldeten Abweichungen bestätigt, angepasst oder als Falsch-Positiv verworfen werden. Die finale Einstufung des Arbeitsprodukts als freigabefähig obliegt ausschließlich dem Menschen.
--	--

6.8.3 Anmerkungen

Der beschriebene Anwendungsfall lässt sich ebenfalls für andere Arbeitsprodukte applizieren.

6.9 VDA-Chatbot

Beschreibung:

KI-gestützter Chatbot zur schnellen, qualifizierten Beantwortung von Fragen rund um die VDA-Regelwerke, z. B. Methoden, hier u. a. 8D, FMEA und Audit-Prüfprozesse, und optional Unternehmens-FAQ. Der Bot liefert geprüfte, qualifizierte Antworten mit Quellenhinweisen (Band/Kapitel/Abschnitt) und, wo sinnvoll, Schritt-für-Schritt-Anleitungen. Die Wissensbereitstellung erfolgt über RAG, eine freie Internetsuche wird unterbunden.

Rahmenbedingungen

- Wissensdatenbank vorhanden (relevante VDA-Bände + Unternehmens-FAQ + verwandte VDA-/Unternehmensstandards).
- Erstellung eines Fragenkatalogs für den VDA-Chatbot zur Verbesserung der Antwortqualität.
- Integration von User-Feedback (z. B. Daumen hoch/runter) zur kontinuierlichen Qualitätsverbesserung.
- Festlegung von Verantwortlichkeiten, Zielgruppen und mehrstufigen Zugriffen.
- Änderungsmanagement für Inhalte und Datenqualität.
- Möglichkeit zur Risikominimierung: Umstellung auf regelbasierte Antworten (Bot darf nur wörtliche Auszüge und geprüfte Zusammenfassungen aus freigegebenen VDA-Abschnitten anzeigen; keine freie Formulierung darüber hinaus).
- Quellenpflicht (RAG): Antworten dürfen nur generiert werden, wenn eine valide Quelle im VDA-Band gefunden wurde.

Mehrwert

Schneller, konsistenter Zugriff auf VDA-/QM-Wissen; reduzierte Such- und Wartezeiten; standortübergreifend einheitliche, geprüfte Antworten.

Herausforderungen

Sicherstellung der Daten-/Inhaltsqualität; Validierung, damit ausschließlich freigegebene Inhalte ausgegeben werden.

Vorgehensweise

- Aufbau eines Chatbots, der relevante VDA-/QM-Dokumente als RAG-Wissensbasis nutzt. Internetsuche unterbinden.
- Masterprompt dazu erstellen, wie und welches Verhalten der Chatbot bei Antworten haben soll.
- Antwortqualität per Fragenkatalog prüfen; Bewertung durch VDA-/QM-Expert:innen.
- In Abhängigkeit von Antwortqualität Nutzerkreis erweitern
→ Skalieren.

Rollen

Mitarbeiter:innen in der Lieferantenqualität, in der Entwicklungsqualität, in der Produktionsqualität, in der Kundenqualität, im Qualitätsmanagementsystem, Qualitäts-Auditor:in/-Assessor:in, Qualitätsführungskraft

6.9.1 Beispiel

Anwender:in meldet sich in einem Web-Frontend an und stellt eine Frage (z. B. „Welche Anforderungen gelten für VDA 2 bei der Produktionsprozess- und Produktfreigabe?“). Das Backend sucht über eine semantische Vektordatenbank die passenden Textstellen aus dem VDA-Band (Retrieval-Augmented Generation). Diese Passagen werden zusammen mit der Frage an das KI-Modul übergeben, das daraus eine Antwort generiert und einen Quellenverweis mit Kapitel-/Abschnittsnummer ergänzt.

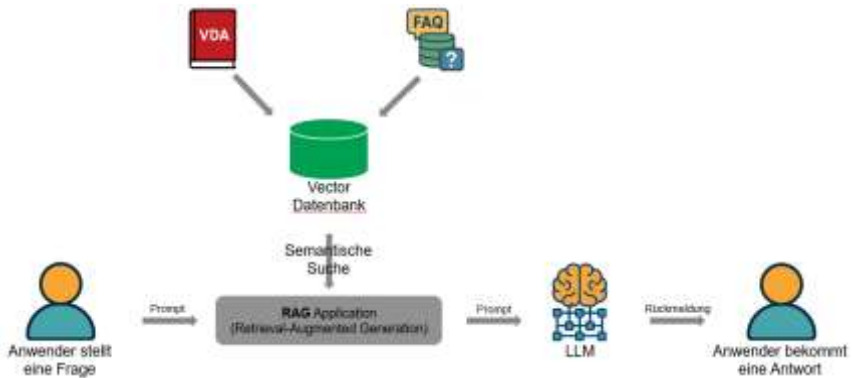


Abbildung 6-8: VDA-Chatbot – RAG-Architektur mit semantischer Vektordatenbanksuche und LLM-gestützter Antwortgenerierung

Beispielhafte Darstellung eines Web-Frontend für den VDA-Chatbot. Sterilisierte Web-Chatbot-Maske mit Login-Header, zentralem Frage-Antwort-Bereich mit verpflichtendem Quellenhinweis sowie Feedback-Funktionen und thematischen Schnellzugriffen.

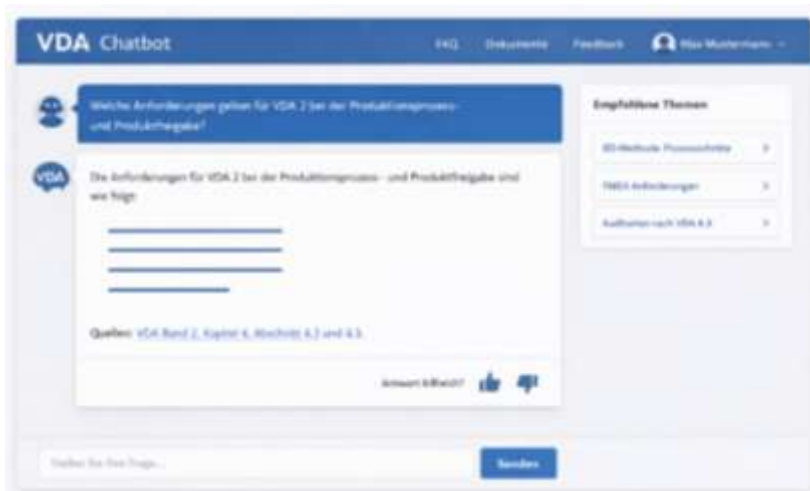


Abbildung 6-9: Beispielhafte Chatbot-Oberfläche mit KI-generierter Antwort und Quellenreferenz

Umgang mit Änderungen	Interpretation und Bewertung des KI-Ergebnisses
<ul style="list-style-type: none"> • Die Konfiguration (erlaubte Quellen, Antwortregeln, Zitierformat) wird laufend gepflegt. • Änderungen durchlaufen vor dem Rollout ein fachliches und technisches Vier-Augen-Review inklusive Test. • Nutzerfeedback wird wöchentlich ausgewertet; kritische Fehler werden umgehend per Hotfix behoben. 	<ul style="list-style-type: none"> • Messung der Abbruchrate (Antworten ohne Quelle). • Protokollierung jeder Interaktion (Frage, Antwort, Quellen-ID, Regelversion). • Regelmäßiger Abgleich mit einem definierten Fragenkatalog mit bekannten korrekten Antworten („Golden Set“).

6.9.2 Anmerkungen

Ein vergleichbares System kann auch die hier vorgeschlagenen Erweiterungen umsetzen. So ließen sich etwa die Inhalte für Normen und Compliance, QM-FAQ-Bot für Lieferanten oder Lessons Learned auf Basis derselben VDA-Chatbot-Architektur realisieren. Durch den modularen Aufbau der Wissensdatenbank und den Einsatz von Retrieval Augmented Generation (RAG) kann das Konzept flexibel auf weitere Qualitätsbereiche übertragen werden.

6.10 Speech Mining für Arbeitsanweisungen

Beschreibung

Mitarbeitende beschreiben während der realen Ausführung ihre Schritte per Sprache. Die Audiodaten werden per Speech-to-Text transkribiert und anschließend via NLP (Natural Language Processing) nach Tätigkeiten, Vorbedingungen, Material/Werkzeugen sowie Sicherheits- und Qualitätsmerkmalen strukturiert. Auf dieser Basis erzeugt die KI eine standardisierte Arbeitsanweisung, die ins Managementsystem überführt und freigegeben werden kann.

Rahmenbedingungen

- Start-Szenario und Prozessgrenzen klar definieren; kritische Schritte vorab klären.
- Rollen festlegen: Prozessverantwortliche/Fachexpert:innen für fachliche Reviews.
- Hardware: Smartphone oder geeignetes Aufnahmegerät; bei Bedarf Headset/Hands-free-Lösung, um Ablenkung bei sicherheitskritischen Tätigkeiten zu vermeiden.
- Datenschutz und Transparenz: Einwilligung zur Sprachaufzeichnung, Zweckbindung, Löschkonzept, Zugriffskontrollen; Vorgaben des Managementsystems beachten.
- Visuelle Ergänzungen (Fotos/Frames) und klare Schrittlisten erhöhen Verständlichkeit und Umsetzbarkeit der Arbeitsanweisung.
- Automatische Ablehnung, wenn Sicherheits- oder Prüfmerkmale fehlen.
- Visuelle Bestätigung: Ergänzung von Fotos/Videos zur Validierung des KI-Textes.

Mehrwert

- Beschleunigte, praxisnahe Prozessdokumentation und Standardisierung direkt aus der Ausführung heraus.

Herausforderungen

- Dialekte oder individuelle Sprechweisen.

- | | |
|---|---|
| <ul style="list-style-type: none"> • Zeitnahe Aktualisierung von Anweisungen bei Prozessänderungen; weniger nachträgliche Interviews/Korrekturen. • Reduktion von nachträglichen Korrekturen. • Kollaborative Erstellung: Mehrere Personen können Beiträge liefern, die konsolidiert werden. • Mehrere Personen können in die Aufnahme einbezogen werden. | <ul style="list-style-type: none"> • Fachterminologie und Konsistenz. • Akustik/Aufnahmesituation (z. B. Produktionslärm, überlappende Sprecher). • Anwender:innen müssen das „sprechende Arbeiten“ erlernen/akzeptieren. Details wie Parameter oder Prüfkriterien dürfen nicht vergessen werden. • Datenschutz: Einwilligung, Zweckbindung, Löschkonzept und Zugriffskontrollen. |
|---|---|

Vorgehensweise

- Datenschutz klären: Rechtsgrundlage, Einwilligungen, Speicher- und Löschregeln festlegen; Betroffene informieren.
- Scope definieren: Konkreten Prozess und Zielformat (Template der Arbeitsanweisung) festlegen.
- Setup bereitstellen: Aufnahmegerät/Smartphone; bei Bedarf Hands-free-Hardware und lärmtaugliches Mikrofon je Sprecher.
- Während der Tätigkeit klar beschreiben: Was? Womit? Warum? Prüfkriterien? Sicherheit?
- Transkription und Strukturierung: Speech-to-Text mit Zeitstempeln.
- Prompts/Extraktionsregeln definieren: Vorgaben für NLP/Generierung festlegen.
- Internetsuche abschalten, damit eingegebene Daten nicht durch externe Informationen beeinflusst werden.
- Im Nachgang muss erzeugter Text von Fachexpert:innen geprüft werden.

