

VDA QMC

German Association of the Automotive Industry
Quality Management Center

Quality Management in the Automotive Industry

Artificial Intelligence in Quality Management

1st edition, March 2026
Online download document

Quality Management in the Automotive Industry

Artificial Intelligence in Quality Management

1st edition, March 2026

Online download document

Verband der Automobilindustrie e. V. (VDA)

ISSN 0943-9412

Copyright 2026 by

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

[Online download document](#)

Non-binding VDA recommendation

The German Association of the Automotive Industry (VDA) recommends that its members apply the following VDA volume when introducing and maintaining QM systems.

Exclusion of liability

VDA publications are recommendations available for general use. Anyone who implements them is responsible for ensuring that they are used correctly in each case.

This VDA publication is based on state-of-the-art technical procedures, current at the time of issue. Implementation of VDA recommendations does not absolve anyone of liability for their actions. In this respect, everyone acts at their own risk.

The VDA and those involved in VDA recommendations shall bear no liability.

If during the use of VDA recommendations, errors or the possibility of misinterpretation are found, it is requested that the VDA be notified immediately so that any possible faults can be corrected.

Copyright

This publication is protected by copyright. Any use outside of the strict limits of copyright law is not permissible without the consent of the VDA and is liable to prosecution. This applies in particular to copying, translation, micro-filming and the storing or processing in electronic systems.

Translations

The German document is the original. In case of questions of interpretation in other language versions, reference should be made to the German version as the original. This publication will also be issued in other languages. The current status must be requested from VDA QMC.

Preface

Artificial intelligence (AI) is no longer a vision of the future – it is a reality. In the automotive industry, it opens up enormous potential for efficiency, precision and predictive control. At the same time, its use raises new questions: How does AI work? What are the risks? How can we use and approve it responsibly and in a way that ensures the necessary quality? And what are the key skills we need to acquire?

This VDA volume "AI in Quality Management" offers a structured guide for specialists and managers in quality assurance, production, development, IT and data science. The aim is to provide an accessible introduction to the topic of AI – not through complex technical explanations, but through comprehensible terms, practical examples and a clear structure. Because only those who understand it can help to shape it responsibly.

Use this volume as a practical tool to become less afraid of using AI, to learn how to make informed assessments of risks, and to grow into a competent, responsible user of AI. The integration of AI into industrial processes is not a question of whether, but of how. This makes it all the more important to familiarize yourself with the basics at an early stage and to strengthen your ability to shape it effectively.

Table of Contents

Preface	4	
Table of Contents	5	
1	Introduction	9
2	Terminology	10
2.1	Introduction	10
2.2	References to existing standards	10
2.2.1	DIN EN ISO/IEC 22989 – Concepts and Terminology for AI	10
2.2.2	EU AI Act – Requirements for quality management	11
2.2.3	ISO 9001 / IATF 16949 / VDA 6.x – Quality management in the automotive industry	11
2.3	Basic terms in artificial intelligence	12
2.3.1	Artificial intelligence (AI), machine learning (ML) and deep learning (DL)	12
2.3.2	Training data, models, inference and agents	13
2.3.3	Supervised, semi-supervised, unsupervised and reinforcement learning	14
2.3.4	Natural language processing (NLP), language models and retrieval-augmented generation (RAG)	15
2.3.5	AI-powered dialog systems (chatbots)	15
2.4	AI-specific terms in quality management	16
2.4.1	Detection of anomalies	17
2.4.2	Auditability	18
2.4.3	Bias	19
2.4.4	Black box	20
2.4.5	Confidence score	21
2.4.6	Drift	22
2.4.7	Explainability	23
2.4.8	Fairness	24
2.4.9	Ground truth	25

2.4.10	Hallucination	26
2.4.11	Causal model	27
2.4.12	Predictive quality	28
2.4.13	Prescriptive quality	29
2.4.14	Robustness	30
2.4.15	Trustworthiness	31
2.5	Regulatory terms in the context of AI systems	32
2.5.1	High-risk AI system	32
2.5.2	Conformity assessment	33
2.5.3	Data quality and data governance	33
2.5.4	Auditability, traceability and transparency	34
2.5.5	Non-high-risk AI systems	34
2.6	Glossary for AI in quality management	35
3	Successfully using AI in QM	36
3.1	Mindset, work culture, motivation and change management	39
3.2	Skills and competences	42
3.3	Data	43
3.4	Organizational structures and processes	44
3.5	Governance, data protection, standards and regulations	45
3.6	Infrastructure	46
3.7	Technical interfaces and integration	47
3.8	Potential applications, use cases, tools and methods	48
4	AI competences in QM	50
4.1	Main competences for AI in quality management	50
4.2	Roles in quality management and relevant AI competences	53
4.2.1	Classical roles in quality management	53
4.2.2	New roles in quality management	54
4.2.3	Role-specific AI competences	55
5	Approving AI systems in QM	57
5.1	Step 1: Determining the project risk class	58

5.2	Step 2: Risk assessment of the AI system	64
5.2.1	Overview of the key questions on the assessment-relevant requirements in the process phases	64
5.2.2	Assessment-relevant requirements for AI systems	67
6	Recommended actions and application examples	93
6.1	AI-powered optical quality control	95
6.1.1	Example	99
6.2	Rule-oriented AI agents to support the 8D process	102
6.2.1	Example	105
6.3	AI-powered audit	111
6.3.1	Example	113
6.4	AI-powered FMEA	116
6.4.1	Example	118
6.5	Predictive process control	121
6.5.1	Example	124
6.6	Preventive maintenance	128
6.6.1	Example	131
6.7	Field data analysis	135
6.7.1	Example	138
6.8	Review of development work products	141
6.8.1	Example	142
6.8.2	Example	143
6.8.3	Notes	145
6.9	VDA chatbot	145
6.9.1	Example	147
6.9.2	Notes	149
6.10	Speech mining for work instructions	149
6.10.1	Example	151
6.10.2	Note	152
6.11	Comparing documents	153
6.11.1	Example	155
6.11.2	Notes	157
6.12	Interactive learning	157

7	Excursion: Risk-based assessment of AI development tools	161
7.1	Explanation of the basic concepts	162
7.2	Performing the risk-based assessment of AI development tools	165
7.2.1	Step 1: Identify risks for selected development tasks	167
7.2.2	Step 2: Tool evaluation and determination of the qualification requirements	170
7.2.3	List of potential error states (example)	173
7.3	Example application of the method	176
7.3.1	Context of the AI-powered SPC evaluation in quality management	177
7.3.2	Context of the fictitious development tool <i>MLtoolExample</i>	177
7.3.3	Step 1: Identify risks for development tasks	177
7.3.4	Step 2: Tool evaluation and determination of the qualification requirements	181
7.4	Overall classification in the context of tool qualification	186

1 Introduction

This volume deals with the practical use of artificial intelligence in quality management as a tool and practical aid. The focus is on describing specific AI systems in quality management, their approval and how to effectively reduce their risks in use.

With identical requirements or requests made to AI systems, their outputs or results may vary. They must therefore be validated by appropriate and traceable mechanisms, such as defined rules, defined limits and guaranteed response times.

This volume is divided into the following chapters, each of which highlights key aspects of AI use in quality management:

- **"Terminology"** introduces the key concepts in AI technology and situates them within the context of quality management.
- **"Successfully using AI in QM"** describes the use of artificial intelligence in companies, taking into account the applicable regulatory requirements and relevant standards. This volume provides an overview of methods, standards and potential applications along the product life cycle.
- **"AI competences in QM"** describes the roles and competences required for the use of artificial intelligence in quality management.
- **"Approving AI systems in QM"** describes a method for determining the risks of an AI system as the basis for granting its approval.
- **"Recommended actions and application examples"** defines, in the form of recommended actions, how AI applications can be operated, adapted and their outputs properly interpreted and evaluated.

In addition to this core content, the excursion **"Risk-based assessment of AI development tools"** describes how to conduct a risk-based assessment of AI development tools.

This volume does not cover applications relating to vehicle functions with an AI component or the practical implementation of the EU AI Act.

This volume provides technical guidance and does not replace any legal assessment or regulatory classification of AI systems.

2 Terminology

2.1 Introduction

The progressive integration of Artificial Intelligence (AI) into quality-relevant processes is changing not only methods and tools, but also the language of quality management. This brings new challenges: Terms and concepts have different meanings in the context of AI, are interpreted differently or are not clearly embedded in existing QM methods.

The aim of this chapter is to establish a practical and uniform understanding of terms and concepts for the use of AI in quality management. It is intended as a supplement to existing standards, in particular DIN EN ISO/IEC 22989 ("Artificial Intelligence – Concepts and Terminology") and to quality-relevant standards such as ISO 9001, IATF 16949 and VDA 6.x. It applies key AI concepts to the context of automotive quality management and provides concrete application examples, defines boundaries and offers instructions for operationalization.

The aim is to facilitate communication between departments through a common language, to improve the auditability of AI systems and to lay the foundation for future standards and regulatory requirements.

2.2 References to existing standards

2.2.1 DIN EN ISO/IEC 22989 – Concepts and Terminology for AI

DIN EN ISO/IEC 22989:2022 defines over 100 terms and concepts related to artificial intelligence and provides an internationally agreed basis for talking about AI systems. Among other things, it describes the life cycle of AI systems, the roles of involved actors (e.g. providers, users, developers) as well as central concepts such as bias, explainability, training, inference and models.

This standard is an important reference point for quality management in the automotive industry, as it provides a systematic definition of terms that can serve as a basis for the assessment, auditing and further development of AI applications. However, it deliberately remains industry-neutral and offers no specific application examples or interpretations for quality-relevant processes in the automotive industry – which is where this volume comes in.

2.2.2 EU AI Act – Requirements for quality management

The European AI Act (EU 2024/1689) obliges providers and manufacturers to implement specific quality management requirements, depending on the risk class of the system. For high-risk systems (see chapter 2.5.1) particularly extensive requirements apply, but even systems with lower risk classes must meet certain requirements from the regulation, such as transparency, robustness or data quality. Article 17¹ of the Regulation requires a documented and systematic quality management system (QMS) for high-risk systems or the integration of the requirements into an existing QMS.

These include but are not limited to:

- Regulatory compliance strategies
- Methods for the development, testing and validation of AI systems
- Data management processes (including data quality, labeling, storage)
- Risk management and post-market monitoring
- Traceable documentation and communication processes
- Monitoring of the system during operation

These requirements overlap in many respects with existing quality management systems based on ISO 9001 or IATF 16949, but go much further in terms of AI-specific aspects such as data processing, model validation and algorithmic transparency. This volume is intended to help translate these requirements into the language and practice of automotive quality management.

2.2.3 ISO 9001 / IATF 16949 / VDA 6.x – Quality management in the automotive industry

The ISO 9001:2015 standard forms the basis for quality management systems worldwide. It emphasizes customer focus, risk-based thinking and continuous improvement. IATF 16949:2016 complements these requirements with automotive-specific aspects, including:

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

- Product safety and traceability
- Error prevention instead of error detection
- Supplier development and evaluation
- Process capability analyses and FMEA

In addition, VDA volumes 6.x further specify the requirements on systems, processes and products for the German automotive industry.

With the use of AI in quality-relevant processes – for example, for automated error classification, anomaly detection or predicting quality deviations – these established systems now require new forms of regulation, including:

- Auditability of AI models during system, process and product audits
- Obligation to declare automated decisions (e.g. classification of scrap)
- Evaluation of data quality and origin as part of process and product safety
- Integration of AI into existing audit strategies, service assessments and supplier assessments

2.3 Basic terms in artificial intelligence

This chapter presents key terms and concepts in AI that are relevant for quality management in the automotive industry. The aim is to create a common understanding that facilitates communication between quality management, data science, IT and production.

2.3.1 Artificial intelligence (AI), machine learning (ML) and deep learning (DL)

Artificial Intelligence (AI) refers to machine-based systems designed for varying levels of autonomous operation, which infer from inputs how outputs, such as predictions, content, recommendations, or decisions, are produced that may influence physical or virtual environments (see Art. 3 No. 1

EU AI Act). In the context of this volume, the term particularly includes systems that perform tasks typically requiring human intelligence – such as text and image generation, pattern recognition, or decision-making.

Machine learning (ML) is a field of artificial intelligence in which systems learn from data without being explicitly programmed. ML models recognize patterns and make predictions based on historical data.

Deep learning (DL) is a special form of machine learning based on artificial deep neural networks. DL is often used for complex tasks such as image or speech recognition.

Relationship to quality management (QM): ML models can be used in quality control, e.g. to classify failure patterns, detect anomalies in process data or predict quality deviations.

2.3.2 Training data, models, inference and agents

Training data is structured or unstructured data² that is used to "train" an AI model. The quality and representativeness of this data is critical to the subsequent performance of the model.

Note: Training data also typically includes the "ground truth," which is critical for training and validating the model later. Ground truth is explained separately in chapter 2.4.9.

A trained AI **model** refers to the mathematical structure that is able to make predictions or support decisions after training.

Inference is the process by which a trained model is applied to new, unknown data to produce a prediction or classification.

Agents in AI systems are autonomous or semi-autonomous software components that perform tasks, make decisions or interact with their environment. They use trained models to process information and take targeted action – for example, by recognizing patterns, making decisions or initiating actions. In complex applications, multiple agents can communicate and collaborate with each other to achieve an overarching goal.

² "Structured data" means clearly organized information, usually in tabular form or in databases, e.g. measurements with time stamps. "Unstructured data" is less formalized and can take the form of images, text, audio recordings or videos, for example.

Relationship to quality management (QM): In quality management, trained models can be used, for example, in visual end-of-line testing to automatically decide whether a component meets the quality requirements. Agents can also be used in production monitoring to continuously analyze data, detect anomalies and automatically initiate quality assurance measures.

2.3.3 Supervised, semi-supervised, unsupervised and reinforcement learning

Supervised learning: A model is trained on data that has been labeled with known outcomes, such as "OK" or "NOK"³.

Unsupervised learning: A model is trained on data without labels in order to detect patterns or structures – for example, clustering in process data.

Semi-supervised learning: The model is trained using a combination of labeled and unlabeled data. This is particularly useful when manually labeling large amounts of data is expensive or time-consuming. The model uses the few labels there are to recognize patterns even in the unlabeled data.

Reinforcement learning: The agent learns by trial and error by receiving feedback from its environment and from the operator, e.g. to optimize testing strategies.

Relationship to quality management (QM): Supervised learning is particularly relevant for classification tasks, e.g. in automated component evaluation. Semi-supervised learning can be used if only some of the quality data has been classified manually, for example in the case of new product variants. Unsupervised learning is useful for anomaly detection, for example with sensor readings. Reinforcement learning can be used to optimize testing strategies or to dynamically adapt quality processes.

³ "OK" means a product or component that meets the quality requirements. "NOK" stands for "not OK" and indicates a product or component that does not meet the quality requirements.

2.3.4 Natural language processing (NLP), language models and retrieval-augmented generation (RAG)

Natural language processing (NLP) is a branch of artificial intelligence that deals with the automated processing, analysis and generation of natural language. The aim is to enable machines to understand human language and respond to it appropriately – both in written and spoken form.

Language models are AI-powered models that are trained to understand, analyze or generate language. This class of models has existed for many years. **Large language models (LLMs)** are an advanced and much larger form of these language models. They are trained on very large volumes of text and use deep neural networks to generate context-related answers, analyze texts or create new content. LLMs form the basis for generative AI applications (GenAI).

Retrieval-augmented generation (RAG) augments the capabilities of LLMs by combining the generative component with targeted information retrieval from external data sources. While an LLM is based on the knowledge it was trained on, RAG provides access to current, verified content – e.g. QM databases, standard documents or complaint histories. This allows answers to be not only linguistically plausible, but also correct in terms of content and context.

Relationship to quality management (QM): In quality management, NLPs, LLMs and RAGs can be used to efficiently process text-based information such as test reports, complaints or standard documents. They facilitate automated reporting, root cause analysis and interactive assistance systems (e.g. chatbots) that support professionals in processing 8D reports or analyzing the cause of errors. RAG is particularly relevant when AI systems need to access up-to-date, validated QM data in order to provide standard-compliant formulations or sound proposals for action.

2.3.5 AI-powered dialog systems (chatbots)

AI-powered dialog systems – often referred to as **chatbots** – are a special form of verbal assistance systems based on artificial intelligence. In contrast to rule-based chatbots, which are based on predefined decision trees and fixed response patterns, AI-powered systems react flexibly and

contextually to user input. They enable interactive communication with users and can take on a variety of tasks in quality management, such as processing complaints, analyzing the root cause of errors or documenting actions taken.

Technological fundamentals such as natural language processing (NLP), large language models (LLMs) and retrieval methods (e.g. retrieval-augmented generation (RAG) are explained in chapter 2.3.2. NLP enables language comprehension, LLMs generate contextual responses based on large volumes of text, and RAG complements this capability by providing access to current, validated data sources such as QM databases or standards documents, so that responses are not only linguistically plausible, but also correct in terms of content. In brief: NLP = understand language, LLM = generate answers, RAG = integrate company knowledge.

Relationship to quality management (QM): In quality management, such systems offer potential for increasing efficiency, especially in repetitive tasks, or in assisting professionals in complex decision-making processes. At the same time, they place new requirements on validation, traceability and acceptance: The boundary between assistance and automated decision-making must be clearly defined, in particular when proposed actions or root cause analyses are generated. An AI-powered chatbot can, for example, assist in processing 8D reports, identify similar error cases or suggest standard-compliant formulations. Successful use requires transparent documentation of the data sources, a clear distribution of roles between man and machine and training of users in the use of AI applications in quality management. **Important:** In generative systems, there is a risk of "hallucinations" – that is, the generation of plausible-sounding but factually incorrect content. This risk must be taken into account during validation and operation (see chapter 2.4.10 "Hallucination"). In addition, there are risks such as misinterpretations, bias in the data or lack of traceability of the answers. These aspects must be taken into account during design and operation.

2.4 AI-specific terms in quality management

This chapter explains key terms from AI that have a special meaning in the context of quality management. The aim is to create a common understand-

ing of terms that are used often in practice but interpreted differently – especially at the interfaces between data science, production and quality assurance.

Each term is explained with a definition, a reference to QM-specific applications and, if necessary, with references to standards or regulatory requirements.

2.4.1 Detection of anomalies

Definition (ISO/IEC 22989):

Detection of anomalies is the identification of data points, patterns or events that deviate from the expected norm. It can be based on statistical methods, machine learning or hybrid approaches.

Relevance in the QM context:

In quality monitoring, anomaly detection can help to detect unusual process flows, faulty components or sensor deviations at an early stage – often before a classical rule violation or scrap occurs – and can therefore be useful in statistical process control (SPC). It is particularly useful in complex, data-rich processes where classical threshold logic reaches its limits.

Typical challenges:

- Process instabilities (e.g. tool wear, temperature fluctuations)
- Sensor errors or calibration issues
- Human intervention or operator error
- New or rare failure patterns not included in the training

Safeguarding measures:

- Select suitable algorithms (e.g. isolation forest, autoencoder, statistical procedures)
- Combine with domain knowledge to prevent false positives
- Establish a feedback process for continuous improvement
- Visualize and contextualize anomalies for specialist users

Related terms:

Predictive quality, drift, data quality, SPC, process monitoring

2.4.2 Auditability**Definition:**

Auditability describes the ability to traceably document, evaluate and review decisions, processes and outputs of an AI system – both internally and by external bodies (e.g. customers, certification bodies, government authorities).