- Text in Vorlage einkopieren und entsprechend aufbereiten.
- Text ins Arbeitsanweisungs-Template übernehmen, visuelle Elemente ergänzen.

Rollen

Mitarbeiter:innen in der Lieferantenqualität, in der Entwicklungsqualität, in der Produktionsqualität, im Qualitätsmanagementsystem

6.10.1 Beispiel

Beim Zusammenbau eines Kugelschreibers beschreibt eine erfahrene Mitarbeiterin oder ein erfahrener Mitarbeiter den Arbeitsablauf während der Tätigkeit kontinuierlich per Sprache. Die Aufnahme wird automatisiert transkribiert und anschließend durch NLP (Sprachverarbeitung) analysiert. Die KI erkennt dabei relevante Tätigkeitsblöcke, verwendete Komponenten (z. B. Mine, Gehäuseteile, Feder) sowie typische Qualitätsmerkmale (z. B. sauberer Lauf der Mechanik) und strukturiert diese Informationen gemäß dem unternehmensweiten Template für Arbeitsanweisungen. Zusätzlich kann die KI aus den extrahierten Tätigkeiten automatisch einen grafischen Prozessablauf erzeugen, der den Montageprozess übersichtlich darstellt. Aus der gesprochenen Beschreibung entsteht so eine vollständige, standardisierte Arbeitsanweisung, die anschließend durch die Prozessverantwortlichen fachlich geprüft und in das Managementsystem übernommen wird.



Abbildung 6-10: Beispielhafter KI-generierter Prozessablauf für Arbeitsanweisungen – Kugelschreibermontage, abgeleitet aus gesprochener Mitarbeiterbeschreibung

Umgang mit Änderungen	Interpretation und Bewertung des KI-Ergebnisses
<ul style="list-style-type: none"> • Vorlagen und Extraktionsregeln werden versioniert. • Kleine Anpassungen können nach Vier-Augen-Check sofort aktiv werden. • Datenschutz (Einwilligung, Speicher- und Löschrufen) bleibt aktuell. • Das Aufnahme-Setup wird standardisiert. 	<ul style="list-style-type: none"> • Prüfung der Vollständigkeit von Pflichtfeldern. • Freigabe durch eine fachkundige Person ist zwingend erforderlich (Human-in-the-Loop). • Protokollierung der Versionierung.

6.10.2 Anmerkung

Der beschriebene Ansatz lässt sich auf beliebige manuelle Montageprozesse übertragen und skaliert gut in Umgebungen mit häufigen Variantenwechseln oder kurzen Produktlebenszyklen.

Die Qualität der Ergebnisse hängt stark von Akustik, Sprechdisziplin und Konsistenz der Fachsprache ab. Eine fachliche Validierung bleibt zwingend erforderlich.

6.11 Vergleichen von Dokumenten

Beschreibung

Der Vergleich von Kundenvorschriften, Verträgen oder Normen erfolgt mittels einer hybriden Architektur aus deterministischen Komponenten und KI-gestützten Analyseverfahren. Die Dokumente werden zunächst strukturiert vorverarbeitet (Parsing, Abschnittserkennung, Klausel-Mapping). Anschließend bewertet eine KI die inhaltlichen Unterschiede semantisch und ordnet sie Bedeutungskategorien zu. Es werden keine Wort- oder Zeichenänderungen dargestellt; stattdessen entstehen kontextualisierte Änderungsobjekte mit Kurzbeschreibung, Relevanz, Risiko- und Impact-Bewertung sowie Empfehlungen.

Rahmenbedingungen

- Dokumente liegen in maschinenlesbaren Formaten vor.
- Unternehmensrichtlinien werden in der KI-Anwendung hinterlegt, um Abweichungen und Empfehlungen fundiert zu bewerten.
- Eine feste Vorverarbeitungslogik stellt sicher, dass Segmentierung, Mapping und Reihenfolge reproduzierbar bleiben – reine KI-Agenten reichen dafür nicht aus.
- Großdokumente werden durch automatisiertes Chunking und Parsing in stabile Abschnitte/Klauseln überführt.
- Hybrider Ansatz: Deterministische Textsegmentierung kombiniert mit KI-Inhaltsanalyse.
- Regelbasierte Bewertung: Einstufung von Risiko und Verbindlichkeit erfolgt über feste Schlüsselwörter (z. B. „muss“, „darf nicht“).

Mehrwert

- Änderungen werden nach Bedeutung und Kontext

Herausforderungen

- Zuverlässige Erkennung kontextueller Bedeutungsänderungen ist anspruchsvoll;

<p>präsentiert; das beschleunigt die fachliche Prüfung.</p> <ul style="list-style-type: none"> • Automatisierte Risiko- und Impact-Bewertung pro Änderung mit klaren Empfehlungen. • Effizienz- und Zeitgewinne durch KI-gestützten Vergleich und strukturierte Ergebnisaufbereitung. • Alle relevanten Änderungen erscheinen in einer klar gegliederten Übersicht. 	<p>Fehlklassifikationen sind möglich.</p> <ul style="list-style-type: none"> • Unstrukturierte Texte, Scans, Tabellen und uneinheitliche Formatierungen verschlechtern Segmentierung und damit die Vergleichsqualität. • Sehr große Dokumente belasten die semantische Analyse; abschnittsweises Vorgehen (Chunking) verbessert die Ergebnisse. Anzahl der maximalen Tokens beachten. • Menschliche Validierung/Prüfung notwendig.
--	---

Vorgehensweise

- Vergleichsdokumente festlegen.
- Strukturierte Vorverarbeitung: Parsing, Klassifizierung, Gliederung in Abschnitte/Klauseln.
- Semantische Analyse der Änderungen mittels Embedding (Texteinbettungen) und Ähnlichkeitsmetriken.
- Abgleich mit Standards und Richtlinien; Einstufung nach Risikoarten/-stufen und Beschreibung der Auswirkungen.
- Erzeugung einer gegliederten Änderungsübersicht mit Bewertung und Empfehlungen.
- Review des Dokumentenvergleichs durch die Fachabteilung.

QM-Rollen

Mitarbeiter:innen in der Lieferantenqualität, in der Entwicklungsqualität, in der Produktionsqualität, in der Kundenqualität, im Qualitätsmanagementsystem, Qualitäts-Auditor:in, Qualitäts-Assessor:in

6.11.1 Beispiel

Der Prozess beschreibt, wie mithilfe von KI Dokumente inhaltlich verglichen werden (z. B. automatisierter Abgleich zweier Versionen einer Kundenvorschrift, Erkennung von Bedeutungsänderungen und Vorschlag von Risikokategorien).

Durch semantische Segmentierung wird das Lost-in-the-Middle-Problem vermieden – ein Effekt, bei dem große Sprachmodelle Informationen aus der Mitte langer Texte schlechter verarbeiten oder übersehen.

So bleibt der inhaltliche Kontext vollständig erhalten und Änderungen können präzise und nachvollziehbar bewertet werden.

Ablaufschritte:

1. Dokumente hochladen – Bereitstellung und Versionierung der zu vergleichenden Dokumente.
2. Semantische Segmentierung – Aufteilung der Dokumente in logisch zusammenhängende Abschnitte.
3. Semantischer Vergleich – Analyse der inhaltlichen Unterschiede auf Basis von Embeddings (Texteinbettungen) und Ähnlichkeitsmetriken.
4. Interpretation & Bewertung – Generative Zusammenfassung und Bewertung der Änderungen nach Relevanz, Normbezug und Risiko.
5. Konsolidierung & Bericht – Zusammenführung aller Ergebnisse in einem nachvollziehbaren, freigabefähigen Bericht.



Abbildung 6-11: KI-gestützter Dokumentenvergleich – semantische Segmentierung und Differenzanalyse ermöglichen eine präzise Erkennung und Risikobewertung inhaltlicher Änderungen zwischen zwei Dokumentversionen

Durch das segmentbasierte Context-Window-Management kann der Lost-in-the-Middle-Fehler vermieden werden.

Mögliche Darstellungsvariante:

Abschnitt	Thema	Änderung	Bewertung	Risiko	Kommentar / Empfehlung
4.2.3	Dokumentierte Information	„Aufbewahrungsfrist“ wurde zu „Archivierungszeitraum“ geändert	Semantisch gleichwertig	Niedrig	Keine Anpassung erforderlich
5.1.2	Kundenorientierung	Neuer Satz: „Die Organisation muss Kundenfeedback systematisch auswerten“	Neue Anforderung	Hoch	Prüfen, ob QMS-Prozess Q-KUN-02 angepasst werden muss
6.3	Planung von Änderungen	Abschnitt erweitert um Bewertung der Auswirkung auf Lieferanten	Bedeutungsänderung	Mittel	Rückprache mit Einkauf empfohlen
8.2.1	Lenkung fehlerhafter Produkte	Absatz um KI-gestützte Prüfprozesse ergänzt	Neue Technologiebezug	Hoch	Bewertung durch Fachbereich QS erforderlich

Umgang mit Änderungen	Interpretation und Bewertung des KI-Ergebnisses
<ul style="list-style-type: none"> • Regeln für Pflicht- und Kann-Klauseln sowie Risikokategorien werden als Konfiguration gepflegt. • Bei neuen Dokumentformaten führt man Regressionstests durch; Mapping wird angepasst. • Neue Normen werden fachlich geprüft; danach sind schnelle Konfigurationsänderungen möglich. 	<ul style="list-style-type: none"> • Messung der Übereinstimmungsquote zwischen KI-Vorschlag und Fach-Review. Die Übereinstimmungsquoten sind vom Unternehmen selbst zu definieren. • Versionierung der Vergleichsberichte mit Freigabevermerk. • Nutzung von Referenzdokumenten mit bekannten Änderungen zur Validierung.

6.11.2 Anmerkungen

Der beschriebene Anwendungsfall lässt sich ebenfalls für die automatisierte Analyse und den Vergleich von Normen- oder Prozessdokumenten einsetzen. Mittels NLP-gestützter Verfahren können neue oder geänderte

Normenanforderungen automatisch identifiziert und inhaltlich mit bestehenden QMS-Prozessen abgeglichen werden. So entsteht eine strukturierte Gap-Analyse, die Änderungen nachvollziehbar darstellt und Handlungsempfehlungen für notwendige Prozessanpassungen ableitet.