Relevance in the QM context:

Auditability is a key principle in the automotive industry – for example, in the context of system audits (ISO 9001, IATF 16949), process audits (VDA 6.3) and product audits (VDA 6.5), as well as in supplier audits. When AI systems make quality-relevant decisions or decision proposals (e.g. scrap classification, supplier evaluation), these decisions must be traceable, explainable and documented. This is also a key requirement in the EU AI Act for high-risk AI systems.

Typical challenges:

- Missing or incomplete documentation of training data, model versions or decision logic
- Use of "black box" models without explainable decision-making paths
- No clear responsibilities for AI systems in the QM system
- Lack of integration into existing audit processes

Safeguarding measures:

- Introduction of AI life cycle management with versioning, documentation and traceability
- Using explainability methods to facilitate auditability
- Anchoring AI systems in the quality management manual
- Training auditors in the use of AI systems
- Utilizing audit trails, logging and automated documentation

Related terms:

Explainability, transparency, re-training, validation, EU AI Act

Reference to chapter 2.5.4 Auditability:

Chapter 2.4.2 describes "auditability" in the context of quality management – i.e. the ability to document AI-powered decisions within the framework of QM audits (e.g. according to VDA 6.3) in a traceable manner.

In chapter 2.5.4, auditability is treated in the regulatory context of the EU AI Act, where it is defined as a mandatory requirement for high-risk systems.

The two perspectives complement each other and should be clearly distinguished in language use by having QM documentation explicitly refer to the audit objectives (system, process, and product audits) and by separately addressing regulatory requirements under the term "auditability" in compliance documents.

2.4.3 Bias**Definition (ISO/IEC 22989):**

Bias refers to a systematic distortion in data, models or decision-making processes that can lead to incorrect or unfair outputs.

Relevance in the QM context:

In quality management, bias can occur in various forms. One example is data bias or label bias, where training data for an AI model for error classification comes predominantly from one specific production line, resulting in poorer detection of errors on other lines. This can lead to distorted error detection and significantly impair the effectiveness of corrective actions.

Likewise, measurement methods themselves may be distorted (measurement or feature bias), model architectures may privilege certain patterns (algorithmic bias) or new distortions may result from feedback effects during operation (feedback loop/deployment bias).

Typical challenges:

- Data bias and label bias: Non-representative data (e.g. day shift only), historical error classifications with human bias, imbalance in error distribution (class imbalance).
- Measurement bias and feature bias: Features or measurement methods are themselves distorted (e.g. sensors with systematic deviation, unsuitable feature selection).

- Algorithmic bias: Model architecture or loss functions disadvantage certain groups or are not suitable for the intended use.
- Feedback loop / deployment bias: The model influences the data it later processes itself (for example, in predictive maintenance, where decisions alter future data distributions).

Safeguarding measures:

- Data analysis for representativeness and quality.
- Use of fairness metrics and bias detection methods.
- Validation by independent QM teams.
- Inspection of the measurement methods and feature selection for distortions.
- Review of model architecture and loss functions for fairness and suitability.
- Post-deployment monitoring to detect and correct feedback loops.

Differentiation from bias in QM test process management:

In quality management, the term "bias" is traditionally used for systematic measurement deviations. The VDA QMC Glossary defines bias as:

"bias/BI of the measurement, estimate of a systematic measurement error."

This definition refers to the accuracy of measurement processes and must be clearly distinguished from the bias discussed here in the context of AI.

Related terms:

Fairness, ground truth, data drift

2.4.4 Black box

Definition:

A black box is an AI system whose internal functioning is not transparent or traceable to users. The inputs and outputs are known, but the decision logic remains hidden.

Relevance in the QM context:

Black box models (e.g. complex neural networks) make auditability and root cause analysis difficult in the event of complaints. They stand in contrast to explainable models, and they are of particular importance in safety-relevant or quality-relevant applications.

Typical challenges:

- Lack of transparency for auditors and QM teams.
- Difficult root cause analysis.
- Regulatory risks in high-risk systems (EU AI Act).

Safeguarding measures:

- Use of XAI methods to partially open the black box.
- Documentation of the model architecture and decision logic.
- Combining black box models with explainable components (hybrid approaches).

Related terms:

Explainability, transparency, auditability.

2.4.5 Confidence score

Definition:

A confidence score indicates how likely a model is to trust its own prediction.⁴ It is a measure of the uncertainty of the decision: the higher the score, the lower the uncertainty. **Important: Confidence score is not the same as accuracy. A model can predict with 99% confidence and still make the wrong decision.**

Relevance in the QM context:

For example, a model classifies a part as "OK" with 92% confidence. With low confidence scores the uncertainty is greater, so manual verification may be necessary – especially for safety-critical components. Uncertainty must

⁴ Note: The term "confidence score" is not currently formally defined in ISO/IEC standards, but it is widely used in industrial practice and in AI frameworks.

also be taken into account when assessing the suitability of the test process (see VDA Volume 5).

Typical challenges:

- Misinterpretation by users
- No thresholds defined
- Uncertainty not documented or not taken into account

Safeguarding measures:

- Establishment of verification thresholds taking into account the level of uncertainty.
- Visualization of scores and uncertainties in dashboards
- Combining with explainability methods for better interpretation
- Documentation of the uncertainty score for audits and for determining the suitability of the test process.

Related terms:

Uncertainty, decision logic, auditability

2.4.6 Drift

Definition (ISO/IEC 22989):

Drift is the change in data distributions or relationships over time that can affect the performance of an AI model. A distinction is typically made between data drift (change in input data) and concept drift (change in the relationship between the input and the target value).

Relevance in the QM context:

A model that has been trained on process data for a specific material or machine condition can become less accurate due to changes in production (e.g., material batch, tool wear, new shift crew). Drift can lead to misclassifications, delayed reactions or wrong decisions.

Typical challenges:

- Process changes (e.g. new batches, new lots (material), new suppliers, machine updates)

- Seasonal effects or shift changes
- Sensor aging or calibration deviations
- Changes in the test process or the preprocessing of the data

Safeguarding measures:

- Monitoring of model metrics over time (e.g. accuracy, error rate)
- Use of drift detection algorithms (e.g. population stability index, Kolmogorov-Smirnov test)
- Re-training or adjustment of the model in the event of significant drift
- Documentation and traceability of changes in the process environment

Drift in other QM contexts:

In classical quality management, "drift" is often used for physical or metrological changes, e.g. temperature fluctuations or the effects of wear on test equipment. This meaning differs from the drift that is discussed here in the context of AI, which refers to changes in data or model relationships.

Related terms:

Re-training, model validation, data quality, robustness

2.4.7 Explainability

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

Explainability refers to the ability to make an AI system's decisions understandable to humans.

Relevance in the QM context:

In audits or complaints, it must be clear why an AI system has classified a part as "NOK". Explainability is also a regulatory requirement for high-risk systems under the EU AI Act.

Typical challenges:

- Complex models (e.g. deep learning) are difficult to explain

- Lack of documentation of the decision logic
- Unexplainable visualizations or scores

Safeguarding measures:

- Use of explainable models (e.g. decision trees)
- Visualization of decision paths
- Integration of XAI methods (explainable AI)⁵

Related terms:

Auditability, confidence score, transparency, black box

2.4.8 Fairness

Definition (ISO/IEC 22989):

Fairness refers to the property of an AI system that prevents it from systematically disadvantaging certain categories of data, groups or stakeholders. It is closely linked to the concepts of bias and transparency.

Relevance in the QM context:

In quality-relevant applications, a lack of fairness can lead, for example, to certain processes, shifts or product variants being systematically evaluated as worse – not based on objective quality data but because of unbalanced training data or unreflective model logic.

Differentiation from bias:

- **Bias** refers to a distortion in the (training) data, models or processes and is **measurable** (e.g. by statistical analysis).
- **Fairness** describes the fairness of a model's outputs and is **assessable** (e.g. through fairness metrics such as "equal opportunity" or "demographic parity").

⁵ Explainable AI (XAI) refers to methods and techniques designed to make the decisions and inner workings of AI systems understandable and transparent to humans. The aim is to create transparency, promote trust and meet regulatory requirements (e.g. the EU AI Act), especially for complex models such as deep learning.

Example:

A model for loans can contain bias, because historical data includes more loans to men. Fairness means checking whether the model still ensures equal opportunities for men and women.

Typical challenges:

- Difficulty in verifying fairness in model output (e.g. equal opportunities for all groups).
- Selecting appropriate fairness metrics and interpreting them.
- Balancing fairness with other objectives (e.g. accuracy, efficiency).
- Transparently communicating the fairness assessment to auditors and stakeholders.

Safeguarding measures:

- Use of fairness metrics (e.g. equal opportunity, demographic parity)
- Analysis and balancing of training data
- Sensitivity analyses and counterexamples in model testing
- Documentation of assumptions and model limitations

Related terms:

Bias, data quality, supplier evaluation, explainability

2.4.9 Ground truth**Definition (ISO/IEC 25012):**

Ground truth refers to the reference data that is considered correct and is used to train or validate an AI model.

Relevance in the QM context:

In quality testing, the ground truth can be defined, for example, using manually tested failure patterns or measurement data from certified test equipment. The quality of the ground truth is decisive for the quality and later auditability of the model.

Note: The ground truth is also commonly used in training data, but it is not the same as the term "training data." While training data constitutes the entire dataset used for training a model, ground truth refers to the correct reference values that are crucial for training, validation and auditing (see chapter 2.3.2 "Training data, models, inference and agents").

Typical challenges:

- Incorrect or inconsistent labeling processes
- Incomplete or unrepresentative data
- Deviations between the reference data defined as correct (ground truth) and the actual process conditions, e.g. when production parameters change or new variants occur
- Lack or insufficient amount of training data required to create and validate a sound ground truth

Safeguarding measures:

- Use of expert labeling
- Validation by independent QM teams
- Documentation of the origin and quality of the ground truth

Related terms:

Bias, data quality, validation

2.4.10 Hallucination

Definition (ISO/IEC 22989):

Hallucination refers to the generation of outputs by an AI system that appear plausible but are factually incorrect, unfounded or not supported by the input data.

Relevance in the QM context:

In quality-critical applications, hallucination can cause an AI system to generate incorrect recommendations for action or classifications – for example, a faulty diagnosis in predictive quality or a wrong cause in a complaint analysis. This endangers process reliability and auditability.

Typical challenges:

- Lack of validation of generated content
- Users placing too much trust in AI outputs
- Use of models in contexts for which they have not been trained
- Inadequate dataset or prompt design in generative AI systems

Safeguarding measures:

- Use of verification mechanisms (e.g. cross-checks, plausibility checks)
- Limiting the use of generative AI to non-safety-critical areas
- Training users in the handling of AI outputs
- Combining with classical QM methods for validation (e.g. random sampling)

Related terms:

Uncertainty, confidence score, explainability, bias

2.4.11 Causal model**Definition:**

A causal model describes the causal relationships between variables in a system. These can be represented in the form of mathematical equations or visually as diagrams, both in qualitative and quantitative terms.

Relevance in the QM context:

These models can be used to understand the effects of controllable interventions and cause-and-effect relationships. When combined with statistical data, these models can facilitate the causal inference of relationships from data that go beyond mere associations. Some QM methods, such as FMEA and the Ishikawa analysis, are based on qualitative considerations of causal relationships.

Typical challenges:

- Development of complete and accurate models

- Challenges in testing and validating assumptions
- Incomplete or unrepresentative data

Safeguarding measures:

- Use of expert knowledge
- Visualization of effect chains
- Evaluation of models and sensitivity analyses

Related terms:

Models, inference, ground truth, explainability, transparency, auditability, drift

2.4.12 Predictive quality

Definition:

Predictive quality⁶ refers to the use of data analysis and machine learning for the early prediction of product and process quality criteria. The aim is to make decisions based on the predictions and to initiate measures before possible quality deviations lead to scrap, rework or customer complaints.

Relevance in the QM context:

Predictive quality enables proactive quality assurance: Instead of reacting to errors, companies can intervene proactively on the basis of process, machine or environmental data. This increases process stability, reduces costs and improves customer satisfaction.

Typical challenges:

- Missing or unstructured data history and data quality
- Complex, non-linear relationships between process parameters and quality criteria
- Low acceptance by departments due to the models' lack of explainability (black box effect)

⁶ Note: The term "predictive quality" is not currently defined in the relevant standards, but it is used in industrial practice to describe data-based methods for predicting quality deviations. It is used in the context of Industry 4.0, data-driven quality management and AI-powered process monitoring.

- Difficulty integrating into existing IT and production systems (e.g. MES, ERP)

Safeguarding measures:

- Build robust data pipelines and ensure data quality
- Select appropriate ML models, e.g. random forest, gradient boosting, long short-term memory (LSTM)
- Close cooperation between Data Science, Production and Quality Assurance
- Visualization of predictions and recommended actions for users
- Piloting with clear KPIs (e.g. scrap rate, false positive rate, early warning time, ROI)

Related terms:

Anomaly detection, bias, data quality, drift, explainability, process capability, regression analysis, re-training, condition monitoring

2.4.13 Prescriptive quality

Definition:

Prescriptive quality refers to the use of data analysis, machine learning and optimization methods to predict quality deviations and to derive concrete recommendations for action to prevent them from occurring. Unlike predictive quality, which only makes predictions, prescriptive quality combines predictions with proposals for preventive measures that can be implemented under human supervision.

Relevance in the QM context:

Prescriptive quality enables proactive and action-oriented quality assurance. Companies can not only recognize that there is a risk, but also how it can be avoided – e.g. by adjusting process parameters, taking maintenance measures or changing materials. This increases process stability, reduces scrap and rework and facilitates continuous improvement.

Typical challenges:

- High complexity in deriving recommendations from predictions.

- Ensuring that the proposed measures are feasible and practical.
- Acceptance by departments (trust in AI-generated recommendations).
- Integration into existing decision and control processes.

Safeguarding measures:

- Developing decision-making models that make recommendations transparent and traceable.
- Combining AI methods with domain knowledge to validate the recommendations.
- Clearly defining the role of human oversight in implementation.
- Pilot projects with measurable KPIs (e.g. reduction of scrap, improvement of process capability).
- Visualization of recommendations and their expected impact for users

Related terms:

Predictive quality, explainability, condition monitoring, process optimization, human oversight, re-training

2.4.14 Robustness

Definition (ISO/IEC 22989):

Robustness refers to the ability of an AI system to deliver stable and reliable outputs even with changing conditions, disruptions or outliers.

Relevance in the QM context:

For AI applications, robustness means the model's resistance to data noise, fluctuating environmental conditions and new variants. This term differs from the definition used in the VDA volume "Robust Production Process," which refers to the overall process stability – i.e. an operation that proceeds reliably according to plan, remains resilient to disruptions and delivers products on time and in the correct quantities in accordance with specifications (⇒ customer). In the context of AI, on the other hand, the term primarily refers to the performance of the model under varying input conditions.

The two concepts complement each other: A robust AI system supports but does not replace process robustness.

Typical challenges:

- Overfitting to training data
- Sensitivity to small data deviations
- Lack of testing under real production conditions
- Insufficient model maintenance in the event of process changes

Safeguarding measures:

- Use of robust model architectures and regularization
- Testing with faulty data, outliers and real process deviations
- Monitoring and continuous validation in operation
- Combining with classical QM methods (e.g. SPC)

Related terms:

Drift, re-training, validation, process capability, resilience

2.4.15 Trustworthiness

Definition (ISO/IEC 22989):

Trustworthiness refers to the property of an AI system that is designed to inspire justified trust among users and stakeholders. It includes aspects such as transparency, robustness, fairness, reliability and traceability.

Relevance in the QM context:

In quality management, trustworthiness is crucial for the acceptance of AI systems. It influences people's willingness to adopt AI-powered decisions and is a prerequisite for audits and regulatory compliance (e.g. EU AI Act for high-risk systems).

Typical challenges:

- Lack of transparency in complex models
- Unclear responsibilities for AI decisions

- Risks of hallucinations or bias
- Insufficient integration into existing QM processes

Safeguarding measures:

- Use of explainable models and documentation of decision logic
- Establishing governance structures for AI systems
- Validation and monitoring over the entire life cycle
- Training of users and auditors

Related terms:

Explainability, auditability, fairness, robustness, reliability

2.5 Regulatory terms in the context of AI systems

The increasing regulation of AI systems – in particular through the EU AI Act (EU 2024/1689) – brings with it new terms that must also be understood and used in quality management. This chapter presents key terms from the regulatory environment and explains their meaning in the context of AI. The aim is to create a common understanding of terms that facilitates communication, especially between departments, auditors and regulatory bodies.

Note on application: This chapter – as well as the entire volume – serves solely as technical and linguistic guidance for working with AI systems in quality management. It does not replace the need for legal assessment and regulatory classification. In order to classify a specific AI system, the intended use, the fundamental rights affected and the technical design must always be examined.

2.5.1 High-risk AI system

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

An AI system that poses a high risk to health, safety or fundamental rights.

Linguistic classification:

The term "high risk" is not to be understood as a technical attribute, but rather as a legal classification. It refers to the purpose and the application environment of a system.

Differentiation:

A technical system can be considered high-risk or non-high-risk depending on the context in which it is used.

Example formulation:

"The AI system for automated scrap classification falls under the category of high risk in accordance with Annex III to the AI Act."

2.5.2 Conformity assessment**Definition (EU AI Act, Art. 19–24):**

Procedure for verifying whether an AI system meets the requirements of the AI Act.

Note on linguistic use:

This term is familiar from product safety regulations (e.g., CE marking), but in the context of AI, it is applied to algorithmic systems.

Typical misunderstandings:

"Conformity assessment" is not the same as a classical QM audit – it also includes technical documentation, risk assessment and, if necessary, external audit bodies.

2.5.3 Data quality and data governance**Definition (EU AI Act, Art. 10):**

Requirements for the quality, representativeness and documentation of the data used.

Linguistic classification:

The term "data quality" is not defined in normative terms the AI Act, but rather is operationalized through specific requirements (e.g. "free of errors", "representative").

Differentiation of terms:

"Data quality" within the meaning of the AI Act is not identical to classical

QM criteria such as measurement accuracy, accuracy, precision, measurement uncertainty or calibration.

2.5.4 Auditability, traceability and transparency

Definition:

In the EU AI Act, the terms "auditability", "traceability" and "transparency" are differentiated as follows:

- **Auditability:** Capability of an AI system to be reviewed by external third parties (e.g. government agencies, certification bodies).
- **Traceability:** The ability to track data, models and decisions throughout the entire life cycle.
- **Transparency:** Disclosure of the inner functioning, logic and limits of an AI system.

Table 2-1: Definition of the concepts of auditability, traceability and transparency

Term	Focus	Typical contents
Auditability⁷	Auditability by third parties	Documentation, testing paths, external assessment
Traceability	Data flow & model history	Data sources, model versions, labeling processes
Transparency	Comprehensibility & disclosure	Purpose, functioning, limits of the system

2.5.5 Non-high-risk AI systems

Definition (implicitly defined in the AI Act):

AI systems not covered by Annex III to the EU AI Act and therefore not subject to the requirements for high-risk systems.

⁷ Chapter 3.4.8 describes "auditability" in the sense of quality management – i.e. the ability to traceably document AI-powered decisions within the framework of QM audits (e.g. according to VDA 6.3).

In chapter 5.4, the term is discussed within the regulatory context of the EU AI Act, where it forms part of the requirements for high-risk systems. The two perspectives complement each other and should be clearly distinguished in language use.

Linguistic classification:

There are no official terms such as "low-risk" or "unregulated AI" – in practice, terms such as "non-high-risk AI" or "limited-risk AI" have come to be used.

Linguistic recommendation:

- Avoid terms such as "harmless" or "unregulated."
- Use phrases like: "This system does not fall under the high-risk classification according to the AI Act."

2.6 Glossary for AI in quality management

Definitions and terms from the VDA publications are presented in a comprehensive online glossary of the VDA QMC:

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

3 Successfully using AI in QM

The use of AI in organizations, especially in quality management, offers myriad opportunities for improving the quality of products and services. Due to ever shorter product life cycles and changing market conditions, the use of AI to support quality management methods is essential to keep pace with these developments.

The use of AI offers the following benefits throughout the product development process, among others:

- **More efficient data analysis:** AI can process and analyze data from multimodal sources (e.g. texts, images, time series data and sensors) that were previously not evaluable by classical methods.
- **New access options and interfaces to internal and external data:** Intuitive interaction using semantic searching and natural language enables much faster and wider access to knowledge, even for users without technical expertise.
- **Improvement and automation of processes:** By using AI, opportunities for improvement in processes (e.g., development, business, and production processes) can be identified, and appropriate measures can be proposed. This can lead to higher productivity and lower costs.
- **Automated inspections:** AI-powered analyses can be used to automatically inspect components to increase the speed, accuracy and robustness of testing.
- **Error detection and prediction:** AI-powered systems can analyze patterns or anomalies in data and predict potential quality deviations (predictive quality) before they lead to more complex problems and higher costs. This enables predictive quality monitoring and control.
- **Predictive maintenance:** AI can help predict maintenance needs by analyzing data about the condition of machines. This minimizes downtime and extends the service life of the machines.

- **Customer feedback analysis:** AI tools can analyze customer feedback and data from the field, such as social media or surveys, to identify trends and problems that could indicate quality defects.

The targeted and successful use of AI in quality management requires consideration of various aspects relating to **people, technology** and the **organization**.

- **People:** Employees are the core of every organization. Their acceptance of and ability to work with new technologies are crucial for the success of AI implementation. Training and education are needed to ensure that employees can understand and effectively use the new systems. In the end, motivation and a positive attitude toward technology play a key role.
- **Technology:** The technical infrastructure must meet the requirements to support AI applications and deploy them in the organization. This includes suitable hardware, software solutions and platforms as well as databases for storing and processing large amounts of data. Defined interfaces and processes enable AI to be embedded in the organization. Tools and methods enable AI to be used for specific purposes.
- **Organization:** Targeted use cases form the basis for defining the necessary specifications of the required resources. Clear organizational structures are needed to define responsibilities and promote communication between different departments and divisions. Governance and standards enable the beneficial use of AI in compliance with data protection law and applicable regulations (e.g. EU AI Act, GDPR).

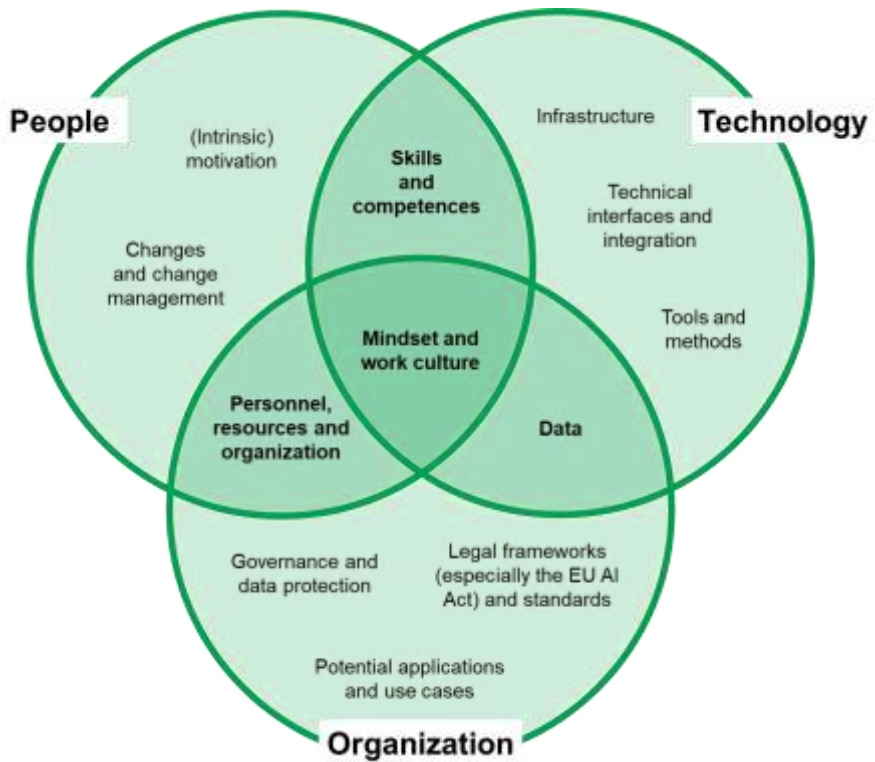


Figure 3-1: Classification of aspects that influence the successful use of AI in the organization.

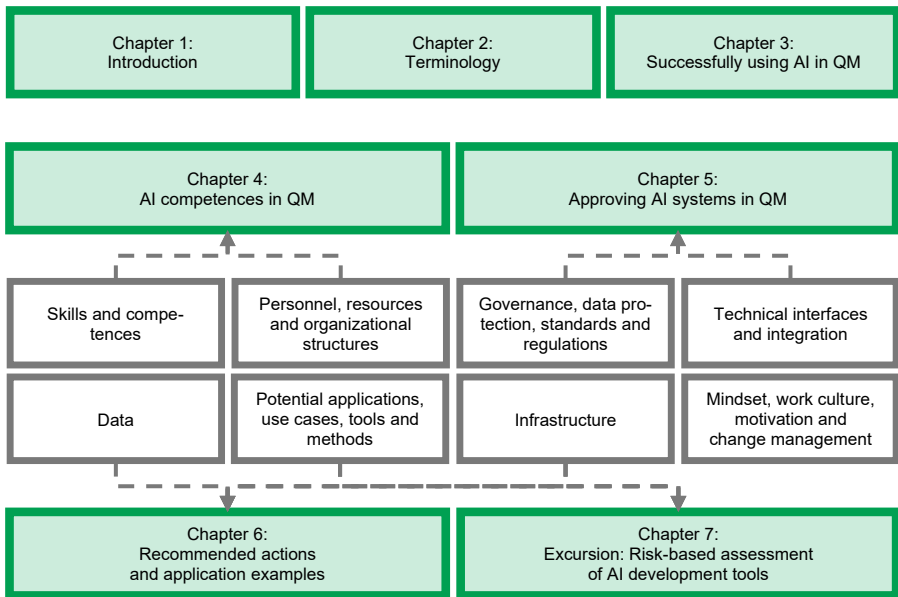


Figure 3-1: Success factors for the use of AI and overview of chapter structure.

The following section provides an overview of these aspects, which are divided into eight groups, each highlighting the success factors and challenges for the successful use of AI within the organization. These groupings will also serve as the basis for a more detailed examination in later chapters.

3.1 Mindset, work culture, motivation and change management

The successful introduction of AI in quality management depends not only on technical solutions, but also to a large extent on the mindsets, the work culture and the willingness to change in an organization. These terms refer to the common attitudes, values and behaviors that determine how people respond to new technologies, uncertainty and learning. These include openness, a willingness to learn, trust in data, interdisciplinary cooperation and the courage to question established processes. An innovation-friendly work culture promotes the acceptance of AI.

These factors are crucial because AI systems do not merely automate existing processes; rather, they can fundamentally transform working methods, roles and decision-making processes. Only if employees understand the benefits, assume responsibility and build trust in AI-powered decisions can

the full potential be realized. Collaboration between man and machine requires new forms of communication, leadership and motivation. This makes the development of appropriate mindsets a key success factor on the path to a sustainable culture of quality.

Success factors

- **Openness and willingness to learn:** Openness to new ideas and perspectives as well as motivation and willingness to acquire and deepen new knowledge
- **Iterative improvement:** Acceptance that AI solutions do not immediately work perfectly and that iterative improvement is necessary.
- **Transparency:** Clear communication about how and why AI is being used and what it can and cannot do (yet)
- **Participation and engagement:** Active involvement of employees in the development and introduction of AI solutions
- **Valuing of experience:** Seeing the combination of AI with human expertise and intuition as a strength
- **Leaders as enablers:** Demonstrating change, providing guidance and being open to dialogue
- **Access to AI coaches:** Enabling local multipliers to drive adoption, support users and disseminate best practices across the company.

Challenges

- **Fear of losing jobs, autonomy or control:** "I'll become superfluous," "AI will replace me," "AI makes the decisions"
- **Distrust in technology:** A lack of understanding of how AI works can lead to rejection or uncertainty.
- **Unrealistic expectations and blind trust:** "AI will solve everything," knowledge gap or lack of understanding among decision-makers at all levels
- **Unclear benefits to the individual:** Lack of understanding how AI will directly benefit the individual's everyday work
- **New tools create barriers to entry:** Increased work for data maintenance, training and getting used to the solutions
- **Unclear responsibilities:** Unclear roles and responsibilities lead to uncertainty in the use of technology ("What can I do with AI, what is allowed, what is not allowed?").
- **Forcing technology:** Introduction without employee involvement
- **Resistance to change:** Adherence to familiar processes and routines

3.2 Skills and competences

The use of AI significantly expands the spectrum of competences in quality management. In addition to classical quality expertise, data analysis and algorithm literacy are now becoming key skills. AI systems require not only technical expertise, but also the ability to interpret models, analyze their outputs and work across disciplines. This creates a new competence profile that integrates quality management more closely with data science, IT and intelligent processes.

Success factors

- **Understanding of AI basics:** Knowledge of subject areas such as statistics, data analyses, ML and algorithms
- **AI tools and platforms:** Basic knowledge of using and building AI tools
- **Interdisciplinary expertise:** Combination of different disciplines, e.g. IT, law, technology and engineering
- **Space for continuous learning:** Expanding skills and building competences
- **Training concepts:** Information conveyed in various media and formats such as online courses, training sessions and workshops

Challenges

- **Lack of data competence:** Limited skills for the creation, interpretation, evaluation and appropriate use of data
- **Technological change:** Rapid changes in digital and AI-powered technologies require ongoing development of skills, methods, tools and roles.
- **Technological complexity:** Breadth and technical complexity of the topic of AI

3.3 Data

Data is the cornerstone for the successful use of AI in quality management. This term encompasses not only the results of measurements and tests, but also information about processes, products and context that makes it possible to assess quality. Their availability, quality and structure determine how reliably AI models can work. At the same time, the introduction of AI is changing the way data is handled, from the collection of data to processing and interpretation. Concepts and strategies for collecting, storing and managing data are essential for high data quality. This makes systematic data management a key factor for deriving sound insights and sustainable improvements from information.

Success factors

- **Data quality:** Completeness, correctness, consistency, timeliness and relevance of the data
References: ISO 8000, ISO/IEC 25012
- **Data management:** Collecting, storing and maintaining data to improve data quality
- **Standardization:** Consistency of formats for collecting, exchanging and managing data
- **Integration:** Integration and aggregation from multiple sources
- **Automation:** Reducing factors that can impair quality
- **System maintenance:** Regular review of data quality and stability (changes in variability over time)

Challenges

- **Variability of data quality requirements:** Different and variable data quality criteria for AI depending on the application
- **Data availability and scalability:** Insufficient data for AI applications
- **Planning of the required data:** Inaccurate estimation of required data volume and quality
- **Data access:** Access restrictions or data protection policies for required data
- **Timeliness of data:** Rapid change in quality requirements due to dynamic market conditions and ever shorter product life cycles

3.4 Organizational structures and processes

In addition to human resources and technical competences, the successful introduction of AI in quality management depends largely on sufficient capacities, clear responsibilities and structures that enable agile working and continuous learning. AI is transforming the way teams are organized, resources are allocated and decision-making processes are structured. This gives rise to new requirements regarding leadership, role structures and interdisciplinary cooperation, all of which are crucial for the sustainable integration of AI systems.

Success factors

- **Resources:** Sufficient resources for the development and use of AI systems
- **Motivation and mindset:** Motivation and willingness to engage with a new AI-driven environment and to improve the system
- **Competences:** Training to enable users to utilize a new AI-driven environment
- **Organization:** Alignment and adaptation of organizational structures and roles to new AI systems
- **Infrastructure:** Familiarity with the interfaces to the AI system for data processing
- **Targeted communication:** Clear information and regular dialogue, including through new communication structures, channels and formats
- **AI and data governance:** Clear and quick decisions, accessible contact persons

Challenges

- **Silo thinking:** Lack of know-how or cooperation needed to build an efficient AI-driven environment
- **Insufficient integration into the organizational structure:** Low degree of digitalization and integration of AI tools into quality management processes
- **Insufficient integration into organizational culture:** Distrust, unrealistic expectations or resistance to AI technologies

3.5 Governance, data protection, standards and regulations

Governance and data protection refer to the framework conditions that ensure that AI is developed and used responsibly. Mechanisms are needed to define compliance requirements that meet legal standards. In particular, the handling of sensitive data requires comprehensive data protection policies. This means companies need clear policies on how data is collected, stored and used. Compliance with national and international standards and regulations is essential for the safe development and use of AI in order to avoid unforeseen risks to quality or safety. These aspects are crucial for trust in AI systems; without them, companies face legal risks and damage to their reputation.

Success factors

- **Clear responsibilities:** Defined roles and responsibilities for the development, operation and monitoring of AI applications
- **Regulatory compliance:** Development of AI systems with regulatory requirements in mind right from the outset
- **Data protection-compliant data processing:** Compliance with data protection policies, e.g. through anonymization, pseudonymization or consent management
- **Standardized processes for AI validation:** Monitoring of AI systems, e.g. through regular model review, audit trails and documentation
- **Use of established standards and frameworks:** Alignment with regulations and standards such as ISO 9001, ISO/IEC 27001,

Challenges

- **Unclear regulatory situation:** Uncertainty about applicable regulations, e.g. EU AI Act, GDPR and Product Liability Directive
- **Lack of internal policies for AI:** Lack of processes for selecting, deploying and controlling AI systems
- **Data protection conflicts:** Unauthorized use of data for training purposes, for example personal data
- **Lack of auditability:** Insufficient traceability and lack of documentation of the AI models
- **Loss of trust due to insufficient transparency:** Insufficient traceability or high uncertainty regarding the outputs and decisions of AI models
- **Lack of uniform standards for AI:** Lack of established benchmarks or validation methods for AI models

Success factors

Challenges

ISO/IEC 23894

(AI risks) and the EU AI Act

- **Transparency and traceability:** Explainability and documentation of model outputs and AI decisions
- **AI ethics guidelines:** Consideration of and compliance with aspects such as fairness, non-discrimination and human-centered development
- **Interdisciplinary governance teams:** Integration of various departments such as QM, IT and Legal

- **Unresolved liability questions:** Responsibilities for inaccuracies or wrong decisions by AI models not clearly defined
- **Slow adaptation of existing QM systems:** Lack of integration into existing QM methods; for example, FMEA (failure mode and effects analysis) or CAPA (corrective and preventive action) not designed for AI risks

3.6 Infrastructure

Infrastructure refers to the basic systems and technologies needed to successfully implement AI in organizations. A modern IT infrastructure forms the backbone for the development and operation of AI applications. This encompasses everything from hardware such as powerful servers and specialized processors, to software solutions and concepts that enable data to be exchanged and processed and algorithms to be run, to cloud platforms and data rooms, all the way to technical solutions and strategies for data sovereignty and cyber security. It also includes various models for how and to what extent the IT infrastructure is managed in-house (on-premises) or by a service provider, for example infrastructure as a service (IaaS), platform as a service (PaaS), and software as a service (SaaS). In addition, the network infrastructure plays a critical role in ensuring a stable connection between different systems and the rapid transfer of large amounts of data. A key success factor for modern infrastructure is the seamless integration of different systems within the organization and with external partners (e.g. for the exchange of quality data).

Success factors

Challenges

- **Data infrastructure:** Collection (sensor technology), storage and management of data in AI-enabled formats
- **Interfaces and integration:** Interoperability of systems and integration into existing system landscape
- **Scalability:** Expandable systems in the organization with adequate access and performance
- **Standardization:** Uniform formats and systems in the organization
- **Established IT infrastructure:** Low flexibility and legacy systems with multiple tools
- **Scalability and performance:** High computing and memory requirements
- **Integration of AI:** Limited integration and interoperability with existing workflows and systems
- **Resources:** High infrastructure investment costs
- **Technological complexity:** Complexity of IT systems and interfaces in the organization
- **Infrastructure localization:** Certain data is subject to country-specific or region-specific regulations that make seamless exchange difficult
- **Cyber security:** Increasing security threats jeopardize infrastructure stability as data protection regulations tighten
- **Data sovereignty:** Complex legal and regulatory conditions make it difficult to exchange data

3.7 Technical interfaces and integration

Interfaces and technical integration refer to the connections between different software and hardware systems that make it possible to exchange data and coordinate processes. Seamless integration with existing systems is

critical to the success of AI applications in organizations. Well-defined interfaces are needed to ensure that different technologies can work together harmoniously and information can flow efficiently. By connecting AI with existing production or quality management systems, companies can use data from various sources, such as machine sensors or customer feedback systems. In addition, a well-integrated infrastructure improves collaboration between departments. For example, shared access to data enables production, quality, and development teams to communicate and collaborate better, resulting in faster solutions to problems.

Success factors

Challenges

- **Defining interfaces:** Clearly defining the interfaces between the different AI systems
- **Networking:** Networking of related AI applications and agents, with the results available as data for further processing
- **Comprehensiveness:** Linking of AI systems in quality management functions with other functions and domains in the organization
- **Integration:** Integration of AI systems into quality management systems, environments and processes
- **Lack of uniformity:** Standardized and comprehensive interfaces not defined or not clearly defined
- **Design concept silos:** Lack of design concepts for AI agents that can be integrated with each other later
- **Data silos:** Lack of data concepts for AI systems that prepare data for other systems

3.8 Potential applications, use cases, tools and methods

The use of AI in quality management opens up new opportunities, from more efficient analyses to predictive quality assurance. But success depends not only on the tools or methods available, but also on their targeted application and the overcoming of key challenges such as data availability, process integration and user acceptance. The successful use of AI

requires a combination of clearly defined goals and requirements for processes, methods and tools. Concrete use cases create transparency about the requirements for the targeted use of AI.

Success factors

- **Clear definition of the problem:** Clear and measurable definition of the intended added value from using AI, for example in the areas of error detection and process optimization
- **Appropriate tool selection:** Selection of AI tools that fit the data situation, the use case and the user group
- **Tool usability:** Promoting acceptance through intuitive operation and good visualization
- **Integration into existing systems:** Identification of potential improvements and additions to existing QM methods such as SPC, FMEA or 8D report
- **Piloting and iterative approach:** Establishing small, controlled tests with clear metrics to assess value
- **Data quality and availability:** Structured, clean and sufficiently large data volumes as a basis for the effective and efficient use of AI
- **Successful use cases:** Promoting visibility and sharing best practices and success stories

Challenges

- **Unclear objectives:** Introducing AI without a concrete goal and anticipated benefits
- **Complexity of AI models:** AI models too complicated for the application or for the current level of expertise
- **Tool overload:** Too many tools without clear demarcation or integration
- **Missing, incorrect or incomplete dataset:** Missing or unusable historical data, e.g. data unstructured or entered manually. Fragmented data, lack of availability or poor data quality
- **Scaling outlay:** Functional pilots can be converted to stable productive operation only to a limited extent or with high outlays.
- **Lack of standardization:** Limited comparability and reproducibility due to different tools and methods
- **Failure to address regulatory requirements:** Limited ability to validate, audit and track AI models

4 AI competences in QM

The use of AI in quality management changes not only tasks and processes, but also the skills and competences required of employees.