Technischer Hinweis

Der Vergleich basiert nicht auf reinen KI-Agenten. Stabilität, Reproduzierbarkeit und Auditierbarkeit entstehen erst durch eine hybride Pipeline aus deterministischer Logik und KI-gestützter Bewertung, analog zu etablierten industriellen Lösungen, die auf einer Kombination aus strukturiertem Parsing, Embedding-Verfahren und einem nachgelagerten Bewertungsschritt beruhen.

6.12 Interaktives Lernen

Beschreibung

Einsatz eines KI-gestützten, virtuellen „QM-Lehrers“, der Mitarbeitenden komplexe Qualitätsmanagement-Themen einfach, interaktiv und jederzeit verfügbar erklärt. Die KI beantwortet Fragen zu Normen, Prozessen, Prüfmethoden, Fehlerbildern oder QS-Tools in natürlicher Sprache und passt die Erklärungen an das Wissensniveau der Nutzerin oder des Nutzers an. Der virtuelle Lehrer dient sowohl zur schnellen Klärung im Arbeitsalltag als auch für strukturierte Schulungen (z. B. VDA-Standards, Prüfprozesse, 8D, Maschinenfähigkeiten, Prüfmittelfähigkeit).

Rahmenbedingungen

Komplexe Anforderungen, Normen und Prozessstandards verlangen eine hohe Genauigkeit in der Wissensvermittlung. Mitarbeitende in Produktion, QM, Engineering und Instandhaltung haben sehr unterschiedliche Vorkenntnisse, benötigen aber schnellen Zugang zu zuverlässigen Informationen im Alltag. Die Dokumentation im QM ist umfangreich, häufig aktualisiert und für viele Nutzer:innen schwer zugänglich, was zu Wissenslücken und wiederkehrenden Rückfragen führt. Gleichzeitig steigen die Anforderungen durch interne Audits, VDA-Standards und kundenspezifische Richtlinien, wodurch eine konsistente Schulung immer wichtiger wird. Viele Unternehmen arbeiten mehrsprachig, was eine multilinguale Wissensvermittlung erforderlich macht.

Der virtuelle KI-Lehrer soll dieses Umfeld unterstützen, indem er jederzeit präzise und freigegebene Informationen vermittelt. Für die Implementierung müssen Datenschutz- und IT-Sicherheitsanforderungen beachtet werden, da QM-Daten oft sensibel sind. Gleichzeitig müssen QM-Expert:innen eingebunden werden, um Inhalte freizugeben und fortlaufend zu pflegen. Insgesamt entstehen die Rahmenbedingungen aus dem Spannungsfeld zwischen hohem Schulungsbedarf, begrenzten Ressourcen, regulatorischen Anforderungen und dem Wunsch, Wissen am Shopfloor schnell verfügbar zu machen.

Mehrwert

- **Schnellere Klärung von QM-Fragen im Alltag** ohne Rückfrage bei Expert:innen oder langes Suchen in Dokumenten.
- **Einheitliches Verständnis** von Normen, Prozessen und Methoden über Standorte hinweg.
- **Bessere Qualität der Entscheidungen** durch konsistente, geprüfte Antworten.
- **Effiziente Einarbeitung neuer Mitarbeitender** in QM-Prozesse, Prüfpläne, Standards und Tools.
- **Reduktion repetitiver Fragen** an QM-Expert:innen und dadurch mehr Fokus auf wertschöpfende Aufgaben.
- **Interaktive Erklärungen von Fehlerbildern, Prüfmethode**n

Herausforderungen

- **Aktualität der QM-Inhalte** bei Änderungen in Normen (VDA, IATF, interne Spezifikationen).
- **Validierung der Antworten:** Die KI darf nur freigegebene QM-Informationen ausgeben.
- **Komplexität der Fachsprache** und standortspezifische Begriffe müssen korrekt verstanden werden.
- **Vertraulichkeit von Qualitätsdaten**, Audit-Findings oder Produktionsproblemen.
- **Akzeptanz bei Auditoren und QM-Teams**, die exakte und regelkonforme Formulierungen erwarten.
- **Integration von bestehendem Wissensmaterial** (Prozessbeschreibungen, Prüfpläne, Lessons Learned).

oder Root-Cause-Methoden (z. B. 5Why, Ishikawa).

- **Schnelle Aktualisierung** bei Änderungen im VDA-Regelwerk, internen Spezifikationen oder Audit-Anforderungen.

Vorgehensweise

- **Scope definieren** → Themen eingrenzen: z. B. 8D, Prüfmittelfähigkeit, Maschinenfähigkeit, VDA 6.3, Fehlerkataloge und Zielgruppen festlegen: Produktion, QM, Maintenance, Engineering.
- **Wissensbasis aufbauen** → Relevante QM-Dokumente sammeln (VDA-Standards, interne Vorgaben, Prüfpläne, FAQs) und Inhalte strukturieren, versionieren und für die KI validieren.
- **Technische Plattform auswählen** → Entscheidung: On-Premise-KI, Cloud-KI mit abgesichertem Unternehmenszugang oder Integration in die Unternehmens-Lernplattform.
- **Pilot entwickeln** → Start mit einem klaren QM-Bereich (z. B. „Virtueller Lehrer für Prüfmittelfähigkeit“). Typische Nutzerfragen definieren (aus Audits, Shopfloor, Engineering).
- **Testen & Validieren** → QM-Expert:innen prüfen die Antwortqualität. Inhalte iterativ verbessern und Lücken schließen.
- **Rollout & Training** → Einführung im Shopfloor und QM. Kurze Tutorials, Live-Demos, „Frag den virtuellen QM-Lehrer“-Sessions.
- **Betrieb & kontinuierliche Verbesserung** → Verantwortlichkeiten definieren (Content Owner, QM-Expert:innen). Nutzung monitorieren, häufige Fragen analysieren, Inhalte laufend aktualisieren.

QM-Rollen

Mitarbeiter:innen in der Lieferantenqualität, in der Entwicklungsqualität, in der Produktionsqualität, in der Kundenqualität, im Qualitätsmanagementsystem, Qualitäts-Auditor:in, Qualitäts-Assessor:in, Qualitätsführungskraft,

AI Q-Data Engineer, AI Q-Data Analyst, AI Q-Data Scientist, AI Q-Data Manager

7 Exkurs: Risikobasierte Bewertung von KI-Entwicklungswerkzeugen

In diesem Abschnitt wird eine Methodik vorgestellt, die eine risikobasierte Bewertung von KI-Entwicklungswerkzeugen ermöglicht. Als KI-Entwicklungswerkzeuge werden hierbei genau die Werkzeuge bezeichnet, die im Rahmen des KI-Lebenszyklus zur Entwicklung von KI-Applikationen, KI-Bestandteilen von Applikationen oder vollständigen KI-Systemen eingesetzt werden. Zur operationalen Bewertung und, wo erforderlich, Qualifizierung von KI-Entwicklungswerkzeugen wird ein risikobasierter Ansatz empfohlen. Ziel ist es, den Bewertungs- und Absicherungsaufwand proportional zum Risiko steuern zu können. Das Vorgehen orientiert sich an in Safety-Domänen etablierten Praktiken (u. a. ISO 26262, DO-330/DO-178C), nutzt deren Kerngrößen Tool Impact (TI)⁸ und Tool Error Detection (TD)⁹ und adressiert zugleich andere regulatorisch relevante Bereiche wie den Datenschutz und die KI-Regulatorik (EU AI Act) mit ihren Anforderungen an Transparenz, Rückverfolgbarkeit und menschlicher Überwachung.

Die Methodik besteht aus zwei unabhängig voneinander durchführbaren Analyseprozessen.

- Einer Analyse des Prozessrisikos für ausgewählte Entwicklungsaufgaben im KI-Entwicklungslebenszyklus.
- Der Bewertung eines oder mehrerer Entwicklungswerkzeuge im Kontext seiner Aufgabe im Entwicklungslebenszyklus unter Berücksichtigung der zuvor ermittelten Prozessrisiken.

Die Trennung in eine Prozessrisikoanalyse und eine werkzeugspezifische Bewertung erhöht die Flexibilität des Ansatzes. Durch die modulare Struktur

⁸ Der Tool Impact beschreibt den Einfluss des Tools auf die Sicherheit des Produkts und lässt sich in TI-0 (Tool hat keinen Einfluss auf die Sicherheit) und TI-1 (Tool kann die Sicherheit des Produkts gefährden) unterteilen.

⁹ Die Tool Error Detection dient der Einschätzung der Wahrscheinlichkeit, einen Fehler des Tools zu erkennen. Man kann die Tool Error Detection in drei Stufen festsetzen. Diese können TD-0 (hohe Entdeckungswahrscheinlichkeit), TD-1 (Toolfehler wird vermutlich erkannt) und TD-2 (Toolfehler wird vermutlich nicht erkannt) sein.

können Prozessrisiken zunächst unabhängig von konkreten Entwicklungswerkzeugen identifiziert und bewertet werden. So entsteht eine stabile, referenzierbare Risikobasis entlang des KI-Entwicklungslebenszyklus.

Die Bewertung einzelner Entwicklungswerkzeuge erfolgt anschließend kontextbezogen auf Basis der zuvor bestimmten Prozessrisiken. Änderungen in der Werkzeugkette, etwa durch Austausch, Update oder Hinzunahme neuer Werkzeuge, erfordern damit keine vollständige Neubewertung des Gesamtprozesses, sondern lediglich eine erneute werkzeugspezifische Analyse im jeweiligen Anwendungskontext.

Darüber hinaus erlaubt die Entkopplung beider Analyseebenen die Entwicklung standardisierter Vorlagen oder Referenzmodelle für typische Entwicklungsprozesse. Solche vordefinierten Prozessrisikoprofile können organisationsweit wiederverwendet und projektspezifisch angepasst werden, was sowohl Effizienzgewinne als auch eine konsistente Bewertungsgrundlage unterstützt.

7.1 Erläuterung der Grundkonzepte

Das hier beschriebene Bewertungsverfahren basiert auf einer formalisierten Beschreibung des KI-Lebenszyklus, der darin stattfindenden Entwicklungsaufgaben, der eingesetzten Werkzeuge sowie der damit verbundenen Risiken. Ziel ist es, Risiken systematisch zu identifizieren, ihren Entstehungskontext transparent zu machen und die Bedeutung von Werkzeugen als Risikofaktor strukturiert zu beschreiben.

Grundlage bildet ein konsistentes Informationsmodell, das die relevanten Konzepte und ihre Abhängigkeiten explizit beschreibt. Dadurch werden Einflussfaktoren auf Risiken, deren Ursachenketten sowie mögliche Gegenmaßnahmen nachvollziehbar und wiederverwendbar dokumentiert. Das Modell unterstützt sowohl die Prozessrisikoanalyse als auch die nachgelagerte werkzeugspezifische Bewertung. Das Modell ist in Abbildung 7-1 dargestellt.