The purpose of this chapter is to provide employees in quality-related roles with an overview of the required **competences** so that they can work more effectively and efficiently in their current roles by leveraging AI. Similarly, the competency requirements serve to foster professional development and prepare individuals for future roles and responsibilities.

AI literacy is defined in the EU AI Act as the skills, knowledge and understanding that enable providers, operators and stakeholders, while taking into account their respective rights and obligations, to **use** AI systems in a **competent manner** and to be aware of the **opportunities and risks** of AI, as well as the potential harm it may cause. AI competence should equip providers, operators and affected persons with the necessary concepts to make informed decisions about AI systems. These concepts may vary depending on the context. They include three areas for understanding the correct use of AI systems:

- in the **development phase** of the AI system,
- the measures to be applied during **use**, and
- the appropriate use of the **AI system's outputs**.

The concepts also include the knowledge needed to understand how decisions made with the help of AI can affect the people impacted by them. The EU AI Act defines the framework conditions for AI competences. Companies must ensure that their employees have the appropriate competences, depending on their role in working with AI. The competence requirements take into account the specifications of the EU AI Act and ISO 42001.

4.1 Main competences for AI in quality management

The **four main competences** for working with AI systems in quality management are:

1. Basic understanding of AI/ML concepts (AI models, data models, model evaluation, validation)

Employees are familiar with typical AI applications in quality management – including predictive quality, anomaly detection, computer vision, and process-oriented optimization approaches – and are able to assess these in terms of their benefits and the prerequisites for their use (e.g., high data quality).

They can explain the difference between classical rule-based systems and machine learning and identify their limitations.

In addition, employees understand basic ML concepts such as supervised and unsupervised learning, can handle training and test data and select suitable model types such as classification, regression or anomaly detection.

2. Basic understanding of data analysis and statistics (big data, visualization, data quality).

Employees understand the need for good data in sufficient amounts (completeness, accuracy and consistency) as a prerequisite for successful AI outputs and can assess the consequences if distorted data is used for analyses.

They understand that inadequate data quality can lead to incorrect decisions. Particularly in quality management, the problem is that errors are rare and often not reproducible, so data from faulty parts does not occur in very large numbers.

Employees are aware that distorted data can lead to distorted AI models and thus to poor predictions ("garbage-in/garbage-out").

3. Working with generative AI LLMs (including prompting)

Employees are able to correctly use appropriate tools for the practical application of generative AI (LLMs) to make work easier, such as structured creation and revision of documentation (8D reports or FMEA de-

scriptions), efficient research of best practices and normative requirements as well as optimization of written communication in the form of e-mails or reports.

They can clearly and precisely formulate instructions in the form of basic prompt engineering in order to obtain reliable and relevant outputs.

Employees are able to critically validate generated content and check for possible hallucinations to avoid misinformation and ensure factually correct usage.

4. Awareness of AI opportunities and risks, AI ethics, AI legal principles and potential harm

Employees are able to interpret the outputs of AI systems by understanding probabilities, reliability and model-based predictions and situate them within the context of the respective application. In doing so, they can critically scrutinize the AI outputs, compare them with expert knowledge and check their plausibility. A basic understanding of explainability is essential to understand why a model has arrived at a certain decision – especially with regard to auditability and compliance requirements.

Employees know the ethical, legal and organizational conditions and implement them responsibly. These include:

1. conscious handling of personal data in accordance with the GDPR,
2. sensitivity to possible bias in AI models – for example, through distorted data that can lead to unfair assessments – and
3. understanding that AI supports decision-making but does not replace human responsibility.

In addition, employees are able to ensure transparency by documenting the use of AI systems in a traceable manner in order to meet regulatory and audit-related requirements.

Note: Due to an increasingly simple tool landscape, AI can also be used to solve more complex problems. Examples:

- Program code can be generated via an LLM and be used to form complete applications.
- In-depth programming knowledge is often no longer necessary for the automation of process flows, as more and more no-code or low-code tools are being used to directly create applications.
- Agentic workflows that act independently and make decisions.

Particular caution is required here and employees must be aware of the associated risks.

4.2 Roles in quality management and relevant AI competences

The four main competences are fundamental for many tasks, and therefore also for employees in quality management. In addition, and to make it easier for employees to identify and apply them, it helps to link the necessary AI competences with the **typical roles in quality management**:

4.2.1 Classical roles in quality management

This volume describes **seven "classical" roles** in automotive quality management that cover the majority of quality activities and in which employees in quality-related roles may find themselves. Example AI-specific competence requirements are defined for these roles, which can help employees to carry out their existing tasks more effectively and efficiently with the help of AI.

Seven classical role profiles:

- **Employee in supplier quality** (SQE, SDE)
- **Employee in development quality** (product development)
- **Employee in production quality** (process development, industrialization, series production)
- **Employee in customer quality** (complaints, warranty, field monitoring)

- **Employee in quality management systems** (QMS owner)
- **Quality auditor** (system, process, product) and **quality assessor** (ASPICE)
- **Quality manager** (management function for personnel, budget, strategy)

The typical tasks associated with these roles will not be discussed in detail here, as they are widely known and described in other VDA regulations (e.g. as part of the maturity level or audit guidelines).

The roles are generic descriptions, not specific position/role descriptions. The exact scope of tasks and responsibilities of individual employees may sometimes span several roles.

4.2.2 New roles in quality management

Due to the increasing use of AI in quality activities, this volume also identifies **four "new" roles** in quality management in the automotive industry that (may) result from the use of AI.

Example AI-specific competence requirements are defined that can help employees to perform their new roles. In practice, employees in quality management will increasingly deal with AI in their role (fulfilling their "classical" role with appropriate AI support) and then sometimes evolving to take on "new" roles. The more extensively AI is used in companies – and thus also in quality assurance – the more roles and subtopics become conceivable. These are not considered in this volume.

The roles are generic descriptions, not specific position/role descriptions. The exact scope of tasks and responsibilities of individual employees may sometimes span several roles.

- **AI Q-Data Engineer** (focus: data acquisition and preparation)

Quality data engineers build and operate the data infrastructure for QM applications. They develop automated data pipelines, integrate heterogeneous sources (sensors, MES, ERP), ensure data quality and implement monitoring systems. They work with cloud technologies, establish DevOps practices and create scalable architectures.

Objective: Providing reliable, high-performance data for analysts and data scientists.

- **AI Q-Data Analyst** (focus: data analysis and visualization)

Quality data analysts analyze quality-relevant data from production and testing processes to identify patterns, deviations and potential for improvement. They apply statistical methods (SPC, Cpk), use AI tools for pattern and anomaly detection, create dashboards and communicate findings to stakeholders. Objective: Data-driven quality improvement and error prevention.

- **AI Q-Data Scientist** (focus: development of data analysis methods)

Quality data scientists develop and train machine learning models for QM-specific applications such as error prediction, anomaly detection or optical quality inspection. They carry out feature engineering, evaluate model performance, ensure interpretability and work with data engineers on deployment. Objective: Develop innovative AI solutions for quality improvement and predictive quality.

- **AI Q-Data Manager** (focus: AI strategy, data governance and compliance)

Quality data managers develop and manage the data strategy in the QM sector, establish data governance and define standards. They build interdisciplinary teams, select technologies, manage budgets and orchestrate collaboration between IT, QM and specialist departments. Objective: Driving data-driven transformation, creating business value and ensuring compliance.

4.2.3 Role-specific AI competences

The **AI competency requirements** include the four main requirements mentioned above (which are identical for all "classical" roles), which are then broken down into specific competency requirements for each role. This helps to connect the typical tasks and activities of each role with specific competences and examples of AI usage. This makes it easier for employees in quality-related roles to **identify their own tasks** and the resulting AI competences. If more complex applications such as large AI-controlled

process automation systems are set up, AI engineers and software engineers are also required. However, these do not require in-depth knowledge of quality management and are therefore not described here.

Appendix 1 describes the above-mentioned roles and their recommended AI competences with examples of AI use.

Of course, there are different characteristics for the individual competence requirements, depending on the exact task requirement, the company-specific characteristics of job and work content and the previous and future necessary competence of individual employees. Basically, it can be said that in the "classical" quality roles, an **AI competence level** from "**basic**" (= employee understands and can use AI) to "**advanced**" (= employee has mastered the use of AI) is usually sufficient. For the "new" roles, the relevant AI competence level then goes up to "**expert**" (= employee designs and trains AI). However, as already mentioned, this level is heavily dependent on the individual job and function profile of the employee.

5 Approving AI systems in QM

This section describes the procedure for determining the potential risks of an AI system that has been developed for a specific purpose and how these can be used as the basis for approval. These risks should be identified and evaluated before the AI system is put into operation, at the latest, but it is recommended that the criteria from this method be applied during the planning and development of the system.

The risks are determined in two steps:

- The first step involves a project risk assessment for the scope of the AI-related task with a view to the impact a fault would have on the company. The result of this activity, i.e. the project risk class, defines the scope of the assessment-relevant requirements for the respective AI system.
- The next step is to evaluate the requirements that are recommended based on the risk class of the project. This involves assessing whether the AI system fulfills the requirements relevant to the assessment.

Step 1: Determining the project risk class

"How critical is a malfunction of the system?"

Defining the risk class AIQM

Step 2: Risk assessment of the AI system

"Are the technical requirements met by the AI system?"

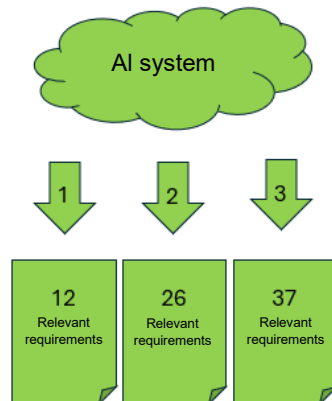


Figure 5-1: Diagram showing the risk assessment of an AI system

The aim is to evaluate all relevant requirements for fulfillment or non-fulfillment and to justify this evaluation in a documented manner. If requirements in the checklist are not applicable (n.a.), the reason for non-application must be documented.

A negative assessment results from the non-fulfillment of the corresponding requirement. This does not automatically constitute a rejection of the AI system. Rather, an assessment of the remaining risks and the implementation of appropriate measures (e.g. risk acceptance, implementation of risk mitigation measures) is expected.

The actual approval procedure based on these assessed requirements and any necessary additional measures must be determined by the company using the system. Depending on the project risk class, it is recommended that the relevant stakeholders be identified and included in the approval process for the AI system.

Even if, due to a low project risk class, only some of the assessment-relevant requirements are applicable and therefore need to be assessed, it is recommended that all other requirements for higher project risk classes are also considered.

Note: The assessment criteria were deliberately narrowed down to the technical aspects of developing AI systems that are designed and trained for a specific business purpose. This method is not a general evaluation of the suitability of AI tools, i.e. tools for the automated development of AI systems.

5.1 Step 1: Determining the project risk class

The AIQM project risk class is determined based on various risk criteria with assessment scales that are divided into seven risk groups. The resulting project risk class reflects both the complexity and the business risk associated with using the AI system.

In each of the risk groups, an assessment must be made as to whether the risk is high (class AIQM-3), moderate (class AIQM-2) or low (class AIQM-1). The overall project risk assessment results from the highest risk level from all seven risk groups.

Example: If one risk group is rated AIQM-3, two others AIQM-2 and all others AIQM-1, the project risk class of the AI system is rated AIQM-3.

No.	Risk group	Risk criterion	Assessment	Risk class
1.	AI regulations	<p>EU: High-risk AI system in accordance with the EU AI Act, but with established control mechanisms or specific homologation requirements.</p> <p>Other markets: High regulatory requirements of the respective target markets</p>	<p>EU: AI system is a high-risk AI system according to the EU AI Act.</p>	AIQM-3
		<p>Moderate regulatory requirements and/or GPAI</p>	<p>EU: AI system is subject to transparency obligations in accordance with EU AI Act Article 50.</p>	AIQM-2
		<p>Low regulatory requirements</p>	<p>EU: AI system is not a high-risk AI system under the EU AI Act and is not subject to any special requirements (see Article 95).</p>	AIQM-1

2.	Data protection law	Data is personal and sensitive or data rights of third parties could be violated.	There is a high risk associated with the use of personal data, data requiring a high level of protection – as information may be leaked from the company or unintended consequences may result – or third-party data, where the data rights of those parties could be violated.	AIQM-3
		Pseudonymized data	Low risk due to technical and organizational measures	AIQM-2
		No personal or sensitive data processed	Only anonymized or synthetic data is used	AIQM-1
3.	Bias & fairness	Fairness is violated or there is an undesirable bias.	Bias will most likely occur	AIQM-3
		Fairness may be violated or there may be an undesirable bias.	Bias may occur, but there are processes for detection and correction	AIQM-2
		No relevant evidence of bias or fairness violation	Training data and model have an acceptable bias	AIQM-1

4.	Transparency	Largely black box, largely not explainable	Important decisions are largely untraceable	AIQM-3
		Partially explainable, some black box aspects	Some decisions are not traceable	AIQM-2
		Model is largely explainable and documented	All decisions are traceable	AIQM-1
5.	Financial risk	High economic losses due to incorrect outputs	Incorrect outputs or the failure of business processes (e.g. sales, maintenance) lead to considerable costs or misallocations.	AIQM-3
		Moderate economic impact in the event of a fault	Errors lead to increased costs, but no critical losses.	AIQM-2
		Low economic impact	Incorrect decisions lead to additional costs, but these are cushioned by processes or are economically manageable.	AIQM-1
6.	Reputational risk	High-profile incident with loss of confidence.	AI system causes scandal or discrimination, resulting in massive reputational damage.	AIQM-3

		Moderate reputational risks	In the event of an incident, affected stakeholders are brought on board; the reputational damage is limited, with the opportunity to mitigate it through open communication and problem-solving.	AIQM-2
		No relevant or low reputational risk.	AI application is only used for internal purposes. No impact outside the company is to be expected.	AIQM-1
7.	Product characteristics	Assessment of the functional safety (ISO 26262) or safety in use (ISO 21448) of the product in question as well as special characteristics (IATF 16949) and cyber security (CS)	High = functional safety ASIL D/C, ISO 21448 type 3/4 or special characteristic or components with active communication (CS requirements)	AIQM-3
			Medium = functional safety ASIL B/A, ISO 21448 type 2, not a special characteristic and low CS relevance	AIQM-2

			Low = functional safety ASIL QM, ISO 21448 type 1, not a special characteristic and no CS relevance	AIQM-1
--	--	--	---	--------

5.2 Step 2: Risk assessment of the AI system

The relevant requirements are assigned to process phases from the development to the commissioning of an AI system. This serves to provide a better overview and readability. The phases or assessment-relevant requirements can also be assigned to project and/or development phases of the AI development processes defined in the company using the AI system.

5.2.1 Overview of the key questions on the assessment-relevant requirements in the process phases

1	Application area
1.1	Have you defined an application area and a realistic goal for the AI application?
1.2	Have special requirements for the explainability of the expected AI behavior been defined?
1.3	Have all relevant roles been defined and the necessary competences ensured to enable a smooth start to the project/activity?
1.4	Has the use of the AI system been coordinated with the customer?
1.5	Are relevant regulatory requirements, internal and external standards and contractual agreements known and are they being adhered to?
2	Understanding of data
2.1	Has it been determined which data is required to develop an AI system for the desired business purpose?
3	Data collection
3.1	Is the collection of data sufficiently documented so that it can be reproduced?
3.2	Does the data meet the requirements of the task?
3.3	Is data versioning ensured?
3.4	Are relevant stakeholders involved in the data collection?

4	Data preparation
4.1	Is the data provided complete, accurate and consistent to enable trustworthy modeling, testing and validation?
4.2	Has it been ensured that the data is fair and representative and that relevant bias is reduced to an acceptable level?
4.3	Are the data preparation steps reproducible, traceable and correct from a technical point of view?
4.4	Have data processing and security-critical aspects been taken into account?

5	AI modeling
5.1	Is the training deterministically configured and reproducible?
5.2	Is there a model calibration in place that reflects the uncertainties?
5.3	Is the training fully documented, including procedures, performance limits and restrictions?

6	Evaluation/testing
6.1	Does the model concept ensure explainability and interpretability for relevant stakeholders?
6.2	Are there metrics in place for reproducibility, system fairness, monitoring, legal compliance and GDPR compliance of the AI tests?
6.3	Are robustness, edge case and sensitivity tests systematic, measurable and documented?
6.4	Are tests, KPIs and reports in place to ensure functional equivalence, consistency and compliance?
6.5	Are incorrect answers reduced to an acceptable level?

7	Preparation for deployment
7.1	Have IT security and cyber security issues been taken into account?

7.2	Has an operational concept including escalation and contingency concepts been drawn up and agreed?
7.3	Is a secure rollout process planned?

8	Application integration
8.1	Has an access and identity management plan been defined and implemented for the AI system, the AI application and relevant datasets?
8.2	Has the documentation for data, algorithm, hardware and software been created and is versioning ensured?

9	Performance verification
9.1	Does the AI system fulfill the planned business purpose?
9.2	How robustly does the AI system react under extreme loads? Are system limits defined and monitored?
9.3	Is the performance of the AI system verified under normal conditions and taking into account different test methods?
9.4	Are validation results for each AI system version documented in a complete, traceable and auditable manner?

10	Go-live
10.1	Has responsibility for the AI system been transferred to the customer?

11	Continuous improvement
11.1	Has data monitoring been set up?
11.2	Has model monitoring been set up?
11.3	Has performance monitoring been set up for the runtime environment in live operation?
11.4	Has a change management system been set up?
11.5	Is a monitoring and retraining/update procedure in place?