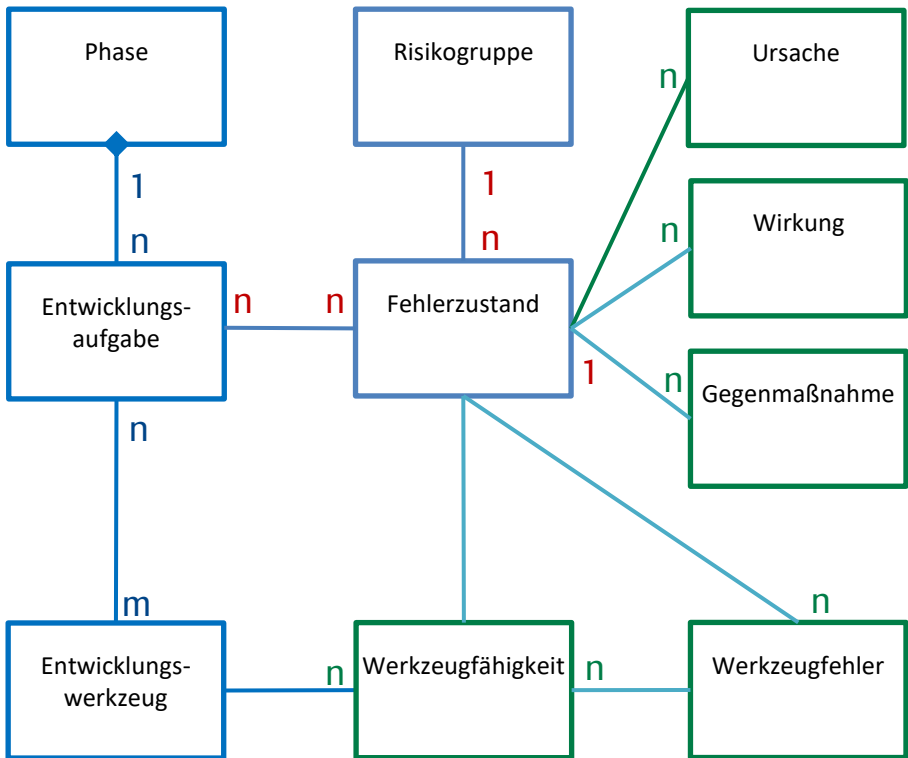


Abbildung 7-1: Informationsmodell

Im Folgenden werden die zentralen Konzepte des Modells aus Abbildung 7-1 erläutert.

- **Phasen:** Der KI-Lebenszyklus wird in strukturierte Phasen unterteilt (siehe hierzu auch Prozessphasen in Kapitel 5.2.1). Jede Phase stellt einen abgegrenzten Kontext dar, in dem spezifische regulatorische, sicherheitsbezogene und datenschutzrechtliche Anforderungen wirksam werden. Die Phasenstruktur dient als Ordnungsrahmen für die Zuordnung von Entwicklungsaufgaben, die Identifikation typischer Fehlerzustände und die systematische Verortung von Risiken.
- **Entwicklungsaufgaben:** Innerhalb jeder Phase werden konkrete Entwicklungsaufgaben definiert. Beispiele hierfür sind Aufgaben wie

„Datenvorverarbeitung und -transformation“, „Feature Engineering und Extraktion“ in der Phase „Datenaufbereitung“ oder „Experiment Tracking & Management“ in der Phase „KI-Modellierung“. Die Entwicklungsaufgaben bilden den zentralen Bezugspunkt der Risikoanalyse. Für jede Aufgabe wird untersucht, welche Fehler auftreten können und welche regulatorischen oder sicherheitsrelevanten Anforderungen dadurch verletzt würden.

Entwicklungsaufgaben stellen somit die Verbindung zwischen abstrakter Prozessphase und konkreter operativer Tätigkeit dar.

- Fehlerzustände: Für jede Entwicklungsaufgabe werden mögliche Fehlerzustände identifiziert. Ein Fehlerzustand beschreibt einen konkreten unerwünschten Prozess- oder Systemzustand, der im Rahmen einer Entwicklungsaufgabe auftreten kann und im Falle seines Eintretens zu einer Nichtkonformität mit regulatorischen, datenschutzrechtlichen oder sicherheitskritischen Anforderungen führt. Beispiele für Fehlerzustände sind u. a.
 - Verlust der Datenintegrität, fehlende Echtzeitverarbeitung, inkonsistente Modellreproduzierbarkeit oder die unzureichende Behandlung von Randfällen, aber auch
 - Fehler bei der Anonymisierung und Pseudonymisierung von Daten, intransparente Verwaltung von Nutzereinigilligungen, Verstöße bei grenzüberschreitenden Datenübermittlungen oder Nichteinhaltung von Bias- und Fairness-Anforderungen.

Jeder Fehlerzustand ist einer Risikogruppe zugeordnet. Damit wird transparent, welche regulatorische oder sicherheitsbezogene Dimension betroffen ist.

- Risikogruppen: Zur Strukturierung werden Fehlerzustände thematischen Risikogruppen zugeordnet (siehe hierzu Kapitel 5.1). Diese Gruppierung erlaubt eine systematische Abdeckung regulatorischer Anforderungen (z. B. EU AI Act) sowie die strukturierte Ableitung von Prüf- und Absicherungsmaßnahmen.
- Ursache, Wirkung und Gegenmaßnahmen: Ausgehend von den Fehlerzuständen und Risikogruppen werden für jeden

Fehlerzustand Ursachen (z. B. Fehler in der Aufgabenabarbeitung, ungenügende Absicherungsmaßnahmen und Kontrollen), Wirkungen (z. B. fehlerhafte oder fehlende Ergebnisartefakte) und Gegenmaßnahmen (z. B. prozessuale Vorkehrungen) identifiziert. Hierdurch entsteht eine nachvollziehbare Ursache-Wirkungs-Kette, die gezielte Maßnahmen zur Risikoreduktion ermöglicht.

- **Entwicklungswerkzeuge:** Entwicklungsaufgaben werden typischerweise durch spezifische Werkzeuge unterstützt, beispielsweise ML-Frameworks, Trainingsumgebungen, Versionsverwaltungssysteme oder Monitoring-Plattformen. Entwicklungswerkzeuge sind stets kontextgebunden zu betrachten. Ihre Bewertung erfolgt nicht isoliert, sondern in Bezug auf die jeweilige Entwicklungsaufgabe und die damit verbundenen Risiken.
- **Werkzeugfähigkeiten:** Für jede Entwicklungsaufgabe wird analysiert, welche Werkzeugfähigkeiten genutzt werden, um eine Entwicklungsaufgabe umzusetzen. Ihre Existenz, Reife und korrekte Konfiguration sind zentrale Bewertungskriterien im Rahmen der Werkzeugqualifikation.
- **Werkzeugfehler:** Ein Werkzeugfehler ist der Zustand, in dem ein Werkzeug bei der Ausübung seiner Fähigkeiten fehlerhaft agiert und damit zur Ursache eines Fehlerzustands wird. Ursachen dafür sind u.a. fehlerhafte Implementierung, falsche Konfiguration oder unzureichende Validierung. Werkzeugfehler können die zuvor genannten Fehlerzustände entweder direkt oder indirekt, durch eine verminderte Werkzeugfähigkeit, begründen.

Der zentrale Nutzen des Informationsmodells liegt in der Bereitstellung eines klar strukturierten Ordnungs- und Analyserahmens, der eine systematische Identifikation und Analyse von Risiken unter expliziter Berücksichtigung von Abhängigkeiten zwischen Entwicklungswerkzeugen, ihren Aufgaben und Fehlern sowie den Entwicklungsaufgaben und den damit verknüpften Prozessrisiken ermöglicht.

Dies macht Wechselwirkungen und Einflussketten transparent und verhindert isolierte Einzelbetrachtungen.

7.2 Durchführung der risikobasierten Bewertung von KI-Entwicklungswerkzeugen

Die risikobasierte Bewertung von KI-Entwicklungswerkzeugen folgt einem strukturierten, zweistufigen Vorgehen, das sowohl die Prozessperspektive als auch die Werkzeugenebene systematisch berücksichtigt. Ausgangspunkt ist die Betrachtung konkreter Entwicklungsaufgaben innerhalb definierter Phasen des KI-Lebenszyklus. Für diese Aufgaben werden potenzielle Fehlerzustände, deren Ursachen und Wirkungen sowie geeignete Gegenmaßnahmen identifiziert und hinsichtlich ihrer Kritikalität bewertet. Auf diese Weise entsteht eine nachvollziehbare Risikoeinschätzung für die Durchführung ausgewählter Entwicklungsaktivitäten.

Auf dieser Grundlage erfolgt anschließend die Bewertung der in diesen Aufgaben eingesetzten Werkzeuge. Dabei wird analysiert, inwieweit werkzeugspezifische Eigenschaften oder Fehlerpotenziale zur Entstehung identifizierter Risiken beitragen können. Im Fokus steht somit die Frage, welchen Einfluss ein Werkzeug auf das Risiko der jeweiligen Entwicklungsaufgabe hat und welcher Absicherungs- oder Qualifizierungsbedarf daraus resultiert.

1. Ermittlung von Risiken für Entwicklungsaufgaben

1.1 Auswahl relevanter Phasen und Entwicklungsaufgaben

1.2. Identifikation potentieller Fehlerzustände

1.3. Ermittlung des Risikos für Fehlerzustände



2. Ermittlung des Qualifizierungsbedarfs für KI-Entwicklungswerkzeuge

2.1. Ermittlung der Bedeutung des Werkzeugs für einen Fehlerzustand (Tool Impact)

2.2. Bewertung der Erkennbarkeit des Werkzeugfehlers (Tool Error Detection) und Bestimmung des Tool Confidence Level (TCL) pro Risikogruppe

2.3. Abschließende Werkzeugbewertung und Handlungsempfehlung auf Basis des Risikos und des TCL

Abbildung 7-2: Vorgehen zur Ermittlung des Qualifizierungsbedarfs für KI-Werkzeuge

Das kombinierte Vorgehen ermöglicht eine konsistente Herleitung des Qualifizierungsbedarfs. Risiken werden zunächst im fachlichen und regulatorischen Kontext der Aufgabe bestimmt und anschließend in Beziehung zu den eingesetzten Werkzeugen gesetzt. Dadurch wird eine transparente, kontextbezogene und methodisch fundierte Bewertung von KI-Entwicklungswerkzeugen entlang des gesamten Lebenszyklus unterstützt.

7.2.1 Schritt 1: Ermittlung von Risiken für ausgesuchte Entwicklungsaufgaben



Abbildung 7-3: Ermittlung von Risiken für Entwicklungsfolien

Ziel des ersten Schrittes ist die systematische Identifikation und Bewertung von Prozessrisiken für ausgewählte Entwicklungsaufgaben innerhalb definierter Phasen des KI-Entwicklungslebenszyklus. Im Mittelpunkt steht die Frage, welche Fehlerzustände im Rahmen einer konkreten Aufgabe auftreten können, welche Auswirkungen diese haben und wie kritisch sie im regulatorischen oder sicherheitsrelevanten Kontext zu bewerten sind. Die Risikoermittlung kann sich dabei an etablierten Verfahren wie der Failure Mode and Effects Analysis (FMEA) orientieren. Das Vorgehen erfolgt in mehreren strukturierten Teilschritten.

7.2.1.1 **Unterschritt 1.1: Auswahl relevanter Phasen und Entwicklungsaufgaben**



Abbildung 7-4: Auswahl relevanter Phasen und Entwicklungsaufgaben

Zunächst werden die Phasen des KI-Lebenszyklus bestimmt, in denen eine Bewertung erfolgen soll. Innerhalb dieser Phasen werden die konkreten Entwicklungsaufgaben identifiziert, die hinsichtlich möglicher Risiken analysiert werden. Die Orientierung an einem definierten Phasenmodell stellt sicher, dass die Analyse konsistent, vollständig und vergleichbar durchgeführt wird. Gleichzeitig ermöglicht die gezielte Auswahl einzelner Aufgaben eine fokussierte Betrachtung besonders kritischer Prozessschritte.

7.2.1.2 **Unterschritt 1.2: Identifikation potenzieller Fehlerzustände**



Abbildung 7-5: Identifikation potenzieller Fehlerzustände

Im nächsten Schritt werden für die ausgewählten Entwicklungsaufgaben potenzielle Fehlerzustände identifiziert. Die Ableitung erfolgt unter Berücksichtigung geeigneter Risikogruppen bzw. Konformitätsbereiche (z. B. Sicherheit, Datenschutz, Transparenz, Robustheit).

Nicht jede Risikogruppe ist in jedem Kontext gleichermaßen relevant. Abhängig von Anwendungsdomäne, Systemkritikalität oder regulatorischem Rahmen können bestimmte Fehlerzustände ein- oder ausgeschlossen wer-

den. Zusätzlich werden für jeden identifizierten Fehlerzustand mögliche Ursachen, potenzielle Wirkungen und mögliche Gegenmaßnahmen systematisch erfasst und dokumentiert.

7.2.1.3 Unterschrift 1.3: Ermittlung des Risikos für Fehlerzustände

Abbildung 7-6: Ermittlung des Risikos für Fehlerzustände



Die identifizierten Fehlerzustände bilden die Grundlage für die anschließende Risikobewertung. Im Rahmen einer FMEA erfolgt die Bewertung beispielsweise anhand der etablierten Faktoren:

- Severity (S) – Schwere der Auswirkung im Fehlerfall (Skala 1–10)
- Occurrence (O) – Wahrscheinlichkeit des Auftretens (Skala 1–10)
- Detection (D) – Wahrscheinlichkeit der Entdeckung vor Wirksamwerden (Skala 1–10)

Aus diesen drei Faktoren lässt sich dann entweder eine Risikoprioritätskennzahl (RPN) oder Aufgabenpriorität (AP) berechnen.

Während die RPN mit $S \times O \times D$ multiplikativ ermittelt wird, realisiert die AP eine prioritätsorientierte Einstufung gemäß definierter Bewertungsmatrix. Dabei wird insbesondere der Severity-Faktor stärker gewichtet, sodass hohe Schadensauswirkungen unabhängig von niedriger Eintrittswahrscheinlichkeit zu einer hohen Priorität führen können.

Das Ergebnis dieses Schrittes ist eine strukturierte, bewertete Risikoliste für die betrachteten Entwicklungsaufgaben. Diese bildet die fachliche und regulatorische Grundlage für die anschließende werkzeugspezifische Bewertung im zweiten Schritt.

7.2.2 Schritt 2: Werkzeugbewertung und Ermittlung des Qualifizierungsbedarfs



Abbildung 7-7: Ermittlung des Qualifizierungsbedarfs für KI-Entwicklungswerkzeuge

Aufbauend auf der zuvor beschriebenen Risikoermittlung erfolgt im zweiten Schritt die Bewertung der eingesetzten Werkzeuge. Ziel ist es, den potenziellen Einfluss der Entwicklungswerkzeuge auf das Zustandekommen oder die Vermeidung identifizierter Fehlerzustände systematisch zu erfassen und zu bewerten. Dabei wird ein methodischer Rahmen bereitgestellt, der sich an den Prinzipien der Werkzeugbewertung aus der ISO 26262 orientiert und als Hilfestellung für die strukturierte Bewertung und Vorbereitung einer möglichen Werkzeugqualifizierung dient.

7.2.2.1 **Unterschritt 2.1: Ermittlung der Bedeutung des Werkzeugs für einen Fehlerzustand (Tool Impact)**

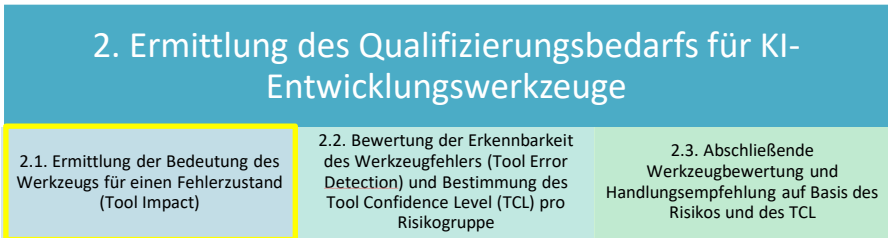


Abbildung 7-8: Ermittlung der Bedeutung des Werkzeugs für einen Fehlerzustand

Im ersten Schritt wird für jeden relevanten Fehlerzustand abgeschätzt, welchen Einfluss ein Werkzeugfehler auf das Eintreten oder die Verstärkung eines Prozessfehlers haben könnte. Diese Bewertung wird als Tool Impact (TI) bezeichnet. Ein hoher Tool Impact liegt vor, wenn ein Fehler im Werkzeug direkt oder indirekt zu sicherheitskritischen, datenschutzrelevanten oder regulatorischen Nichtkonformitäten führen kann. Ein geringer Tool Impact bedeutet, dass der Einfluss des Werkzeugs auf den betrachteten Fehlerzustand marginal oder leicht kompensierbar ist.

7.2.2.2 **Unterschritt 2.2: Bewertung der Erkennbarkeit des Werkzeugfehlers (Tool Error Detection) und Bestimmung des Tool Confidence Level (TCL) pro Risikogruppe**

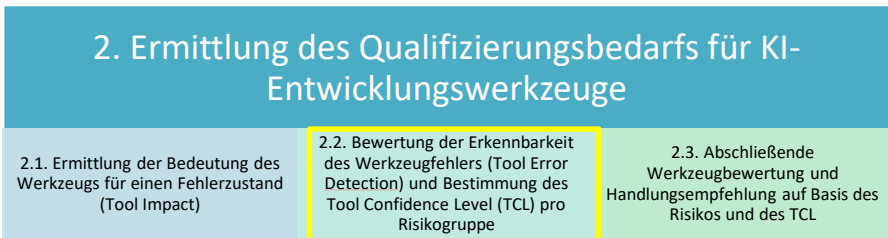


Abbildung 7-9: Bewertung der Erkennbarkeit des Werkzeugfehlers und Bestimmung des TCL

Anschließend wird beurteilt, wie wahrscheinlich es ist, dass ein durch das Werkzeug verursachter Fehler entdeckt wird, bevor er zu einem

Fehlverhalten oder einer Nichtkonformität führt. Diese Einschätzung wird als Tool Error Detection (TD) bezeichnet.

Ein hoher TD-Wert zeigt an, dass Werkzeugfehler schwer erkennbar sind und zusätzliche Prüfmechanismen oder Prozessmaßnahmen notwendig sein könnten. Ein niedriger TD-Wert hingegen signalisiert, dass Fehler im Werkzeug typischerweise schnell erkannt oder durch nachgelagerte Prüfungen identifiziert werden können.

Aus der Kombination von TI und TD ergibt sich ein qualitatives Vertrauensmaß, das als Tool Confidence Level (TCL) bezeichnet wird. Dieses Maß dient nicht der formalen Einstufung im Sinne der ISO 26262, sondern der Orientierung für die Beurteilung des erforderlichen Vertrauens in ein Werkzeug:

1. TCL 1: Werkzeug kann ohne besondere Maßnahmen eingesetzt werden; Risiken sind gering oder durch Prozesse hinreichend kontrolliert.
2. TCL 2–3: Werkzeug weist einen höheren Vertrauensbedarf auf; zusätzliche Prüfungen oder begleitende Maßnahmen sind empfohlen.

Für jede Kombination aus Aufgabe, Fehlerzustand und Werkzeug können TI- und TD-Werte vergeben werden. Der resultierende TCL wird aus den ermittelten Werten berechnet und mit der zuvor bestimmten Risikoprioritätskennzahl (RPN) aus der FMEA verknüpft.

7.2.2.3 **Unterschnitt 2.3: Abschließende Werkzeugbewertung und Handlungsempfehlung auf Basis des Risikos und des TCL**

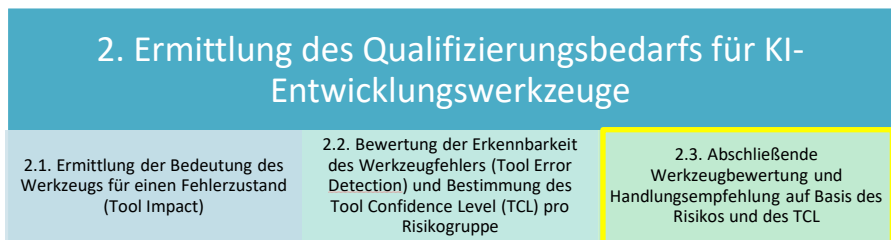


Abbildung 7-10: Abschließende Werkzeugbewertung und Handlungsempfehlung

Durch die Kombination aus dem prozessbezogenen Risiko und dem TCL (werkzeugbezogener Vertrauensbedarf) entsteht eine priorisierte Übersicht über die kritischsten Werkzeuge im jeweiligen Entwicklungskontext.

Beide Bewertungsgrößen können über Schwellwerte gefiltert werden, sodass nur jene Kombinationen berücksichtigt werden, bei denen sowohl ein erhöhtes Prozessrisiko als auch ein hoher Vertrauensbedarf vorliegt.

Diese integrierte Sicht bildet die Grundlage dafür, gezielt Maßnahmen zur Prozessverbesserung und Werkzeugabsicherung abzuleiten.