5.2.2 Assessment-relevant requirements for AI systems

Process phase 1: application area				
1.1	Have you defined an application area and a realistic goal for the AI application?	Applicable to risk class		
		AIQM-1	AIQM-2	AIQM-3
		Yes	Yes	Yes
Assessment-relevant requirements		Examples for implementation		
<p>The application area and the use of the AI application have been defined and aligned with the strategy (company, IT, etc.). Measurable goals and use cases for the AI system have been defined:</p> <ul style="list-style-type: none"> - mathematically or - on the basis of example data or - semantically/verbally (e.g. scenario catalog, deliberately limited operational scope). <p>Scalability must be defined where possible (i.e. applicability to different lines, plants, IT systems). Robustness requirements and IT security requirements must be described if necessary. Where required by internal specifications, a return on investment (ROI) analysis or a benchmarking analysis must be prepared, as well as the necessary cost allocation models (e.g. licenses, tokens).</p> <p>If there are requirements for the end-of-life planning of the AI system (e.g. long-term archiving of all data and models), these must be integrated into the application area.</p> <p>Generative AI: Generative AI requires clear specifications for content generation. The use case restrictions are crucial. It is also necessary to define which high-risk content (e.g. violence, hate) should be filtered out.</p>		<ul style="list-style-type: none"> • Project description • Contracting • Specification • Objective achievement sheet 		

Process phase 1: application area				
1.2	Have special requirements for the explainability of the expected AI behavior been defined?	Applicable to risk class		
		AIQM-1	AIQM-2	AIQM-3
		No	Yes	Yes
Assessment-relevant requirements		Examples for implementation		
<p>The desired function of an expected AI behavior (explainability) is described, documented and traceable. The relationship between inputs and outputs can be described in a level of detail appropriate to the task.</p> <p>The limits of the system and foreseeable misuse must be taken into account.</p> <p>The influence of the outputs of the AI system on other processes or decisions must be taken into account.</p> <p>Any necessary verification of the AI system outputs must be planned (e.g. human-in-the-loop).</p> <p>Note: According to Art. 50 of the EU AI Act, transparency is required with regard to explainability and traceability.</p> <p>Generative AI: Explainability is a challenge when using generative AI, but is essential for understanding the outputs and building trust.</p>		<ul style="list-style-type: none"> • Project description • Contracting • Specification • Objective achievement sheet 		

Process phase 1: application area				
1.3	Have all relevant roles been defined and the necessary competences ensured to enable a smooth start to the project/activity?	Applicable to risk class		
		AIQM-1	AIQM-2	AIQM-3
		Yes	Yes	Yes
Assessment-relevant requirements		Examples for implementation		
<p>A project manager/principal owner and relevant experts for software, AI, systems and domains are involved in the project/activity and are deployed with the necessary capacities. The required competences of those involved must be ensured (including requirements for AI literacy in accordance with the EU AI Act).</p> <p>Additional roles required may include the open source officer, the legal department, the cyber security department, the data protection officer or a member of product compliance.</p> <p>Note: Further role descriptions and competences can be found in chapter 4 of this volume.</p>		<ul style="list-style-type: none"> • Project description • Contracting • Organization chart • Training matrices 		

Process phase 1: application area				
1.4	Has the use of the AI system been coordinated with the customer?	Applicable to risk class		
		AIQM-1	AIQM-2	AIQM-3
		Yes	Yes	Yes
Assessment-relevant requirements		Examples for implementation		
<p>The use of the AI system is clarified and agreed with the customer/stakeholder at an early stage, if required.</p> <p>The monitoring during the period of use and, if necessary, the updating concept for the AI system are coordinated with the customer/stakeholder.</p> <p>EU AI Act Art. 50: There is a transparency obligation for all AI systems intended for direct interaction with persons. For high-risk AI systems, transparency is required in all cases.</p>		<ul style="list-style-type: none"> • Project description • Contracting • Contract • Specification 		

Process phase 1: application area				
1.5	Are relevant regulatory requirements, internal and external standards and contractual agreements known and are they being adhered to?	Applicable to risk class		
		AIQM-1	AIQM-2	AIQM-3
		Yes	Yes	Yes
Assessment-relevant requirements		Examples for implementation		
<p>Data and algorithms must be the property of the company, licensed by the company or freely accessible for commercial use.</p> <p>It must be clarified who is liable for the behavior of the AI within the framework of contracts (e.g. with suppliers).</p> <p>It must also be clarified which additional requirements, e.g. safety standards and product liability, need to be taken into account.</p> <p>At a minimum, check the requirements regarding:</p> <ul style="list-style-type: none"> • EU AI Act and/or other regulatory provisions in the target markets • Export control regulations • Data complies with GDPR/regional data protection requirements. The protection of personal data is ensured in all phases. The implementation of erasure concepts and rights is taken into account (right to erasure, retention policies). • Data law requirements, e.g. EU Data Act or Chinese Data Security Law • Software and data set licensing • Open source software management • Patent search: No patents are infringed 		<ul style="list-style-type: none"> • Project description • Contracting • Contract • Specification 		

-
- When working with suppliers and customers, liability must be clarified and included in the contract
 - Consideration of existing and incoming data with regard to its legal usability, in particular with regard to copyrights and GDPR.

Where applicable, ethical principles and fairness criteria must be taken into account.

Process phase 2: Understanding data				
2.1	Has it been determined which data is required to develop an AI system for the desired business purpose?	Applicable to risk class		
		AIQM-1	AIQM-2	AIQM-3
		Yes	Yes	Yes
Assessment-relevant requirements		Examples for implementation		
<p>The data required for training, testing, validation and operation of the AI system must be determined. Particular attention must be paid to the following properties:</p> <ul style="list-style-type: none"> • Type, scope and variations of the data (e.g. timeliness) • Data formats • Expected data transfer rates • Data schemas • Attribution/indexing • Metadata • Origin of the data: Data sources (for using existing data or collecting new data) are accessible from a technical point of view and documented. License conditions and rights of use are documented for third-party sources. • Raw data is stored unchanged and versioned in unalterable storage to ensure complete traceability • Define requirements for archiving • Data compression vs. loss of data 		<ul style="list-style-type: none"> • Specification • Overview of data formats and contents 		

Process phase 3: Data collection				
3.1	Is the collection of data sufficiently documented so that it can be reproduced?	Applicable to risk class		
		AIQM-1	AIQM-2	AIQM-3
		No	Yes	Yes
Assessment-relevant requirements		Examples for implementation		
<p>The methodology and system for collecting data is fully described and traceably documented.</p> <p>The procedures for collecting and annotating the collected data are specified.</p> <p>The reproducibility of data collection is guaranteed.</p>		<ul style="list-style-type: none"> • Specification • Work instruction • Overview of data formats, sources and contents 		

Process phase 3: Data collection				
3.2	Does the data meet the requirements of the task?	Applicable to risk class		
		AIQM-1	AIQM-2	AIQM-3
		Yes	Yes	Yes
Assessment-relevant requirements		Examples for implementation		
<p>The data collected is suitable for fulfilling the purpose of the AI system defined in the "Application area " process phase, in particular for training, testing/verification and performance validation.</p> <p>If additional data is required: Generating synthetic data is possible under certain circumstances:</p> <ul style="list-style-type: none"> • Simulate the data, e.g. defective parts, using available data and subject-area expertise, e.g. physical models. Subject-area experts document the reasons 		<ul style="list-style-type: none"> • Overview of data formats, sources and contents • Risk assessment • Result of the data examination 		

<p>for using synthetic data and the differences to real data (see also VDA5.3 chapter 6).</p> <ul style="list-style-type: none"> • Synthetic data must not be generated by the AI system that is to be trained or tested (risk of bias). • The synthetic data must be checked before use. 	
---	--

Process phase 3: Data collection				
3.3	Has data versioning been ensured?	Applicable to risk class		
		AIQM-1	AIQM-2	AIQM-3
		No	No	Yes
Assessment-relevant requirements		Examples for implementation		
Any changes to the datasets or raw data are transparently tracked and documented by an established versioning system.		<ul style="list-style-type: none"> • Work instruction • Database or data fields for version data 		

Process phase 3: Data collection				
3.4	Have relevant stakeholders been involved in data collection?	Applicable to risk class		
		AIQM-1	AIQM-2	AIQM-3
		No	Yes	Yes
Assessment-relevant requirements		Examples for implementation		
The relevant stakeholders (e.g. data protection officers, customers, internal departments) were informed about and involved in the data collection.		<ul style="list-style-type: none"> • Contracting • Contract • Organization chart • Proof of notification, e.g. e-mail 		

Process phase 4: Data preparation				
4.1	Is the data provided complete, accurate and consistent to enable trustworthy modeling, testing and validation?	Applicable to risk class		
		AIQM-1	AIQM-2	AIQM-3
		Yes	Yes	Yes
Assessment-relevant requirements		Examples for implementation		
<p>Once the planned data has been collected, the following criteria are checked:</p> <ul style="list-style-type: none"> • Completeness (sufficient data available, gaps have been identified) • Uniqueness (each data point / each dataset is identifiable and assignable, duplicates are recognized) • Consistency (value ranges, formats) • Quality limits for data defined and documented 		<ul style="list-style-type: none"> • Inspection report with result • Definition of the quality limits of the data • Automated checks during data collection or further processing 		

Process phase 4: Data preparation				
4.2	Has it been ensured that the data is fair and representative and that relevant bias is reduced to an acceptable level?	Applicable to risk class		
		AIQM-1	AIQM-2	AIQM-3
		No	No	Yes
Assessment-relevant requirements		Examples for implementation		
<p>A bias analysis has been carried out and documented.</p> <p>Datasets represent the target population (distributions documented).</p> <p>Measures for bias reduction defined and implemented.</p> <p>Fairness metrics are defined and fulfilled.</p>		<ul style="list-style-type: none"> • Analysis report with results • Documentation of measures with responsibilities, deadline and status • Effectiveness of measures taken • Definition of fairness metrics • Automated checks during data collection or further processing 		

Process phase 4: Data preparation				
4.3	Are the data preparation steps reproducible, traceable and correct from a technical point of view?	Applicable to risk class		
		AIQM-1	AIQM-2	AIQM-3
		No	Yes	Yes
Assessment-relevant requirements		Examples for implementation		
<p>All changes to the dataset are recorded in a traceable manner.</p> <p>Outlier analyses are in place and evaluated.</p> <p>Feature selection and generation are justified and documented.</p>		<ul style="list-style-type: none"> • Documentation of changes • Result of outlier analysis • Documentation regarding feature selection and generation 		

Process phase 4: Data preparation				
4.4	Have data processing and security-critical aspects been taken into account?	Applicable to risk class		
		AIQM-1	AIQM-2	AIQM-3
		No	No	Yes
Assessment-relevant requirements		Examples for implementation		
<p>Safety-critical data with reference to special characteristics (see IATF16949) are secured with special test measures.</p>		<ul style="list-style-type: none"> • List of special characteristics according to IATF16949 • Documentation of measures with responsibility, deadline and status • Effectiveness of measures taken 		

Process phase 5: AI modeling				
5.1	Is the training deterministically configured and reproducible?	Applicable to risk class		
		AIQM-1	AIQM-2	AIQM-3
		Yes	Yes	Yes
Assessment-relevant requirements		Examples for implementation		
<p>Deterministic processes are implemented, e.g. seed management.</p> <p>Training processes are reproducibly validated on different development environments.</p> <p>Reproducibility and traceability are ensured (e.g. storage of hyperparameters, tool-supported documentation).</p> <p>Generative AI: These requirements cannot be implemented with generative AI. In this case, additional measures must be provided, e.g. for monitoring and testing.</p>		<ul style="list-style-type: none"> • Result of tests • Documentation of hyperparameters 		

Process phase 5: AI modeling				
5.2	Is there a model calibration in place that reflects the uncertainties?	Applicable to risk class		
		AIQM-1	AIQM-2	AIQM-3
		No	Yes	Yes
Assessment-relevant requirements		Examples for implementation		
<p>Suitable calibration methods have been applied and documented.</p> <p>Evaluation of the prediction uncertainties has been carried out.</p> <p>Calibration is regularly checked and updated.</p> <p>Generative AI: These requirements cannot be implemented with generative AI. In this case, additional measures must be provided, e.g. for monitoring and testing.</p>		<ul style="list-style-type: none"> • Calibration methods such as flat scaling, isotonic regression, temperature or matrix scaling, label smoothing • Evaluation of the prediction uncertainty • Documentation of the calibration 		

Process phase 5: AI modeling				
5.3	Is the training fully documented, including procedures, performance limits and restrictions?	Applicable to risk class		
		AIQM-1	AIQM-2	AIQM-3
		No	No	Yes
Assessment-relevant requirements		Examples for implementation		
<p>Suitable training and test procedures are in place and their results are documented. Performance metrics and their limits are documented. Known limitations and risks are explicitly described.</p> <p>Generative AI: These requirements cannot be implemented with generative AI. In this case, additional measures must be provided, e.g. for monitoring and testing.</p>		<ul style="list-style-type: none"> • Procedural instructions • Performance metrics • Risk assessment • Description of the application area • Specification 		

Process phase 6: Evaluation/test				
6.1	Does the model concept ensure explainability and interpretability for relevant stakeholders?	Applicable to risk class		
		AIQM-1	AIQM-2	AIQM-3
		No	No	Yes
Assessment-relevant requirements		Examples for implementation		
<p>Requirements for explainability are checked and fulfilled. Interpretability methods (e.g. SHAP, LIME) are specified and documented. Critical features and their influence are traceably presented.</p>		<ul style="list-style-type: none"> • List of explainability requirements (e.g. safety, EU AI Act) • List of critical features 		

Process phase 6: Evaluation/test				
6.2	Are there metrics in place for reproducibility, system fairness, monitoring, legal compliance and GDPR compliance of the AI tests?	Applicable to risk class		
		AIQM-1	AIQM-2	AIQM-3
		No	Yes	Yes
Assessment-relevant requirements		Examples for implementation		
<p>Clear, quantifiable performance metrics for monitored procedures and unmonitored procedures are defined and validated on benchmark datasets.</p> <p>Detailed bias analysis of the model is performed on clearly defined groups whose data coverage and sample sizes are checked. Fairness metrics have been recorded and evaluated.</p> <p>The evaluation process is standardized with visualization of group disparities and peer review.</p>		<ul style="list-style-type: none"> • Metrics for monitored procedures such as F1, precision, recall, accuracy • Metrics for unsupervised procedures such as silhouette, Davies-Bouldin • Result of the bias analysis of the model • Result of the fairness metrics such as error rate differences, disparate impact / disparity index, demographic parity gap, equalized odds (TPR-FPR gaps) and calibration gap. • Result of the evaluation • Participants of the peer review 		

Process phase 6: Evaluation/test				
6.3	Are robustness, edge case and sensitivity tests systematic, measurable and documented?	Applicable to risk class		
		AIQM-1	AIQM-2	AIQM-3
		No	Yes	Yes
Assessment-relevant requirements		Examples for implementation		
<p>Standardized sensitivity analyses have been systematically introduced with measurable KPIs. Edge case tests under extreme inputs have been carried out and evaluated.</p>		<ul style="list-style-type: none"> • Definition of KPIs and acceptance criteria • Reports of tests and analyses 		

Process phase 6: Evaluation/test				
6.4	Are tests, KPIs and reports in place to ensure functional equivalence, consistency and compliance?	Applicable to risk class		
		AIQM-1	AIQM-2	AIQM-3
		No	No	Yes
Assessment-relevant requirements		Examples for implementation		
<p>Tests demonstrate the functional equivalence of the AI model using clearly defined input/output sets, if available or feasible.</p> <p>Consistency metrics / KPIs are continuously monitored: Deviation, error rate, response times are within defined tolerances.</p>		<ul style="list-style-type: none"> • Definition of KPIs and acceptance criteria • KPI tracking • Reports of tests and analyses 		

Process phase 6: Evaluation/test				
6.5	Are incorrect responses in systems with generative AI reduced to an acceptable level?	Applicable to risk class		
		AIQM-1	AIQM-2	AIQM-3
		No	No	Yes
Assessment-relevant requirements		Examples for implementation		
<p>Special tests for creativity, consistency and plausibility (benchmarking) are integrated for generative AI.</p> <p>There is an additional check of generated content for stereotypical or unintentional prejudices (bias).</p>		<ul style="list-style-type: none"> • Definition of KPIs and acceptance criteria • KPI tracking • Reports of tests and analyses 		

Process phase 7: Preparation for use				
7.1	Have IT security and cyber security issues been taken into account?	Applicable to risk class		
		AIQM-1	AIQM-2	AIQM-3
		Yes	Yes	Yes
Assessment-relevant requirements		Examples for implementation		
<p>Where necessary and relevant, measures are in place to ensure the integrity, robustness and general cybersecurity of the AI system or AI application against potential attacks throughout its life cycle.</p> <p>Where appropriate, cybersecurity officers are involved if there is an AI security risk in the use case.</p> <p>Possible risks may include:</p> <ul style="list-style-type: none"> • Prompt injection – Direct injections override system prompts, while indirect injections manipulate input from external sources. • Data poisoning – In data poisoning attacks, manipulated data is deliberately injected to contaminate the training data and thus compromise the model training process. • Circumvention – Circumvention attacks find small glitches in the AI/ML model input that lead to significant changes in the output. • Model extraction – In model extraction attacks, AI/ML algorithms are extracted or copied. Malicious actors could use model extraction attacks to steal the model. • Inference – Attempting to determine whether or not the information of a particular dataset, e.g. a person, was part of the training data of a trained ML model. • Sensitive information disclosure – Inclusion of sensitive information in the training data can lead to information disclosure. 		<ul style="list-style-type: none"> • Cybersecurity risk analysis 		

<ul style="list-style-type: none">• Model denial of service – Attack by overloading the system.• Supply chain vulnerabilities – using third-party datasets, pre-trained models and plugins increases security vulnerabilities.• Over-reliance – systems or individuals that rely excessively and without oversight on LLMs may be exposed to misinformation, misunderstandings, legal issues and security vulnerabilities due to incorrect or inappropriate content generated by LLMs.• Prompt injection	
---	--

Process phase 7: Preparation for use				
7.2	Has an operational concept including escalation and contingency concepts been drawn up and agreed?	Applicable to risk class		
		AIQM-1	AIQM-2	AIQM-3
		Yes	Yes	Yes
Assessment-relevant requirements		Examples for implementation		
<p>An operational concept for the planned use of the AI system must be drawn up.</p> <p>This should take into account the following aspects:</p> <ul style="list-style-type: none"> • Define and implement the escalation and contingency processes for incidents during operation, depending on the security risk. • A contingency plan should be taken into account, in addition to other failures in the business. • Define a support and maintenance plan • Defined responsible persons (RASIC), defined contract design with suppliers and service providers • Update strategy • Defined monitoring concept • Defined channel for feedback and information on system properties • Defined rollback conditions • Implemented strategy for maintenance after approval of the AI system • Definition of handover and handshakes to service providers • Documentation of the status at the time of approval as a reference for subsequent updates • Availability of licenses for the product 		<ul style="list-style-type: none"> • Agreed operating concept • Maintenance schedules • Contingency plans • Responsibilities • Authorizations 		

Process phase 7: Preparation for use				
7.3	Is a secure rollout process planned?	Applicable to risk class		
		AIQM-1	AIQM-2	AIQM-3
		No	Yes	Yes
Assessment-relevant requirements		Examples for implementation		
<p>A rollout process for the AI system must be created.</p> <p>This should take into account the following aspects:</p> <ul style="list-style-type: none"> IT infrastructure of the future live system (including protection of the cloud against data outflow / data loss) Training concept 		<ul style="list-style-type: none"> Rollout planning Responsibilities Authorizations IT infrastructure planning Training 		

Process phase 8: Application integration				
8.1	Has an access and identity management plan been defined and implemented for the AI system, the AI application and relevant datasets?	Applicable to risk class		
		AIQM-1	AIQM-2	AIQM-3
		No	No	Yes
Assessment-relevant requirements		Examples for implementation		
<p>The access rights and identity management for the AI system are defined and implemented in an access and identity management plan. This takes particular account of:</p> <ul style="list-style-type: none"> Roles AI application Datasets Documentation 		<ul style="list-style-type: none"> Overview of roles and persons Responsibilities Authorizations Examples of access rights and identity management: Who has access to training data, which confidentiality level is involved, who modifies hyperparameters, conducts training, etc. GDPR requirements must be ensured in particular. 		