Für jede priorisierte Kombination aus Aufgabe, Fehlerzustand und Werkzeugbewertung werden zwei Typen von Maßnahmen hinterlegt:

- Possible Process Mitigations: organisatorische oder prozessuale Gegenmaßnahmen, die zur Minderung des identifizierten Risikos beitragen (z. B. zusätzliche Validierungsschritte, manuelle Reviews, Freigabeprozesse).
- Testable Tool Capabilities: funktionale Anforderungen an das Werkzeug, die eine technische Absicherung ermöglichen (z. B. deterministische Versionierung, Audit-Logging, automatische Fehlererkennung)

Prozessmaßnahmen und testbare Werkzeugfähigkeiten können einzeln ausgewählt und dokumentiert werden. Das Ergebnis ist eine nachvollziehbare, aufgabenspezifische Zusammenstellung aller identifizierten Risiken, ihrer Bewertungen (RPN, TCL) sowie der abgeleiteten Gegenmaßnahmen und Werkzeuganforderungen.

7.2.3 Liste potenzieller Fehlerzustände (beispielhaft)

Tabelle 7-1: Zuordnung Fehlerzustände zu Risikogruppen und Entwicklungsaufgaben

Risikogruppe (Hauptgruppe + weitere)	Fehlerzustand	Entwicklungsaufgaben
Produkteigenschaften (Regulatorik, Finanzielles Risiko)	Verlust der Datenintegrität: Beschädigte oder inkonsistente Daten führen zu fehlerhaften Modellvorhersagen und gefährden sicherheitskritische Systeme.	Data Storage & Management; Data Preprocessing & Transformation
Produkteigenschaften (Finanzielles Risiko, Reputationsrisiko)	Fehlende Echtzeitverarbeitung: Nichteinhaltung von Echtzeit-Inferenzanforderungen in autonomen oder sicherheitskritischen Anwendungen.	Deep Learning & ML Frameworks; Real-time & Batch Inference
Regulatorik (Produkteigenschaften, Finanzielles Risiko)	Inkonsistente Modellreproduzierbarkeit: Nicht-deterministische oder undokumentierte Trainingsprozesse führen zu unvorhersehbarem Modellverhalten.	Experiment Tracking & Management; Model Evaluation & Explainability; Data Versioning & Management; Data Collection & Integration
Produkteigenschaften (Bias & Fairness, Reputationsrisiko)	Unzureichende Behandlung von Randfällen: KI-Modelle behandeln seltene, aber sicherheitskritische Szenarien nicht korrekt.	Data Labeling & Annotation; Feature Engineering & Extraction; Synthetic Data Generation
Produkteigenschaften (Regulatorik, Finanzielles Risiko)	Nicht erkannter Modell-Performance-Drift: Modellleistung verschlechtert sich im Zeitverlauf ohne Korrekturmaßnahmen.	Experiment Tracking & Management; Model Monitoring & Drift Detection; Real-time & Batch

		Inference; Anomaly Detection
Produkteigenschaften (Finanzielles Risiko, Reputationsrisiko)	Unzureichende Redundanz- & Failover-Mechanismen: Fehlende Backup-Modelle oder Ausfallstrategien in kritischen Systemen.	CI/CD for ML (MLOps); Cloud Deployment; Edge & IoT Deployment
Regulatorik (Transparenz, Finanzielles Risiko)	Unzureichende Fehlerbehandlung & Logging: Unzureichende Protokollierung erschwert Fehleranalyse und Ursachenidentifikation.	Deep Learning & ML Frameworks; Model Evaluation & Explainability
Produkteigenschaften (Finanzielles Risiko, Reputationsrisiko)	Modellinstabilität & schlechte Generalisierung: Modell verhält sich unvorhersehbar oder generalisiert nicht zuverlässig.	Data Preprocessing & Transformation; Feature Engineering & Extraction; Synthetic Data Generation; Hyperparameter Tuning; Algorithm Prototyping & Development; Deep Learning & ML Frameworks; Distributed Training & Optimization
Datenschutz (Regulatorik, Reputationsrisiko, Finanzielles Risiko)	Unbefugter Datenzugriff: Fehlende Zugriffskontrollen führen zu Datenschutzverletzungen.	Data Storage & Management; Data Collection & Integration; Cloud Deployment
Datenschutz (Regulatorik, Reputationsrisiko)	Unzureichende Umsetzung des Rechts auf Löschung: Nichteinhaltung des DSGVO-Rechts auf Vergessenwerden.	Data Storage & Management

Datenschutz (Reputationsrisiko, Finanzielles Risiko)	Offenlegung personenbezogener Daten: ML-Modelle legen unbeabsichtigt personenbezogene Daten offen.	Deep Learning & ML Frameworks; Anomaly Detection; Cloud Deployment
Datenschutz (Regulatorik, Finanzielles Risiko)	Fehlende Datenminimierung: Unverhältnismäßige Erhebung und Speicherung personenbezogener Daten.	Data Collection & Integration; Data Quality & Bias Detection
Datenschutz (Regulatorik, Reputationsrisiko)	Fehler bei Anonymisierung & Pseudonymisierung: Unzureichende De-Identifizierung führt zu re-identifizierbaren Daten.	Feature Engineering & Extraction; Data Storage & Management; Data Preprocessing & Transformation; Synthetic Data Generation
Datenschutz (Transparenz, Regulatorik)	Intransparente Einwilligungsverwaltung: Fehlende transparente Nachverfolgung von Nutzereinzwilligungen.	Data Collection & Integration; Regulatory Compliance Monitoring
Datenschutz (Regulatorik, Finanzielles Risiko)	Verstöße bei grenzüberschreitender Datenübermittlung: Nichteinhaltung internationaler Datenübermittlungsregeln.	Cloud Deployment; Data Storage & Management
Transparenz (Regulatorik, Reputationsrisiko)	Mangelnde Transparenz von KI-Systemen: Fehlende Erklärbarkeit von KI-Entscheidungen.	Feature Engineering & Extraction; Model Evaluation & Explainability; Data Labeling & Annotation; Synthetic Data Generation; Deep Learning & ML Frameworks; Edge & IoT Deployment

Bias & Fairness (Regulatorik, Reputationsrisiko, Finanzielles Risiko)	Bias- & Fairness-Nichteinhaltung: Verstärkung von Diskriminierung und Verstoß gegen Fairness-Vorgaben.	Bias & Fairness Audits; Deep Learning & ML Frameworks; Data Labeling & Annotation; Data Quality & Bias Detection; Feature Engineering & Extraction; Distributed Training & Optimization
Regulatorik (Transparenz, Reputationsrisiko)	Fehlende Modell-Risikobewertung: Fehlende Einstufung nach AI-Act-Risikoklassen.	Regulatory Compliance Monitoring; Risk Management
Produkteigenschaften (Regulatorik, Reputationsrisiko, Finanzielles Risiko)	Sicherheitslücken in KI-Systemen: Anfälligkeit für typische KI-Angriffsvektoren (adversarielle Angriffe, data poisoning etc.)	Data Storage & Management; Data Preprocessing & Transformation; Deep Learning & ML Frameworks; Cloud Deployment; Edge & IoT Deployment
Regulatorik (Transparenz, Reputationsrisiko)	Nicht-konforme Human-Oversight-Mechanismen: Fehlende Human-in-the-Loop-Kontrollen.	Regulatory Compliance Monitoring; Observability & Performance Tracking
Bias & Fairness (Reputationsrisiko, Regulatorik)	Fehlende ethische Schutzmechanismen: Ethische Prinzipien nicht ausreichend berücksichtigt.	Bias & Fairness Audits; Model Evaluation & Explainability
Regulatorik (Finanzielles Risiko, Reputationsrisiko)	Unzureichende KI-Folgenabschätzung: Fehlende Risiko-Nutzen-Analyse.	Bias & Fairness Audits; Model Evaluation & Explainability

7.3 Beispielhafte Anwendung der Methode

Zur besseren Veranschaulichung wird die vorgestellte Methode exemplarisch zur Bewertung des fiktiven Werkzeugs *MLtoolExample* angewendet. In diesem Beispiel wird die Entwicklung einer KI-gestützten Auswertung von SPC-Daten im Qualitätsmanagement betrachtet.

7.3.1 Kontext KI-gestützte SPC-Auswertung im Qualitätsmanagement

Die betrachtete Anwendung dient der automatisierten Analyse von Prozesskennzahlen in der automobilen Serienfertigung. Ziel ist es, Abweichungen, Trends und Anomalien frühzeitig zu erkennen, um Prozessstabilität sicherzustellen und die Produktkonformität zu gewährleisten. Der Einsatz erfolgt innerhalb eines Qualitätsmanagementsystems, wie beispielsweise IATF 16949 und ISO 9001, und ist in APQP-/RGA-Prozesse eingebunden.

Fehlerhafte oder nicht nachvollziehbare Modellbewertungen können zu falschen Prozessfreigaben, unnötigen Produktionsunterbrechungen oder zur Freigabe nicht konformer Bauteile führen. Damit sind insbesondere Anforderungen an dokumentierte Information, Änderungsmanagement, Validierung, Requalifikation und Auditfähigkeit betroffen. Machine-Learning-Modelle erweitern klassische SPC-Methoden, indem sie komplexe multivariate Zusammenhänge und schleichende Prozessveränderungen erkennen. Ihre Reproduzierbarkeit und kontinuierliche Überwachung sind daher entscheidend für die Einhaltung von Qualitäts- und Konformitätsanforderungen.

7.3.2 Kontext des fiktiven Entwicklungswerkzeugs *MLtoolExample*

Das fiktive Entwicklungswerkzeug *MLtoolExample* ist eine Open-Source-Plattform zur Unterstützung des ML-Lebenszyklus. Für die Werkzeugbewertung sind insbesondere die Funktionen zur Experimentverfolgung, Versionierung von Modellständen und Verwaltung von Trainingsartefakten relevant. Das Werkzeug ermöglicht die strukturierte Erfassung von Parametern, Metriken, Modellversionen und Umgebungsinformationen und schafft damit die technische Grundlage für Nachvollziehbarkeit, Reproduzierbarkeit und auditierbare Modellfreigaben.

7.3.3 Schritt 1 Ermittlung von Risiken für Entwicklungsaufgaben

Im ersten Schritt der risikobasierten Bewertung werden für ausgewählte Entwicklungsaufgaben systematisch potenzielle Fehlerzustände identifiziert und hinsichtlich ihrer Auswirkungen auf Qualität, Konformität und Produktfähigkeit bewertet, um eine belastbare Grundlage für die anschließende werkzeugspezifische Analyse zu schaffen.

7.3.3.1 Unterschritt 1.1 Auswahl relevanter Entwicklungsaufgaben

Im Rahmen der risikobasierten Bewertung werden jene Entwicklungsaufgaben betrachtet, in denen das fiktive Werkzeug *MLtoolExample* einen unmittelbaren Einfluss auf Nachvollziehbarkeit und Qualitätssicherung besitzt. Hierzu zählen insbesondere

- Experiment Tracking & Management,
- Data Versioning & Management sowie
- CI/CD for ML (MLOps).

Für das vorliegende Beispiel wird die Analyse auf die Entwicklungsaufgaben „Experiment Tracking & Management“ und „CI/CD for ML“ fokussiert, da hier die Grundlage für reproduzierbare Modellstände und dokumentierte Modellfreigaben geschaffen wird.

7.3.3.2 Unterschritt 1.2 Auswahl relevanter Fehlerzustände

Für die gewählte Entwicklungsaufgabe werden zwei relevante Fehlerzustände betrachtet.

Tabelle 7-2: Ausgewählte Fehlerzustände für die Entwicklungsaufgabe „Experiment Tracking & Management“

Risikogruppe (Hauptgruppe + weitere)	Fehlerzustand	Entwicklungs- aufgaben
Regulatorik (Produkteigenschaften, Finanzielles Risiko)	Inkonsistente Modellreproduzierbarkeit: Nicht-deterministische oder undokumentierte Trainingsprozesse führen	Experiment Tracking & Management; Model Evaluation & Explainability; Data Versioning &

	zu unvorhersehbarem Modellverhalten.	Management; Data Collection & Integration
Produkteigenschaften (Finanzielles Risiko, Reputationsrisiko)	Unzureichende Redundanz- & Failover-Mechanismen: Fehlende Backup-Modelle oder Ausfallstrategien in kritischen Systemen.	CI/CD for ML (MLOps); Cloud Deployment; Edge & IoT Deployment

Der erste Fehlerzustand adressiert die fehlende Reproduzierbarkeit von Trainings- und Modellständen infolge unvollständiger oder inkonsistenter Erfassung von Trainingsparametern, Datenständen oder Umgebungsinformationen. Dadurch können die Modellstände in der Entwicklung automatisierter Analyseanwendungen nicht eindeutig rekonstruiert, validiert oder formal freigegeben werden. Der Fehlerzustand verletzt damit unmittelbar normative Vorgaben zur Rückverfolgbarkeit, Änderungsfreigabe und Prozessvalidierung aus der Hauptrisikogruppe „Regulatorik“. Sekundär besteht ein Bezug zur Risikogruppe „Produkteigenschaften“, da nicht eindeutig rekonstruierbare Modellstände zu fehlerhaften Prozessbewertungen führen können. Wird beispielsweise ein nicht validierter Modellstand produktiv eingesetzt, kann dies die Bewertung der Prozessfähigkeit und damit mittelbar die Produktkonformität beeinflussen.

Der zweite Fehlerzustand adressiert unzureichende Redundanz- und Failover-Mechanismen im Betrieb der KI-gestützten Analyseanwendung. Er beschreibt den Fall, dass bei Ausfall eines Modells, einer Deployment-Instanz oder einer Infrastrukturkomponente keine definierte Backup-Strategie oder Rückfalllösung existiert. In einem solchen Szenario kann die automatisierte Prozessüberwachung unterbrochen werden oder es wird auf veraltete oder nicht validierte Modellstände zurückgegriffen.

Die Hauptrisikogruppe ist „Produkteigenschaften“, da die kontinuierliche Bewertung der Prozessfähigkeit Bestandteil der qualitätsrelevanten Produktionsüberwachung ist. Fällt das System ohne definierte Ausweichstrategie aus, kann die Fähigkeit zur frühzeitigen Erkennung von Prozessabweichungen eingeschränkt sein. Dadurch besteht das Risiko, dass nicht konforme Produkte produziert oder freigegeben werden.

Sekundär ist die Risikogruppe „Finanzielles Risiko“ betroffen, da Systemausfälle zu Produktionsunterbrechungen, erhöhtem Prüfaufwand oder verspäteter Fehlererkennung führen können. Ebenso besteht ein Bezug zur Risikogruppe „Reputationsrisiko“, insbesondere wenn Liefertermine oder Qualitätskennzahlen beeinträchtigt werden und sich dies auf Kundenbewertungen oder Lieferantenstatus auswirkt.

7.3.3.3 Unterschritt 1.3 Bewertung der Fehlerzustände

Für die automatisierte Analyse von Prozesskennzahlen dürfen nur Werkzeuge eingesetzt werden, deren Entwicklung selbst nachvollziehbar, kontrolliert und vollständig dokumentiert ist. Die Qualität der Analyseanwendung hängt nicht allein vom späteren Modellverhalten ab, sondern maßgeblich von der Strenge, mit der Trainingsprozesse, Modellstände und zugehörige Parameter im Entwicklungsprozess erfasst und versioniert werden.

Werden Modellstände, Trainingsparameter oder Umgebungsbedingungen nicht konsistent protokolliert, ist im Nachhinein nicht belastbar nachweisbar, unter welchen Bedingungen ein bestimmter Analysezustand entstanden ist. Damit fehlt der formale Beleg, dass das Analysewerkzeug mit der erforderlichen methodischen Sorgfalt entwickelt, validiert und freigegeben wurde. Im Fehlerfall, etwa bei Auditabweichungen, internen Qualitätsanalysen oder Kundenbeanstandungen, kann somit nicht nachgewiesen werden, dass die Analyseanwendung gemäß den Anforderungen an dokumentierte Information, Rückverfolgbarkeit und Änderungslenkung entwickelt wurde. Dieses Beispiel ist in der zweiten Zeile der Tabelle in Abbildung 7-11 dargestellt. Für diesen Fehlerzustand wird der Faktor Severity mit 9 bewertet, da die regulatorische Konformität der Anwendung im Fehlerfall unmittelbar infrage gestellt werden kann. Der Faktor Occurrence wird mit 5 eingeschätzt, da Inkonsistenzen häufig aus unvollständigem Logging oder fehlender Versionierung resultieren. Der Faktor Detection wird mit 7 bewertet, da Defizite typischerweise erst bei Reproduktionsversuchen, Freigabeprüfungen oder Audits sichtbar werden. Daraus ergibt sich ein RPN von 315 ($RPN = Severity \times Occurrence \times Detection = 9 \times 5 \times 7 = 315$).

Der zweite Fehlerzustand betrifft unzureichende Redundanz- und Failover-Mechanismen im Betrieb der KI-gestützten Analyseanwendung. Hierbei ist

insbesondere zu berücksichtigen, dass das fiktive Werkzeug *MLtoolExample* als Bestandteil der CI/CD- und Deployment-Pipeline agiert und Modellstände verwaltet, versioniert und für produktive Umgebungen bereitstellt. Ist das fiktive Werkzeug *MLtoolExample* nicht in eine robuste Backup- oder Rollback-Strategie eingebunden oder fehlen definierte Ausweichmechanismen für Modell- oder Infrastrukturkomponenten, kann im Störfall kein validierter Ersatzmodellstand bereitgestellt werden. In einem solchen Szenario wird die kontinuierliche Prozessüberwachung unterbrochen oder es wird auf nicht freigegebene beziehungsweise veraltete Modellstände zurückgegriffen. Dieses Beispiel ist in der ersten Zeile der Tabelle in Abbildung 7-11 dargestellt. Für diesen Fehlerzustand wird der Faktor Severity mit 8 bewertet, da Auswirkungen auf Produktqualität und Lieferfähigkeit möglich sind. Der Faktor Occurrence wird mit 6 eingeschätzt, da Infrastruktur- oder Deployment-Fehlkonfigurationen im industriellen Umfeld nicht ausgeschlossen werden können. Der Faktor Detection wird mit 6 bewertet, da Schwächen in Backup- oder Rollback-Strategien häufig erst im Störfall oder bei Belastungstests sichtbar werden. Daraus ergibt sich ein RPN von 288 ($RPN = Severity \times Occurrence \times Detection = 8 \times 6 \times 6 = 288$).

Risk Group	Phase	Task	Failure Mode	S	O	D	RPN
Product characteristics (financial risk, reputational risk)	Deployment	CI/CD for ML (MLOps)	Insufficient Redundancy & Failover Mechanisms	8	6	6	288
Regulatory requirements (product characteristics, financial risk)	AI-Modelling	Experiment Tracking & Management	Inconsistent Model Reproducibility	9	5	7	315

Abbildung 7-11: Bewertung der Fehlerzustände im Beispiel

7.3.4 Schritt 2 Werkzeugbewertung und Ermittlung des Qualifizierungsbedarfs

Im zweiten Schritt wird auf Basis der zuvor bewerteten Fehlerzustände untersucht, welche Bedeutung das eingesetzte Entwicklungswerkzeug für das Zustandekommen dieser Fehler besitzt und welches Vertrauensniveau in das Werkzeug erforderlich ist, um die identifizierten Risiken angemessen zu beherrschen.

7.3.4.1 Unterschritt 2.1 Ermittlung der Bedeutung des Werkzeugs für die Fehlerzustände (Tool Impact)

Im Hinblick auf die fehlende Reproduzierbarkeit von Trainings- und Modellständen besitzt das fiktive Werkzeug *MLtoolExample* einen hohen Tool Impact (TI=2), da das Werkzeug unmittelbar für die strukturierte Erfassung, Versionierung und Speicherung von Parametern, Modellartefakten und Umgebungsinformationen verantwortlich ist. Werden Logging- oder Versionierungsmechanismen fehlerhaft konfiguriert oder unvollständig genutzt, entstehen Defizite in der Entwicklungsdokumentation. Diese wirken sich zunächst technisch auf die Rekonstruierbarkeit einzelner Trainingsläufe aus, führen jedoch in einem weiteren Schritt zu einem Dokumentationsproblem auf Anwendungsebene. Da die Nachweisführung gegenüber internen und externen Audits maßgeblich auf der Vollständigkeit dieser Informationen beruht, kann ein Werkzeugfehler unmittelbar die regulatorische Konformität des gesamten Analysewerkzeugs beeinträchtigen. Der Tool Impact ist daher als hoch einzustufen.

Beim Fehlerzustand unzureichender Redundanz- und Failover-Mechanismen ist der Tool Impact differenziert zu betrachten. Das fiktive Werkzeug *MLtoolExample* ist zwar Bestandteil der CI/CD- und Deployment-Pipeline und unterstützt die Verwaltung von Modellversionen sowie deren Bereitstellung für unterschiedliche Betriebsumgebungen, trägt jedoch nur indirekt zur Gesamtrobustheit bei der Ausfallsicherheit bei. Die eigentlichen Mechanismen zur Sicherstellung der Systemverfügbarkeit, z. B. Redundanz, Container-Orchestrierung, Monitoring sowie automatisierte Rollback- oder Fall-back-Prozesse, werden in der Regel durch die zugrunde liegende Deployment- und Infrastrukturplattform bereitgestellt. Das fiktive Werkzeug *MLtoolExample* stellt in diesem Zusammenhang lediglich die organisatorische Grundlage zur Verwaltung der Modellstände bereit, ohne selbst die

operative Failover-Logik zu implementieren. Da somit mehrere unabhängige technische und organisatorische Schutzmechanismen existieren, die die Robustheit der Backup- und Wiederherstellungsstrategie gewährleisten, ist der direkte Einfluss eines Werkzeugfehlers auf das Auftreten des betrachteten Fehlerzustands begrenzt. Der Tool Impact wird daher als niedrig eingestuft (TI=1), da das fiktive Werkzeug *MLtoolExample* zwar zur Nachvollziehbarkeit und Bereitstellung von Modellversionen beiträgt, jedoch nicht primär die Stabilität oder Verfügbarkeit der zugrunde liegenden Systemarchitektur bestimmt.