Process phase 8: Application integration				
8.2	Has the documentation for data, algorithm, hardware and software been created and is versioning ensured?	Applicable to risk class		
		AIQM-1	AIQM-2	AIQM-3
		No	No	Yes
Assessment-relevant requirements		Examples for implementation		
<p>The documentation of the data, algorithms, hardware and software has been completed, including the versions of the codes and packages used for the final AI solution in operation. The scope of the required documentation depends on the risk assessment and the ability to recreate the solution based on the test and training datasets used for the approval.</p> <p>The documentation must be complete.</p> <p>It must be ensured that the current operational version is documented for each period.</p>		<p>Documentation of the AI system</p> <p>The most important aspects are:</p> <ul style="list-style-type: none"> • Data origin and characteristics of the datasets should be documented (structured data sheets with detailed information on data processing, data collector and data collection method). • Characteristics of the AI system should be documented. • Architecture or model graph (number of layers, parameters, connectivity, input-output dimensions) • Expected input data (structure of the data entered into the AI system) • Expected output data (structure of the AI system output) • Parameter accuracy (e.g. 8/16/32 bit) • Hardware requirements • Training method (e.g. online/offline/etc.) • System architecture • Information flow 		

Process phase 9: Verification of performance				
9.1	Is the planned application area fulfilled by the AI system?	Applicable to risk class		
		AIQM-1	AIQM-2	AIQM-3
		Yes	Yes	Yes
Assessment-relevant requirements		Examples for implementation		
<p>It must be determined whether and how the AI system fulfills the functions and objectives planned for the AI system (see also question 1.1 on defining the functions and objectives). This verification is carried out, for example, on the basis of suitable datasets that are independent of the test and verification data. The results achieved by the AI system are compared with the planned results in accordance with the application area and evaluated.</p>		<ul style="list-style-type: none"> • Test results • Reports • Assessment of deviations 		

Process phase 9: Verification of performance				
9.2	How robustly does the AI system react under extreme loads? Are system limits defined and monitored?	Applicable to risk class		
		AIQM-1	AIQM-2	AIQM-3
		No	Yes	Yes
Assessment-relevant requirements		Examples for implementation		
<p>Stress tests are documented. The following tests are recommended:</p> <ul style="list-style-type: none"> • Infrastructure stress • Test model performance under extreme conditions • Confront model with borderline and edge cases • Adversarial examples specifically designed to confuse the model 		<ul style="list-style-type: none"> • Test results • Assessment of robustness 		

Process phase 9: Verification of performance				
9.3	Is the performance of the AI system verified under normal conditions and taking into account different test methods?	Applicable to risk class		
		AIQM-1	AIQM-2	AIQM-3
		No	Yes	Yes
Assessment-relevant requirements		Examples for implementation		
Test plans are created and implemented: <ul style="list-style-type: none"> • Supervised testing • Unsupervised testing • Comparison with reference systems Deviations have been addressed		<ul style="list-style-type: none"> • Test plans with acceptance criteria • Test results • Measures with responsibilities, deadlines and status 		

Process phase 9: Verification of performance				
9.4	Are verification results for each AI system version documented in a complete, traceable and auditable manner?	Applicable to risk class		
		AIQM-1	AIQM-2	AIQM-3
		No	No	Yes
Assessment-relevant requirements		Examples for implementation		
Complete, auditable documentation is available for all verification results, metrics, explanation activities and fairness activities.		<ul style="list-style-type: none"> • Documentation of the verification results, e.g. in a model card 		

Process phase 10: Go-live				
10.1	Has responsibility for the AI system been transferred to the customer?	Applicable to risk class		
		AIQM-1	AIQM-2	AIQM-3
		Yes	Yes	Yes
Assessment-relevant requirements		Examples for implementation		
<p>Inform the customer about the use of AI, depending on the company's focus and legal requirements. Ensure that the customer handover process complies with the transparency requirements of the EU AI Act and provides customers with clear and understandable information about the functionality, limitations and potential risks of the AI system. This includes establishing agreed procedures for customer support and addressing potential issues or concerns.</p> <p>If the AI system or AI outputs are reused, customers and their use of the AI outputs in the new context must be covered by the AI system specification.</p>		<ul style="list-style-type: none"> • Documentation of the AI system • Documentation of transfer of responsibility, e.g. by contract, handover report 		

Process phase 11: Continuous improvement				
11.1	Has data monitoring been set up?	Applicable to risk class		
		AIQM-1	AIQM-2	AIQM-3
		No	Yes	Yes
Assessment-relevant requirements		Examples for implementation		
<p>A data monitoring process is in place for the outputs of the AI system and is active.</p> <p>Metrics for data quality, target values and mechanisms for implementing corrective measures in the event of deviations are defined.</p>		<ul style="list-style-type: none"> • Process description • Definition of KPIs and targets 		

Process phase 11: Continuous improvement				
11.2	Has model monitoring been set up?	Applicable to risk class		
		AIQM-1	AIQM-2	AIQM-3
		No	No	Yes
Assessment-relevant requirements		Examples for implementation		
<p>A model monitoring process is in place for the AI system and is active.</p> <p>Metrics for model quality, target values and mechanisms for implementing corrective measures in the event of deviations are defined.</p>		<ul style="list-style-type: none"> • Process description • Definition of KPIs and targets 		

Process phase 11: Continuous improvement				
11.3	Has performance monitoring been set up for the runtime environment in live operation?	Applicable to risk class		
		AIQM-1	AIQM-2	AIQM-3
		No	Yes	Yes
Assessment-relevant requirements		Examples for implementation		
<p>A performance monitoring process for the runtime environment in live operation is in place and active.</p> <p>Performance metrics, target values and mechanisms for implementing corrective measures in the event of deviations in the runtime environment are defined.</p>		<ul style="list-style-type: none"> • Process description • Definition of KPIs and targets 		

Process phase 11: Continuous improvement				
11.4	Has a change management system been set up?	Applicable to risk class		
		AIQM-1	AIQM-2	AIQM-3
		No	Yes	Yes
Assessment-relevant requirements		Examples for implementation		
<p>Changes to the operational AI system are made in accordance with a defined change process and a rollout plan. A rollback to the previous version is possible at any time.</p>		<ul style="list-style-type: none"> • Process description • Defined procedure for rollback • Roles and authorizations 		

Process phase 11: Continuous improvement				
11.5	Is a monitoring and retraining/update procedure in place?	Applicable to risk class		
		AIQM-1	AIQM-2	AIQM-3
		No	Yes	Yes
Assessment-relevant requirements		Examples for implementation		
<p>Define or estimate how often the process changes and how often the model should be updated or tested.</p> <p>Check whether the input data is still available in the defined data distribution.</p> <p>Updating means that the model must be re-trained or replaced by a new model.</p> <p>As a rule, a model update is required if the database and/or the system has changed (due to new findings or new functions), for example:</p> <ul style="list-style-type: none"> - Change in the computing environment/measurement system - Update of the machine firmware - Introduction of new technologies - Change in the KPIs - Need for security and privacy enhancements due to new customer/regulatory requirements or cyber security findings. <p>Generative AI: When using generative AI, models need to adapt to evolving data, requiring retraining strategies and degradation monitoring.</p>		<ul style="list-style-type: none"> • Process description • Defined procedure for roll-back • Roles and authorizations 		

6 Recommended actions and application examples

The aim of this chapter is to define recommendations for how AI applications can be operated and adapted and how their outputs can be properly interpreted and evaluated.

AI is not viewed as an isolated technology, but rather as an integral part of a modern, data-based QM system – from the provision of knowledge to process analysis and automated evaluation.

In order to make the use and validation of these application examples traceable, they are each described according to the categories listed below.

The **recommended actions for the example applications** are divided into the following structure:

- **Description:** What is the use case? Brief and specific: objective, functioning and output of the AI system.
- **Framework conditions:** What requirements must be met for the use case to function correctly and reliably (e.g. content, sources, responsibilities, infrastructure)? What measures can reduce risk in the use case?
- **Added value:** The concrete benefits in terms of quality, time, costs, safety or collaboration. What is improved compared to the previous procedure?
- **Challenges:** Typical stumbling blocks, risks and difficulties that can arise (e.g. functional, technical, organizational, legal).
- **Procedure:** How to implement and operate the use case (e.g. setup, configuration, test/validation, pilot, rollout, maintenance).
- **Example:** A tangible, real usage situation with a concrete process.
- **Dealing with changes (based on the example):** How are changes made quickly and in a controlled manner without requiring a new overall approval (e.g. change classification, guard rails, process description for changes, etc.).
- **Interpretation and evaluation of the AI output (based on the example):** How is the AI output reviewed for correctness and evalu-

ated according to clear guidelines so that it can be used in a transparent, contextually appropriate and methodologically sound manner (e.g., through transparency of AI-generated content, validation using reference examples, feedback mechanisms, etc.)?

The AI processes described below are intended solely for guidance and for the systematic classification of possible solutions. The list does not claim to be exhaustive and does not necessarily reflect the current state of the art. Which AI methods are suitable for a specific application must always be evaluated on a project-specific and case-by-case basis (depending on the objective, data situation, risks, regulatory requirements and the availability and maturity of the technology, among other things).

The following AI methods are considered:

- **Language and text-based AI (natural language processing (NLP), large language models (LLM), retrieval augmented generation (RAG))**

These systems enable the semantic processing, structuring and generation of texts. They understand natural language input, classify content, recognize correlations and provide contextually appropriate answers. Technologies such as chatbots fall into this category and can be used, for example, to provide qualified answers to questions about legal regulations.

- **Multimodal AI:**

Multimodal AI solutions process different types of input – such as spoken language, images, audio, or text data – within a single model. One example of this is "speech mining," in which spoken descriptions of activities are transcribed and linked with visual or procedural data to generate work instructions.

- **Assistive AI:**

This refers to an assistive form of artificial intelligence that interactively guides users through tasks, asks for missing information and provides context-specific suggestions. Assistive AI systems go beyond traditional analysis functions by generating new content or recommending actions. Multi-agent architectures allow complex tasks to be broken down into specialized sub-steps.

- **Governance and rule-oriented AI:**
This technology combines AI processes with predefined assessment and decision-making logics. Outputs are often produced according to fixed, traceable rules (deterministic). It forms the backbone for the structured application of quality methods, with a focus on explainability, traceability and compliance with regulatory requirements.
- **Agentic AI:**
These systems extend generative models by adding the ability to plan actions and use tools. They can independently manage complex workflows within the QM process by not only creating content but also actively performing actions via defined interfaces (tools) – such as independently compiling complaint data from various systems to prepare for a decision-making process.
- **Computer vision (CV):**
CV is an AI process for the automated evaluation of image and video data. In quality management, CV is primarily used to detect characteristics, deviations and failure patterns objectively, reproducibly and at high speed – either as assistance for a manual inspection or as a fully automated inspection process.
- **Data and process analytics AI:**
Data from processes and other sources, e.g. specifications, sensor measurements and quality inspections, can be used to develop data models for quality assurance in production processes. The patterns recognized in the data enable the detection of anomalies, the prediction of quality characteristics and root cause analyses. In this way, predictive measures can be derived to improve production processes.

6.1 AI-powered optical quality control

Description

Automated machine vision systems are used for optical quality inspection in production. Camera systems capture image data of components or assemblies and then evaluate it in order to detect faults, deviations or irregularities. The image data can be evaluated using both conventional image processing methods and data-driven artificial intelligence methods.

The aim of these systems is to enable robust, reproducible and fast quality control that supports or replaces human inspections and at the same time increases process reliability.

In **rule-based processes**, image data is evaluated using deterministic, pre-defined rules. For example, characteristics such as brightness, contrast, shape or size are analyzed in order to identify objects or structures in the image. These methods are particularly suitable for simple and clearly defined inspection tasks. However, rule-based methods often reach their limits when it comes to more complex image content or widely varying conditions, for example due to different lighting, positions or surface properties.

AI-powered methods are increasingly being used alongside classical image processing methods. These include the following methods:

Object classification: An image or a section of an image is assigned to a known category, for example "weld seam" or "screw." The system therefore recognizes which object class is present in the inspected section, but does not determine the exact position of the object in the image.

Object detection: In this process, the system not only detects which objects are present, but also determines their position in the image. The position of the objects is typically represented by bounding boxes.

Segmentation: Each pixel in the image is assigned to a class or object area. The result is a segmented representation, often in the form of a color mask, which shows which sections of the image belong to which object classes.

Anomaly detection: The system learns, based on examples, the normal state of a product or process. Deviations that indicate unusual or faulty states can then be detected. This method is particularly suitable if rare faults are to be detected and only a few or no sample images of faulty products are available.

There are also other specialized processes.

Depending on the data to be tested, the framework conditions of the image acquisition situation and the requirements for the test, some of the approaches described above may be more or less suitable. The methods

must therefore be selected and assessed based on the respective application.

Framework conditions

Optical quality control is used in a production environment in which high quantities, short cycle times and stable processes are crucial. Components can have different variants, surfaces, materials and manufacturing tolerances, which makes image evaluation challenging. The failure patterns that occur can be quite varied, for example scratches, dimensional deviations or assembly errors. The inspection system must therefore be sufficiently flexible in its configuration. Production lines often only offer limited installation space, so camera and lighting concepts must be precisely adapted to the installation situation. At the same time, external influences such as lighting fluctuations, soiling or vibrations can affect image quality and must be taken into account and compensated for on a technical level when designing the system.

In addition, the solution must be integrated into the existing system and IT landscape. This includes, for example, production control systems such as PLCs, manufacturing execution systems (MES), traceability solutions and, where applicable, edge or cloud infrastructures for data processing and storage. If image data is stored or used for training AI models, data protection and data security requirements must also be taken into account.

The user perspective also plays an important role. Operators and quality staff need comprehensible operating and visualization concepts in order to be able to understand inspection results and adapt the system if necessary. Another prerequisite for the successful operation and continuous improvement of such systems is the availability of specialist expertise in the areas of image processing, quality and production processes.

Overall, the framework conditions for optical quality control include technical restrictions, process-specific requirements and the goal of achieving a robust, stable and scalable inspection solution.

Added value

- **Reproducible and objective quality inspection**, regardless of individual employees' daily performance or experience.
- **Error reduction through early detection** of deviations in the process.
- **Greater process reliability** and lower scrap and rework costs.
- **Seamless documentation** of inspection results and image data for audits and traceability.
- **Stable inspection** even with high cycle times and complex inspections.
- **Relieve employees** from monotonous, tiring visual inspections.
- **Quick adaptation of inspections** in the event of product or process changes (especially with AI support).

Challenges

- **High expenditure for lighting & optics**, as lighting conditions have a significant influence on image quality and fault detection.
- **Variety of variants** that requires flexible inspection strategies and robust training.
- **Quality of training data** for AI-powered systems (failure patterns, good parts, borderline cases).
- **Changes in the product or production process** that can influence the vision parameters.
- **Maintenance during operation** (cleaning of optics, calibration, wear and tear).
- **Integration into the automation environment**, including real-time capability.
- **Acceptance on the shop floor**, especially if AI systems are perceived as difficult to understand (black box).
- **Scaling costs**, especially for multiple lines or locations.
- **Data management and storage of large amounts of image data**, especially for long-

	<p>term archiving for traceability purposes or AI training and validation</p> <ul style="list-style-type: none"> • Validation and safeguarding of AI-powered inspection systems, particularly with regard to robustness, traceability and approval processes.
<p>Procedure</p> <p>VDA 5.3 must be taken into account for the proof of suitability.</p> <p>Note: It is recommended to use the "proof of capability for attribute inspection" described in chapter 6. An AI output (e.g. a confidence score) is not a measured value and is therefore not suitable as a quantitative criterion for demonstrating capability.</p>	
<p>QM roles</p> <p>Employees in production quality</p>	

6.1.1 Example

When inspecting a structural strut variant, the AI needs to evaluate an image, localize the structural strut in the image and then output which variant of the strut is installed. The AI thus handles the image analysis and variant classification, while the human evaluates and uses the output for further quality assurance.

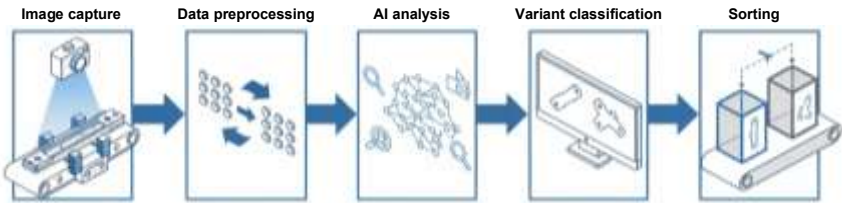


Figure 6-1: Schematic representation of the five-stage process chain in AI-powered optical quality control

Dealing with changes	Interpretation and evaluation of the AI output
<ul style="list-style-type: none"> • Minor changes (class A): In the event of changes to the model, to weights or to the evaluation logic, these must be versioned and the changes documented. If the test result is affected, a new proof of capability must be carried out. • Targeted testing of functionally relevant changes (class B): Significant changes, e.g. in the object to be tested, sensor technology or external influences, should be considered on a case-by-case basis and may require a re-evaluation of the selected test procedure and the required safeguarding measures. • Treat fundamental changes (class C) as a new system version: When new data types are checked with a 	<ul style="list-style-type: none"> • Risk: AI recognizes and classifies a different component instead of the strut → Before the variant decision, the process should check whether the recognized position of the strut is plausible, e.g. by means of a rule-based position comparison with a reference feature such as a screw or a defined fastening point in the image. Only if the position of the AI detection matches the expected position of the strut should the classification be accepted. • Risk: Different camera perspective (viewing angle/zoom) compared to the training/validation set → When interpreting the AI outputs, the process must check whether all relevant reference features

new method or with an established method. Also generally for initial implementations, e.g. new supplier or first application in the organization.

(e.g. screws, edges, drill holes) are visible at the expected positions in the image. If reference features are greatly displaced, only partially visible or not visible at all, the AI output should be evaluated as uncertain and checked manually.

- **Risk: Classification is hallucinated due to light, soiling or occlusion**

→ To confirm the AI output, the size, shape and/or color intensity (e.g. RGB values) of the detected strut can be compared with known expected values from validated reference data. If these characteristics deviate significantly from the expected range (e.g. area too dark due to shadows, atypical color values), the output should be classified as critical and subjected to an additional check.

→ Instead of looking at just one region, the AI should classify several independent areas of the strut (e.g. head area, middle section, connection area). If the classification results of these sub-regions match, this increases credibility. If the results are contradictory (e.g. one area is variant A, another variant B),

	<p>the overall result should be evaluated as uncertain.</p> <p>→ The evaluation of the AI output should be based on several redundant features (e.g. contour, hole pattern, position relative to screw, color/surface feature).</p> <p>If these features match (position match and shape match), this indicates a plausible result. If they do not match, it is a good idea to check them manually or to repeat the image capture.</p>
--	--

6.2 Rule-oriented AI agents to support the 8D process

Description

This application example describes the use of an AI-powered agent system to support the creation and quality assurance review of 8D reports. The system combines generative AI (for drafts and formulations) with rule-oriented process control (for completeness, consistency, verification and release).

Essentially, specialized agents provide assistance in areas such as:

- Collecting information (e.g. from complaint data, blocking/sorting reports, CAQ/DMS)
- Creating structured drafts (in particular problem description D2, optionally also D3/D4 etc.)
- Researching similar historical cases and linking references
- Recognizing information gaps/contradictions and asking specific questions

A higher-level guard agent, which is to be understood as an automated supervisor, checks the outputs based on defined rules (e.g. mandatory content per discipline, consistency between D2–D4, evidence/verification, risk indicators, escalation and release criteria) and controls the process via release points ("gates"). In addition, firmly programmed workflow rules are introduced (e.g. mandatory fields, status logic, blocking in the event of missing evidence, automatic escalation for risk indicators, dual control release), which work independently of prompts and prevent circumventions.

The aim is to reduce recurring documentation effort and at the same time improve the quality, standardization, auditability and release of 8D documentation. The procedure is designed in such a way that the AI provides support but does not replace a specialist's decision. The responsibility and power to release remain with people with the appropriate authorizations and roles.

Framework conditions

- Process & standards: Defined 8D standard in the company (templates, mandatory content per discipline, roles/releases, escalation paths), linked to QM specifications/procedural instructions.
- Data and knowledge base: Pre-defined, testable and versioned rules for completeness, consistency, evidence/verification, risk indicators, escalation and release.
- Roles & governance: Clear user roles, access rights, responsibilities (including release competences).
- Continuous feedback/improvement process (e.g. user evaluations or lessons learned).
- Data protection & information security: Regulations for personal data, confidentiality levels, masking/anonymization as well as specifications for external communication.