7.3.4.2 Unterschritt 2.2 Bewertung der Erkennbarkeit des Werkzeugfehlers und Bestimmung des Tool Confidence Level (TCL)



Abbildung 7-12: Ableitung des Tool Confidence Level (TCL)

Werkzeugfehler im Bereich der Reproduzierbarkeit sind in der Praxis häufig nicht unmittelbar sichtbar. Unvollständige Protokollierung, fehlerhafte Versionierung oder inkonsistente Metadaten fallen typischerweise erst bei gezielten Reproduktionsversuchen, im Rahmen von Modellvergleichen oder bei internen und externen Audits auf. Solange kein konkreter Anlass zur Nachverfolgung besteht, kann ein Defizit in der Entwicklungsdokumentation unentdeckt bleiben. Da das fiktive Werkzeug *MLtoolExample* hier eine

zentrale Rolle bei der Sicherstellung der Entwicklungsnachweise einnimmt und der Tool Impact als hoch bewertet wurde (TI=2), führt die eingeschränkte unmittelbare Erkennbarkeit von Werkzeugfehlern (TD=3) zu einem erhöhten Tool Confidence Level (TCL=3) in der Hauptrisikogruppe „Regulatorik“ (siehe zweite Zeile in der Tabelle in Abbildung 7-13). Das erforderliche Vertrauensniveau in das Werkzeug ist entsprechend hoch, da ein Dokumentationsdefizit transitiv die regulatorische Konformität der gesamten Analyseanwendung gefährden kann.

Im Zusammenhang mit unzureichenden Redundanz- und Failover-Mechanismen ist die Erkennbarkeit von Werkzeugfehlern begrenzt. Schwächen in Rollback-Konfigurationen, Modellkennzeichnungen oder Backup-Pfaden werden häufig erst im tatsächlichen Störfall oder im Rahmen gezielter Ausfalltests sichtbar. Solange keine entsprechenden Tests durchgeführt werden, können Defizite in der Deployment- und Wiederherstellungslogik unentdeckt bleiben. Die Entdeckbarkeit möglicher Werkzeugfehler ist als TD=2 (mittel) einzustufen, da Konfigurations- oder Integrationsprobleme typischerweise erst im Zusammenspiel mit anderen Komponenten der Deployment-Pipeline sichtbar werden. Aus der Kombination von niedrigem Tool Impact (TI=1) und mittlerer Fehlererkennbarkeit (TD=2) ergibt sich ein entsprechend geringer Tool Confidence Level (TCL=1) für die Hauptrisikogruppe „Produkteigenschaften“ (siehe erste Zeile in der Tabelle in Abbildung 7-13).

Risk Group	Phase	Task	Failure Mode	TI	TD	TCL
Product characteristics (financial risk, reputational risk)	Deployment	CI/CD for ML (MLOps)	Insufficient Redundancy & Failover Mechanisms	1	2	TCL1
Regulatory requirements (product characteristics, financial risk)	AI-Modelling	Experiment Tracking & Management	Inconsistent Model Reproducibility	2	3	TCL3

Abbildung 7-13: Ermittlung des Tool Confidence Level im Beispiel

7.3.4.3 Unterschritt 2.3: Abschließende Werkzeugbewertung und Handlungsempfehlung auf Basis der RPN und des TCL

Die abschließende Bewertung des Werkzeugs ergibt sich aus der Kombination der im ersten Schritt ermittelten Risikoprioritätskennzahlen (RPN) und der im zweiten Schritt bestimmten Tool Confidence Level (TCL). Während die RPN die inhärente Kritikalität des Fehlerzustands im Prozesskontext beschreibt, gibt der TCL an, welches Maß an Vertrauen und Absicherung für das Werkzeug erforderlich ist.

Für die Entwicklungsaufgabe „Experiment Tracking & Management“ wurde der Fehlerzustand „Inkonsistente Modellreproduzierbarkeit“ mit einem hohen RPN von 315 bewertet. In Verbindung mit einem hohem Tool Impact (TI=2) und einer hohen Erkennbarkeit (TD=3) von Werkzeugfehlern ergibt sich ein hoher TCL (TCL=3) in der Hauptrisikogruppe „Regulatorik“ (siehe zweite Zeile in den Tabellen in Abbildung 7-11 und Abbildung 7-13). Daraus folgt, dass das Werkzeug zwingend Funktionen bereitstellen muss, die eine vollständige Erfassung und Versionierung aller experimentrelevanten Informationen gewährleisten. Hierzu zählen insbesondere die systematische Protokollierung von Random Seeds (Startwerte für Zufallszahlengeneratoren zur Reproduzierbarkeit von Trainingsergebnissen) und

Umgebungsparametern, die Versionierung von Modellartefakten und Hyperparametern sowie die konsistente Erfassung von Abhängigkeiten wie Bibliotheksversionen. Zusätzlich sind organisatorische Maßnahmen erforderlich, etwa verbindliche Vorgaben zur Seed-Fixierung und eindeutige Versionskennzeichnungen aller Experimente. Das Werkzeug muss diese Anforderungen technisch unterstützen, indem es automatische Artefaktversionierung und lückenlose Protokollierung ermöglicht.

Für die Entwicklungsaufgabe „CI/CD for ML (MLOps)“ wurde der Fehlerzustand „Unzureichende Redundanz- und Failover-Mechanismen“ mit einem RPN von 288 bewertet. In Kombination mit einem niedrigen Tool Impact (TI=1) und einer mittleren Erkennbarkeit von Konfigurationsfehlern (TD=2) ergibt sich hier ein niedriger TCL (TCL=1) in der Hauptrisikogruppe „Produkteigenschaften“ (siehe erste Zeile in den Tabellen in Abbildung 7-11 und Abbildung 7-13). Daraus leitet sich die Anforderung ab, dass das Werkzeug die kontrollierte Bereitstellung mehrerer Modellversionen, definierte Rollback-Mechanismen sowie validierte Fallback-Strategien unterstützen muss. Insbesondere sind automatisierte Validierungsprüfungen vor der Produktivsetzung, klar definierte Deploymentszenarien sowie die Simulation von Ausfall- und Stressszenarien empfohlen. Das fiktive Werkzeug *MLtoolExample* sollte daher so in eine CI/CD-Infrastruktur eingebunden sein, dass konditionale Deployments, Rollback-Trigger und die parallele Validierung mehrerer Modellstände unterstützt werden. Dadurch kann sichergestellt werden, dass bei Ausfall ein geprüfter und freigegebener Ersatzmodellstand verfügbar ist.

Die kombinierte Betrachtung von RPN und TCL zeigt, dass beide Fehlerzustände trotz unterschiedlicher Risikoperspektiven unterschiedliche Anforderungen an die Absicherung des Werkzeugs stellen. Beim ersten Fehlerzustand steht die Nachweisfähigkeit der Entwicklungsstrenges im Vordergrund, da Defizite in der Protokollierung und Versionierung unmittelbar die regulatorische Konformität der Analyseanwendung beeinträchtigen können. Daraus ergibt sich ein erhöhter Absicherungsbedarf des Werkzeugs im Hinblick auf Nachvollziehbarkeit und Dokumentation.

Beim zweiten Fehlerzustand liegt der Schwerpunkt hingegen auf der Resilienz und Beherrschbarkeit des Betriebszustands. Da der direkte Einfluss des Werkzeugs auf die Systemverfügbarkeit begrenzt ist und zusätzliche

infrastrukturelle Sicherheitsmechanismen existieren, ergibt sich hier ein geringerer werkzeugspezifischer Absicherungsbedarf. Die Handlungsempfehlung besteht daher darin, das Werkzeug im ersten Fall gezielt zu qualifizieren, beispielsweise durch definierte Testfälle, Referenzdatensätze oder eine dokumentierte Validierung der Werkzeugergebnisse, und durch verbindliche Konfigurations- und Dokumentationsrichtlinien abzusichern. Im zweiten Fall sind vor allem eine korrekte Integration in die bestehende CI/CD- und Deployment-Infrastruktur sowie regelmäßige Tests der Failover-Mechanismen sicherzustellen.

7.4 Zusammenfassende Einordnung im Kontext der Werkzeugqualifizierung

Die Methode ermöglicht eine ganzheitliche Betrachtung von Prozess- und Werkzeugrisiken entlang des KI-Lebenszyklus. Sie unterstützt Entwicklerinnen und Entwickler dabei, Risiken nicht nur zu erkennen, sondern gezielt auf die Zuverlässigkeit und Vertrauenswürdigkeit der eingesetzten Werkzeuge zu beziehen.

Das integrierte Beispiel verdeutlicht die Logik der risikobasierten Werkzeugbewertung im Qualitätsmanagement. Ausgangspunkt ist nicht das Werkzeug selbst, sondern die Identifikation konkreter Fehlerzustände im Kontext definierter Entwicklungsaufgaben. Erst auf dieser Grundlage wird analysiert, welchen Einfluss das Werkzeug auf das Entstehen oder Beherrschen dieser Fehler besitzt und wie gut potenzielle Werkzeugfehler erkannt werden können.

Die beiden betrachteten Fehlerzustände zeigen, dass unterschiedliche Risikodimensionen adressiert werden, obwohl dieselbe Entwicklungsaufgabe und dasselbe Werkzeug betrachtet werden. In beiden Fällen ergibt sich aufgrund des hohen Tool Impact und der nur eingeschränkt unmittelbaren Fehlererkennbarkeit ein erhöhter Tool Confidence Level.

Damit wird deutlich, dass der Qualifizierungsbedarf nicht abstrakt aus der Art des Werkzeugs abgeleitet wird, sondern aus der systematischen Verknüpfung von Fehlerzustand, Risikogruppe, Prozesskontext und Erkennbarkeit von Werkzeugfehlern.

Qualitätsmanagement in der Automobilindustrie

Den aktuellen Stand der veröffentlichten VDA-Bände zum Qualitätsmanagement in der Automobilindustrie finden Sie im Internet unter <https://www.vda-qmc.de>.

Auf dieser Homepage können Sie auch direkt bestellen.

Bezug:

Verband der Automobilindustrie e. V. (VDA)

Qualitäts Management Center (QMC)

10117 Berlin, Behrenstr. 35

Telefon +49 (0) 30 89 78 42-235, Telefax +49 (0) 30 89 78 42-605

E-Mail: info@vda-qmc.de, Internet: www.vda-qmc.de

VDA QMC

Verband der Automobilindustrie
Qualitäts-Management-Center