<p>Added value</p> <ul style="list-style-type: none"> • Process acceleration through faster drafts and reduced manual documentation. • Higher data and result quality thanks to structured, complete and consistent content for each discipline. • Uniform formulations, traceable documentation of sources and changes improve standardization and auditability. • Hard process rules prevent circumventions. • Missing data is systematically requested. • Knowledge retention through reuse of structural modules, lessons learned and comparability of similar cases. 	<p>Challenges</p> <ul style="list-style-type: none"> • Rules must be unambiguous, consistent, testable and versioned. • Finding the balance between gates that are too strict and block outputs too frequently and gates that are too loose and allow errors. • Strict source and evidence requirements as well as suitable guard mechanisms are necessary to prevent hallucinations and pseudo-accuracy • AI assists but does not decide, which poses a risk of overconfidence. • Data quality and data availability.
<p>Procedure</p> <ul style="list-style-type: none"> • Define use case & scope: Which disciplines are supported, which outputs are binding and which decisions remain with humans. • Build up data/knowledge base: Connect data sources, clarify authorizations, ensure referencing (IDs/links) for evidence. 	

- Define agent roles: For example, recording agent (input/similarity search), formulation agent (D2 draft), analysis agent (optional D4 hypotheses), guard agent (checking/control).
- Set up & version rules: Define completeness, consistency, evidence, risk indicators, escalation and release and determine corresponding owners and the change process.
- Implement hard process gates: Mandatory fields, status logic, blocking without evidence, automatic escalations.
- Validation mechanisms: Plausibility checks, contradiction detection, "is/is not" logic, source requirement for critical outputs.
- Pilot & rollout: Test cases/regression tests, evaluation of rule matches/false positives, training, gradual expansion.

QM roles

Employees in supplier quality, employees in customer quality, employees in production quality

6.2.1 Example

A customer complains about sporadic functional failure. The responsible QM employee starts an 8D report and initially enters a few key pieces of data (series, component, function, customer error description).

- The recording agent automatically searches for similar cases in released historical 8D reports and complaints and returns hits including references. The QM employee marks which cases are actually relevant.
- The formulation agent creates a structured D2 draft based on the available data (e.g. with occurrence conditions, frequency, delimitation, symptoms/warning messages). It recognizes missing information and makes specific queries to the QM employee (e.g. time period/production status, variants, framework conditions).

- The guard agent checks the draft against rules (e.g. mandatory content, consistency, source/evidence references) and sets a gate status if necessary (e.g. "open," "additional information required" or "ready for release").
- Hard workflow rules prevent the process from continuing if mandatory fields/evidence are missing (e.g. blocking certain status changes or dual control review).

All AI suggestions remain editable. Changes are logged transparently, including information on whether they were generated by AI or edited by humans. Only authorized persons can approve changes.



Figure 6-2: Schematic representation of the rule-oriented AI agent workflow in the 8D process

Dealing with changes	Interpretation and evaluation of the AI output
<ul style="list-style-type: none"> • Define change classes: Subdivide edits to the AI system into classes – based on the process described above involving the recording agent, formulation agent, guard agent and strict workflow rules: • A = purely linguistic/ cosmetic (does not change hit list, D2 content or gate status) 	<ul style="list-style-type: none"> • AI-generated content should be regarded as drafts/suggestions (not automatically carried over to the 8D report). • Systematic use of historical knowledge: The agent suggests similar cases. The user actively selects which cases are actually comparable/relevant (e.g.

- B = functionally relevant within the same framework (e.g. improves search/rules/prompts, but no new role for the AI, no new risk profile)
- C = fundamentally changes behavior or risk profile (e.g. AI evaluates causes/actions or changes guard rails/escalation logic). Class C triggers a new approval process.
- **Minor changes (class A):** Changes of a purely linguistic or cosmetic nature (e.g. clearer wording, additional examples, typo corrections) that do not change the functional behavior of the system.

Example: Rewording of text modules that the formulation agent uses for the D2 draft without changing the content structure/test logic.

Example: Adjusting help texts/labels in the input mask (series, component, function, error description).

Can be used following a brief dual-control review on a functional basis with subsequent documentation, without the

with regard to change status, variant, line/system, framework and environmental conditions). This selection (including justification/labeling) is documented in the system so that it is clear that past experience has been deliberately included.

- **Statements with relevance for release or risk must be referenceable** (source/evidence/reference to case, document, measurement value, ticket, etc.).
- **Critically examine hits from similar cases** (e.g. change statuses, variants, lines, systems or environmental conditions) and only accept after plausible comparability.
- **Transparent labeling and confirmation of AI suggestions:** All text parts suggested by the AI are clearly recognizable as such in the system. They must be actively adopted or adapted before they become part of the 8D report. The history

need for additional formal approval.

Primarily serve to improve comprehensibility and usability, not to expand content.

- **Functionally relevant changes (class B):**

Changes that extend or sharpen the functional assistance behavior without changing the basic purpose or risk profile of the system (e.g. additional data sources for similar cases, new checking rules within the existing framework).

Examples in context

The recording agent uses additional approved data sources (e.g. additional internal complaints database) for similarity searches, but stays with "deliver hits + human marks as relevant."

The formulation agent asks more precise queries (e.g. production status/period/variants/framework conditions) or generates a more consistent D2 draft without "deciding" on new 8D steps.

The guard agent receives additional rules (e.g. stronger consistency check,

shows which AI suggestions have been adopted, changed or rejected.

- **Validation with test cases:** Before and during deployment, known complaint cases are used as tests.

AI-powered problem descriptions are compared with proven reference 8Ds. Experts evaluate completeness, clarity and methodological consistency. Deviations are documented and used for improvement (e.g. rules/prompts/test catalog).

clearer evidence/source references), but still only sets gate status and requests additions (no automatic release).

Must be requested via a defined change process, evaluated on a functional basis and tested with representative 8D test cases (including regression test, e.g. with historical cases of "sporadic functional failure").

After documented testing and release, these can be carried over to the live application.

- **Fundamental changes (class C):** Changes that fundamentally alter the behavior or risk profile of the system (e.g. new use case in which the AI evaluates or proposes causes or actions, changed guard rails, new escalation logic).

Examples in context:

AI evaluates causes or actions (D4/D5) or prioritizes actions in a binding manner instead of only providing assistance.

Change in guard rails, e.g. softening of "release exclusively by humans" or "AI suggestions are labeled."

Guard/workflow logic is changed so that status changes are also possible without evidence/dual-control or, conversely, a new escalation is introduced (e.g. automatic notification/stop).

Treated like a new system version and require a complete reassessment (risk, approval criteria, tests, documentation).

Guard rails such as "AI does not decide on causes/actions," "AI suggestions are labeled" and "release exclusively by humans" may only be changed with special justification and advanced release.

- **Note:** The change process (how changes are classified, checked, tested, documented and transferred to the application) should be described in a separate process – including logging whether content was AI-generated or adapted by humans, and with clear roles/responsibilities for release.

6.3 AI-powered audit

Description

An AI agent analyzes the audit findings of an internal process audit carried out in accordance with VDA Volume 6.3, for example. Using an existing knowledge database, the AI can suggest possible corrective measures for the respective deviations.

Rule- and governance-oriented AI supplements the analysis with fixed evaluation logic and defined decision rules. As a result, findings are interpreted consistently, rule violations are clearly assigned and proposed measures are documented in a traceable manner. Audit trails and rule references ensure traceability, while governance mechanisms support the maintenance of rules, roles and knowledge levels.

The result is a structured, reproducible and explainable AI-powered audit process.

Framework conditions

- Consistent knowledge database and clearly defined audit rules required (e.g. VDA 6.3, internal audit standards). The knowledge base includes, in particular, the company's understanding of the process, previous audit reports and a central company FAQ. In addition, the AI requires structured and, if possible, standardized audit documents in order to reliably evaluate findings.
- Definition of responsibilities, user groups and multi-level access based on expertise and feedback quality
- The competence to evaluate the AI's proposed actions should be retained.
- Integration into the existing audit management system (e.g. to monitor effectiveness)
- Integration of user feedback via thumbs up/down or active text feedback
- Change management for content and data maintenance (e.g. updating FAQs)

<ul style="list-style-type: none"> • Definition of fixed categories and priorities. Each recommendation must have a rule reference and documents from the history. 	
<p>Added value</p> <ul style="list-style-type: none"> • Faster, consistent derivation of actions based on the organization's experience. • Rule- and governance-oriented decision-making logics make assessments more consistent, traceable and reproducible. • Identification of actions that have proven to be particularly effective in comparable cases. • Automated prioritization possible. • Automatic exchange and building of knowledge, also with other locations 	<p>Challenges</p> <ul style="list-style-type: none"> • Quality of existing audit reports is crucial. • Inconsistent audit reports can make documentation more difficult. • Traceability of the selected corrective actions must be ensured. • Maintenance of the knowledge database and the underlying rules requires clear responsibilities and regular updating.
<p>Procedure</p> <ul style="list-style-type: none"> • Definition and collection of relevant data. • Creation of a knowledge database with the relevant data including rules and regulations, history of actions taken and company FAQs. • Development of the corresponding AI functions such as text analysis of audit reports and suggestion generation for rule-based derivation of actions. • Implementation of a pilot project with a selected group of users to collect user feedback and adapt the model. • Comparison between the actions proposed by the AI and the auditor assessment to ensure consistency and traceability. 	

- Integration into the audit tool. Provision of the proposed actions directly in the audit software.
- Implementation of a continuous change and improvement process to maintain the knowledge base and decision-making logic.

QM roles

Quality auditor / quality assessor

6.3.1 Example

Internal process audits are carried out in the company in accordance with VDA 6.3. The auditor assigns each finding directly to the corresponding question (e.g. P5, P6, P7).

Since the findings were already assigned to the VDA 6.3 questions during the audit, the AI uses this information as an input value.

The AI then carries out the following tasks:

- Analyzing the text of the findings
- Classifying the deviation based on predefined rules (e.g. systemic, documentation-based)
- Comparing with historical audit cases from the knowledge database
- Proposing suitable actions including rule references
- Providing traceable justifications (audit trail)

Presentation of the results:

Audit findings: production management "Work instruction for setup process not up to date and not available at the workplace."

AI analysis

- Classification: documentation deviation
- Historical cases detected: similar deviations in 2023/03 (audit 2311) and 2024/01 (audit 8512)
- Proposal of the AI:
 - Update and release the work instruction
 - Ensure availability at the workplace
 - Check in the next review cycle

Justification: "Action complies with established rules for document management; effectiveness proven in previous cases."

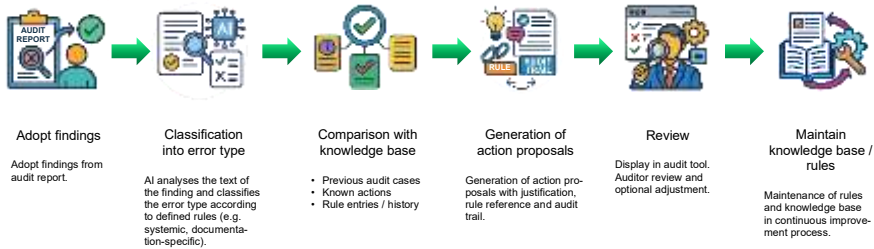


Figure 6-3: Six-step workflow for AI-powered processing of audit findings

Dealing with changes	Interpretation and evaluation of the AI output
<ul style="list-style-type: none"> • Define change classes: Subdivide adjustments to the AI system into classes (e.g. A: purely linguistic/cosmetic, B: functionally relevant but within the same framework, C: fundamental change in behavior). Class C would trigger a new approval process. • Minor changes (class A): In the case of minor changes (e.g. clearer wording, additional examples, typo corrections), a brief dual-control review on a functional basis is sufficient. After documentation, these changes can be applied directly without a new approval. • Targeted testing of functionally relevant changes (class B): Additions that more precisely specify the behavior but do not change the role of the AI – e.g. new categories, refined 	<ul style="list-style-type: none"> • Work with clear rules and a knowledge base: The AI is based on defined rule sets (e.g. VDA 6.3, internal audit standards), fixed categories (e.g. systemic, documentation-based) and a verified knowledge database with historical audit cases and FAQs. Each recommendation contains a rule reference and historical examples. This makes it clear which rules and historical cases a recommendation is based on. • Use structured, standardized input data: Findings are already assigned to the relevant VDA question during the audit (e.g. P5, P6, P7) and are available in a structured format. The AI uses this structure, classifies the deviation according to predefined criteria and searches specifically for suitable cases. This reduces the

classification rules, additional standard actions, new rule references – are checked with a small set of test audit cases. Auditors assess whether classification, proposals and justifications are still accurate and consistent.

- **Treat fundamental changes (class C) as a new system version:** Adjustments that change the risk profile of the system (e.g. automatic assessment of audit questions) require a new approval.
- **Define guard rails:** These guardrails should not be changed lightly. Typical guard rails would be, for example, the AI only makes suggestions, not final assessments; each recommendation contains a rule reference and references to historical cases; actions are always decided and approved by auditors. If these guard rails remain unchanged, the application does not require a new approval.
- **Note:** The change process, i.e. how changes are classified, checked, tested, documented and implemented in the application, should be described in a process.

scope for interpretation and ensures reproducible results with the same data basis.

- **Traceable suggestions with source:** The AI provides a justification for each recommendation (e.g. "Documentation deviation – actions complies with document management rule; assessed as effective in cases X and Y").
- **Human review and release of the actions:** The auditor sees the AI suggestions including classification, rule references and history. The auditor decides which actions to accept, adapt or reject.
- **Feedback mechanisms as quality control:** Auditors can rate suggestions with "thumbs up/down" or provide text feedback ("suggestion too general," "very appropriate"). This feedback is used to improve the rules and knowledge base in a targeted manner.
- **Validation with reference audits and key figures:** In pilot phases and regularly thereafter, AI-powered suggestions for action are compared with existing, manually created audit assessments. Experts check whether the AI classifies

	<p>consistently, proposes suitable actions and whether the rule references are correct. In addition, key figures can be evaluated (e.g. percentage of suggestions accepted by auditors, number of improvements after external reviews).</p>
--	---

6.4 AI-powered FMEA

<p>Description</p> <p>An AI-powered solution for the assisted creation of FMEAs in accordance with VDA guidelines and company-specific specifications.</p> <p>The agent-based system assists moderators and teams with knowledge-based suggestions on structure and formulation and ensures consistent documentation. By providing context-specific assistance, the solution helps with data entry and to comprehensively describe and minimize risks. It also provides suggestions for actions and formulations based on historical FMEA data sources.</p>
<p>Framework conditions</p> <ul style="list-style-type: none"> • Availability a validated FMEA knowledge base • Availability of a standardized FMEA data model • Development of a methodical model to ensure the methodical correctness of the FMEA analysis • Integration of feedback options for users to improve the solution • Definition of access rights and responsibilities in the system • Implementation of logic rules (e.g. failure sequence matches the cause).

- Only rule-compliant entries are accepted.

Added value

- Higher quality risk analyses and minimization strategies, in accordance with VDA standards
- Structured input processes to assist the user
- Acceleration of the process by using existing knowledge of the risk analysis methodology to support the team
- Improved data quality through more consistent and precise risk derivations and minimizations
- Less manual effort for the initial description of the structural functional and failure analysis

Challenges

- Ensuring that the data provided by the user to build the AI system is accurate.
- Finding the right balance between automation and necessary reviewing by experts
- User acceptance and training in the use of the AI solution
- Data protection and security for sensitive quality data

Procedure

- Use of an AI solution that is based on a structured knowledge database (historical FMEAs, guidelines, company standards) and serves as an assistance system to support FMEAs
- Development of a structured knowledge base with historical FMEAs
- Implementation of pre-formulated prompts to guide moderators and teams through the FMEA creation process
- Piloting with selected experts, FMEA moderators and decision-makers to collect feedback and make iterative improvements to the solution
- Gradual expansion of the user group depending on the confirmed response quality and user satisfaction

- Scaling of the deployment after a successful pilot phase and continuous adaptation of the knowledge base and the user guidance
- Implementation of a continuous change and improvement process to maintain the knowledge base

Roles:

Employees in development quality, employees in production quality

6.4.1 Example

A development team is working on a new control unit for a vehicle. Before the product goes into series production, it needs to be determined where possible failures can occur and how they can be avoided. The team uses an FMEA, i.e. a structured risk analysis, to do this: They describe what the component is supposed to do, consider what can go wrong, what effects this would have, why this could happen and what measures are necessary to prevent failures or detect them in good time.

Instead of rebuilding everything from scratch, the FMEA moderator uses an AI-powered assistance system. She briefly enters what the issue is – such as "control unit for vehicle dynamics control in vehicle platform Z." The AI accesses a knowledge base from previous FMEAs, guidelines and internal standards and makes suitable suggestions: typical functions of the control unit, possible failures, probable causes, typical effects and proven countermeasures. The team selects, adapts and adds what is relevant for the specific case.

For example, the AI could suggest that a "communication breakdown to the sensor" can occur, with the effect that "vehicle dynamics control not available, warning message for the driver" and root causes such as "loose plug connection" or "cable break." It suggests measures such as more robust connectors or a full test in the end-of-line test. The team reviews these suggestions, decides what to adopt and determines who implements which measure and by when.

Meanwhile, the AI checks in the background whether the FMEA is logical and complete: whether failures have been recorded for important functions, whether critical failures have causes and countermeasures and whether the

combinations of failures, causes and effects are plausible. It points out to the moderator if something is missing or does not fit together. The end result is an FMEA that was created more quickly, is based on tried-and-tested knowledge and is methodologically consistent – but all expert decisions remain with the team; the AI only provides support with suggestions and checks.

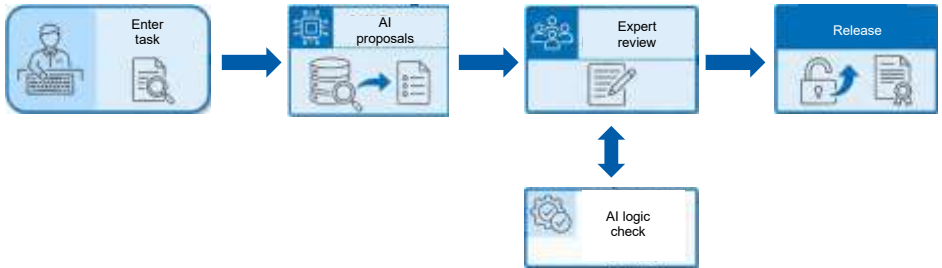


Figure 6-4: AI as an assistance system in FMEA

Dealing with changes	Interpretation and evaluation of the AI output
<ul style="list-style-type: none"> • Define change classes: Subdivide adjustments to the AI system into classes (e.g. A: purely linguistic/cosmetic, B: functionally relevant but within the same framework, C: fundamental change in behavior). Class C would trigger a new approval process. • Minor changes (class A): In the case of minor changes (e.g. clearer wording, additional examples, typo corrections), 	<ul style="list-style-type: none"> • Documentation of rule checks: The system logs which content-based and logical rules have been checked (e.g. "There is at least one potential failure for each function," "Each cause is linked to a failure," "Each error with high significance has at least one measure for prevention or detection.") • Documentation in the audit: Which rules are defined; which instance of the FMEA was checked and what was the result?

a brief dual-control review on a functional basis is sufficient.

After documentation, these changes can be applied directly without a new approval.

- **Targeted testing of functionally relevant changes (class B):**
Changes that expand the assistance behavior but do not change the role of the AI (e.g. new standard modules for typical failures/causes) are tested and documented with defined test cases and expert reviews.
- **Treat fundamental changes (class C) as a new system version:**
Adjustments that change the risk profile of the system (e.g. new use case in which the AI independently proposes/determines assessments) require a new approval.
- **Define guard rails:**
These guardrails should not be changed lightly.

- **Moderator releases:** Each FMEA has a clear release history (e.g. who released which version and when, with which release status).
- **Labeling of AI contributions:**
The system labels which entries (failures, causes, effects, measures) are based on AI suggestions; at the same time, it is documented that a human (moderator/expert) has approved or adapted these suggestions.
- **Validation of the results by comparison with reference FMEAs:** Existing FMEAs that are accepted as being of good quality are used as a reference. (The AI generates suggestions for a known use case and these are compared by the experts and moderators. Does the AI cover the same or more relevant failures/causes/effects? Do the proposed measures make comparable sense?).
- **Validation of the results by independent expert reviews:**
Independent experts (not involved in the project) randomly check FMEAs that were created with AI assistance (assessment of completeness (are significant risks covered?); assessment of plausibility (do causes and

<p>Typical guard rails include, for example, the AI only makes suggestions, does not decide; it does not make any final evaluations; every FMEA entry is checked by a moderator/expert; only rule-compliant entries are accepted; FMEA releases are only made by authorized persons. If these guard rails remain unchanged, the application does not require a new approval.</p> <ul style="list-style-type: none"> • Note: The change process, i.e. how changes are classified, checked, tested, documented and implemented in the application, should be described in a process. 	<p>measures fit?)), including review reports as proof.</p> <ul style="list-style-type: none"> • Possible key figures: Number of subsequent FMEA changes due to complaints / returns from the field; time required for FMEA creation (before/after AI use); completeness indicators (e.g. average number of errors in logic and consistency before/after introduction) → Improvements in these key figures over time offers proof of process reliability. • Tests of the methodology rules: Can you create fake FMEAs with deliberately included errors and check whether the system recognizes them? Are violations of the VDA methodology reliably reported? Documentation of the test cases and test results as proof of process qualification for the tool.
--	--

6.5 Predictive process control

Predictive process control (e.g. predictive quality) enables proactive quality assurance in production. Statistical methods such as SPC, ML methods or a combination of different approaches can enable early corrective measures. The resulting data models can be used for root cause analysis, for example with the help of classical tools such as Ishikawa, previously developed causal models and/or explainable AI methods (XAI).

Description

Predictive process control can be implemented in different ways. A key aspect of the design is the time between a prediction and the initiation of corrective or improvement measures based on the findings. The aim is to improve process stability and capability while reducing the amount of testing required. The added value of predictive process control is at its highest with complex or destructive testing and where there is a long time between the prediction of a quality characteristic or detection of the deviation and the opportunity to take countermeasures.

Framework conditions

- **Definition of use case:** Objectives, metrics (e.g., scrap rate, accuracy, ROI), applicability, scope, requirements, team and departments.
- **Expert knowledge and training:** Involvement of experts to assess data and forecast quality and to check assumptions, plausibility of model results and derived insights. Training of the relevant stakeholders with regard to the development, use and maintenance of the solution.
- **Data availability and quality:** Completeness, correctness, consistency, timeliness and relevance of data (e.g. specifications, sensor measurements and quality checks). The data must be available in sufficient quantity for modeling.
- **Automated collection of measurement data:** Successful predictive control is based on the correct collection of sensor data, which enables precise predictions and appropriate measures.
- **Data integration:** Integration and aggregation of data from different sources for analysis and modeling.
- **Standardization and data management:** Uniformity of formats for collecting, exchanging and managing data. Continuous maintenance of data to maintain and/or improve data quality.
- **Threshold logic:** Combination of AI prediction with predefined thresholds.

- **Supervision and review:** Automated interventions only occur after threshold is exceeded and human approval is obtained (justification required).
- **Validation:** Implementation of pilot projects to test the effectiveness of the solutions in controlled environments.
- **Governance and data protection:** Definition of responsibilities and compliance with existing standards and regulations.
- **Infrastructure and technical interfaces:** Suitable hardware and software for solution development and ensuring successful technical integration into existing systems.

Added value

- **Improving product and process quality:** Predictive initiation of improvement measures based on data.
- **Improved understanding of the process:** Possible insights into interrelationships within the process and optimal parameter settings.
- **Resource efficiency:** Reduction in inspection quantities (for random sample inspection), required resources and associated costs.

Challenges

- **Data quality:** Process changes influence data quality, e.g. calibration and sensor wear.
- **Unbalanced datasets:** Datasets may be biased in terms of the balance between states to be classified or predicted, e.g. defects, which could affect the uncertainty of the models.
- **Product and process complexity:** Complex, non-linear relationships between process parameters and quality characteristics.
- **Traceability and acceptance:** Low acceptance by specialist departments due to lack of

	<p>model explainability (black box effect).</p> <ul style="list-style-type: none"> • Technical integration and interfaces: Integration into existing IT and production systems (e.g. MES) and processes.
<p>Procedure</p> <ul style="list-style-type: none"> • Definition of use case: Objectives, metrics (e.g., scrap rate, accuracy, ROI), scope, coverage, requirements, team and departments. • As-is assessment: Analysis of affected processes, process parameters and quality characteristics as well as initial analysis and evaluation of existing data. • Data preparation and modeling: If necessary, adaptation or expansion of the data requirements, collection of relevant data and subsequent training of the predictive model in a test environment. • Testing and validation: Implementation of the solution in production and derivation of improvement measures. Iterative improvement of the trained model with comparison of the defined metrics. 	
<p>QM roles</p> <p>Employees in production quality</p>	

6.5.1 Example

An interdisciplinary team (e.g. production management, production quality, process development, technology experts, IT, data scientist) works to minimize rework and destructive testing in welding processes, e.g. in resistance spot welding (car body construction) or laser beam welding (e.g. transmission parts or battery packs). By predicting critical points, the number of manual inspections of welded joints is to be reduced, the scrap rate lowered and plant downtimes avoided. This application requires data from quality inspections (e.g. ultrasonic testing), measurable environmental data (e.g. operating equipment, material batch, temperature) and relevant controllable

process parameters of the plant (e.g. welding time, welding speed, welding current and laser power, depending on the technology).

The required data is collected from, for example, quality inspections, sensor data and/or MES systems and can be stored separately for the development of test applications. During data preparation, initial analyses of data distributions and outliers are carried out. The quality of the data is evaluated (e.g. completeness, interpretability and assignability of the quality data). With the help of experts, the data is cleaned and prepared for modeling (depending on the data format, this can be a time-consuming step).

In a classification analysis, the data is labeled (OK and NOK), thresholds and metrics (e.g. prediction intervals and classification metrics) are defined for the AI models and, if necessary, additional features are developed from the existing data in order to better map the existing process and goals of the use case. Different algorithms are selected (e.g. neural networks, random forests) to train the models and evaluate the prediction accuracy (uncertainty).

The resulting models are optimized by testing various hyperparameters (model parameters) and tested using additional historical data. Sensitivity analyses or XAI methods can be used to determine the influences of various parameters in the model on the prediction results (e.g. parameter limits for a prediction as OK or NOK) and discussed with the team.

After successful iterative validation in a test environment, measures for automated correction of the identified critical process parameters are derived. These can include, for example, automatic adjustment of the welding source parameters or automatic error alarms to scrap the faulty part at an early stage. Integration into production is piloted in the existing IT architecture in collaboration with the team and continuously monitored for effectiveness and potential improvements.



Figure 6-5: AI-powered control loop for predictive process control

Dealing with changes	Interpretation and evaluation of the AI output
<ul style="list-style-type: none"> • Define guard rails: These guard rails can be application-specific thresholds such as prediction intervals, classification metrics or tolerance fields for the statistical variability of the input variables. Exceeding or falling below these thresholds can trigger a change according to change classes A, B or C. • Define change classes: A: retraining, updating the model, B: adjustment of features and/or thresholds; 	<ul style="list-style-type: none"> • Monitoring the detection and false alarm rates (checking false positives). • Monitoring of values exceeding or falling below defined guard rails and metrics. • XAI methods can be used for each prediction to indicate the influence of a specific parameter on the prediction.

C: adjustment of model architecture or scope of application.

- **Change class A (retraining, updating the model):** Retraining of the model with new data. After documentation, these changes can be applied directly without a new approval.
- **Change class B (adjustment of features and/or threshold values):** Addition or modification of features in the model (e.g. humidity). After documentation, testing of the improvement in a test environment and an expert review, these adjustments can be released.
- **Change class C (adjustment of model architecture or scope of the application):** Change of the model class, if not previously tested in development (e.g. from neural networks to random forests) or change of scope (e.g. prediction of other downstream process variables). After documentation, testing of the new model or application in a test environment and a more extensive

- Plausibility and consistency check of the results. Deviations are reviewed by experts to check for possible changes of class A, B or C.
- Documentation of manual spot checks and comparisons with measurement results.

expert review, these adjustments can be approved..

- **Documentation:** Versioning of data and models, documenting changes according to change classes: Storage of the tests performed and hyperparameters of the model, documentation of changes to features and thresholds as well as documentation of the algorithms and comparisons used.
- **Data monitoring:** Regularly check whether data scope or data quality has changed significantly (e.g. missing values, new value ranges). Calibration of sensors and monitoring of data quality. If there are any noticeable changes, identify the cause.

6.6 Preventive maintenance

Data from plants, machines, processes and other sources, e.g. technical specifications, can be used to develop data models for detecting anomalies and predicting possible malfunctions and downtimes in production. This approach enables the implementation of preventive measures to increase plant and machine availability, as well as the identification of root causes for warranty claims.

Description

Predictive maintenance is a maintenance strategy in which the condition of plants and machines is continuously monitored and maintenance measures are carried out in good time on the basis of predictions.

The data models used for this are based on historical data and can include various data sources and sensor data such as temperature, vibration and pressure. Using statistical and AI-powered methods, data can be analyzed to identify trends and predict potential failures. Added value includes the ability to plan maintenance measures according to the actual condition of the plants and machines as well as the reduction of associated costs and downtimes.

Framework conditions

- **Definition of use case:** Objectives, metrics (e.g., scrap rate, accuracy, ROI), applicability, scope, requirements, team and departments.
- **Specifications of the plant manufacturer:** Consideration of manufacturer specifications, data interfaces and possibly involving manufacturers regarding existing models and data.
- **Expert knowledge and training:** Involvement of experts to assess data and forecast quality and to check assumptions, plausibility of model results and derived insights. Training of the relevant stakeholders with regard to the development, use and maintenance of the solution.
- **Data availability and quality:** Completeness, correctness, consistency, timeliness and relevance of the data (e.g. machine parameters and sensor measurements). The data must be available in sufficient quantity for modeling.
- **Automated collection of measurement data:** Successful anomaly and failure prediction is based on the correct collection of sensor data, which enables precise predictions and appropriate measures.
- **Data integration:** Integration and aggregation of data from different sources for analysis and modeling.

- **Standardization and data management:** Uniformity of formats for collecting, exchanging and managing data. Continuous maintenance of data to maintain and/or improve data quality.
- **Threshold logic:** Combination of AI prediction with fixed thresholds and understandable reasoning by the AI (e.g. "Vibration > threshold").
- **Supervision and review:** Mandatory review within defined timeframes for safety-relevant applications.
- **Validation:** Implementation of pilot projects to test the effectiveness of the solutions in controlled environments.
- **Governance and data protection:** Definition of responsibilities and compliance with existing standards and regulations.
- **Infrastructure and technical interfaces:** Suitable hardware and software for solution development and ensuring successful technical integration into existing systems.

Added value

- **Improving product and process quality:** Predictive initiation of improvement measures based on data.
- **Improved machine and plant behavior:** Possible insights into correlations between machine parameters and service life.
- **Resource efficiency:** Increased machine and plant availability, maximization of service life and machine performance.

Challenges

- **Data quality:** Process changes influence data quality, e.g. calibration and sensor wear.
- **Unbalanced datasets:** Datasets may be biased in terms of the balance between states to be classified or predicted, e.g. defects, which could affect the uncertainty of the models.
- **Product and process complexity:** Complex, non-linear relationships between machine parameters and machine conditions.
- **Traceability and acceptance:** Low acceptance by specialist

<ul style="list-style-type: none"> • Cost reduction: Reduction of downtimes and faults and optimization of maintenance work. 	<p>departments due to lack of model explainability (black box effect).</p> <ul style="list-style-type: none"> • Technical integration and interfaces: Integration into existing IT and production systems (e.g. MES).
<p>Procedure</p> <ul style="list-style-type: none"> • Definition of use case: Objectives, metrics (e.g. false alarm rate, early warning time, ROI), applicability, scope, requirements, team and departments. • As-is assessment: Analysis of the affected plants, machines and processes as well as initial analysis and evaluation of existing data. • Data preparation and modeling: Preparation and analysis of the data, adapting or expanding data requirements if necessary and collecting corresponding data, and subsequent data modeling and training of the predictive model in a test environment. • Testing and validation: Implementation of the solution in production and derivation of improvement measures. Iterative improvement of the trained model with comparison of the defined metrics. 	
<p>QM roles</p> <p>Employees in production quality</p>	

6.6.1 Example

An interdisciplinary team (e.g. production management, production quality, process development, technology experts, IT) works to minimize plant downtimes for repair and maintenance, e.g. for assembly robots (body construction or final vehicle assembly) or milling processes (e.g. transmission parts or engine components). By predicting critical load and performance points (anomaly detection), downtimes are to be reduced, maintenance

work optimized and ultimately the associated costs lowered. Examples include the prediction of failures or position deviations over time in the event of joint wear (assembly robots) and premature wear or tool breakage (milling process).

For these applications, the potential is first evaluated (e.g. recording the number of unplanned failures, downtimes and associated costs). The potential is determined on the basis of the plant manufacturer's technical specifications and the targets and requirements are defined. Data from quality tests (e.g. tolerance measurements and deviations), environmental data (e.g. vibrations, temperature) and relevant controllable process parameters of the plant (e.g. torques or speeds, depending on the technology) as well as technical specifications from the manufacturer are required.

The required data is collected from quality tests, sensor data and operating data and can be stored separately for the development of test applications. During data preparation, initial analyses of data distributions and outliers are carried out. The quality of the data is evaluated (e.g. completeness, interpretability, relevance). With the help of experts, the data is cleaned and prepared for modeling (depending on the data format, this can be a time-consuming step).

The data is labeled (e.g. values in time series outside defined operating limits), threshold values and metrics (e.g. prediction intervals, deviation metrics) are defined for the AI models and, if necessary, additional features are developed from the existing data in order to better map the existing process and goals of the use cases. Different algorithms are selected (e.g. neural networks, random forests) to train the models and evaluate the prediction accuracy (uncertainty).

The resulting models are optimized by testing various hyperparameters (model parameters) and tested using additional historical data. Using sensitivity analyses or XAI methods, the influences of various parameters in the model on the prediction results (e.g. parameter limits for a prediction outside the defined threshold values) can be determined and discussed with the team.

After successful iterative validation in a test environment, measures for automated correction of the identified critical process parameters are derived.

This can include, for example, the automatic creation of a maintenance order when defined wear limits are reached or the adjustment of process parameters such as torque or feed speed with a warning signal. Integration into production is piloted in the existing IT architecture in collaboration with the team and continuously monitored for effectiveness and potential improvements.



Figure 6-6: Process chain of AI-powered predictive maintenance

Dealing with changes	Interpretation and evaluation of the AI output
<ul style="list-style-type: none"> • Define guard rails: These guard rails can be application-specific thresholds such as prediction intervals, operating limits or tolerance fields for the statistical variability of the input variables. Exceeding or falling below these thresholds can trigger a change according to change classes A, B or C. 	<ul style="list-style-type: none"> • Monitoring the detection and false alarm rates (checking false positives). • Monitoring of values exceeding or falling below defined guard rails and metrics. • XAI methods can be used for each prediction to indi

- **Define change classes:**
A: retraining, updating the model, B: adjustment of features and/or thresholds; C: adjustment of model architecture or scope of application.
- **Change class A (retraining, updating the model):**
Retraining of the model with new data. After documentation, these changes can be applied directly without a new approval.
- **Change class B (adjustment of features and/or threshold values):**
Addition or modification of features in the model (e.g. humidity). After documentation, testing of the improvement in a test environment and an expert review, these adjustments can be released.
- **Change class C (adjustment of model architecture or scope of the application):** Change of the model class, if not previously tested in development (e.g. from neural networks to random forests) or change of scope (e.g. prediction of other downstream process variables). After documentation,

cate the influence of a specific parameter on the prediction.

- Plausibility and consistency check of the results. Deviations are reviewed by experts to check for possible changes of class A, B or C.
- Mandatory human confirmation for safety-relevant applications.
- Documentation of every maintenance decision.

testing of the new model or application in a test environment and a more extensive expert review, these adjustments can be approved..

- **Documentation:** Versioning of data and models, documenting changes according to change classes: Storage of the tests performed and hyperparameters of the model, documentation of changes to features and thresholds as well as documentation of the algorithms and comparisons used.
- **Data monitoring:** Regularly check whether data scope or data quality has changed significantly (e.g. missing values, new value ranges). Calibration of sensors and monitoring of data quality. If there are any noticeable changes, identify the cause.

6.7 Field data analysis

Data from the use phase and other sources, e.g. warranty claims, customer feedback, telemetry data and data from diagnostic equipment, can be used to develop data models for monitoring product quality in the field. The patterns identified in the data enable early detection of failure trends, analysis of usage profiles and root cause analysis of complaints. In this way, measures can be derived to reduce warranty costs and improve current and future product generations.

Description

Field data analysis is a strategy for monitoring product quality in the usage phase, in which data from real customer operations is continuously evaluated. The data models used for this are based on a combination of structured data (e.g. telemetry, logs from diagnostic devices) and unstructured information (e.g. damage descriptions, customer feedback). With the help of AI-powered methods such as natural language processing (NLP) and anomaly detection, these heterogeneous data sources can be analyzed to identify fault patterns and correlate causes of failure.

Framework conditions

- **Definition of use case:** Objectives, metrics (e.g., accuracy, ROI), applicability, scope, requirements, team and departments.
- **Definition of decision basis:** Use of transparent models for justification, classification into criticality levels using fixed rules and decision by expert committee in case of high criticality.
- **Data availability and quality:** Large, structured data volumes in real-time transmission from diagnostic devices enable the use of data models for proactive analysis. Lower data quality in terms of consistency, timeliness, and completeness, as well as reactive measures for unstructured data, such as customer feedback or error descriptions.
- **Data integration:** Integration and aggregation of heterogeneous data formats (e.g. unstructured customer feedback and structured logs from diagnostic devices) from different sources for root cause analysis.
- **Expert knowledge:** Complex cause-effect relationships in the collected field data require a high degree of expertise in order to derive potential causes of errors, interpret usage behavior or identify product optimization potential.
- **Data protection and compliance:** Meeting legal requirements and maintaining customer trust with regard to the storage and processing of personal data.

<ul style="list-style-type: none"> • Infrastructure and technical interfaces: Taking into account the requirements for real-time transmission and high data processing capacity in the IT infrastructure. 	
<p>Added value</p> <ul style="list-style-type: none"> • Proactive product optimization: Continuous monitoring of trends and patterns enables the initiation of targeted innovations and improvement measures. • Preventive error handling: Taking proactive measures to prevent the occurrence of functional errors. • User behavior analysis: Derive insights into actual user behavior. • Reduction of warranty/service costs: Initiation of preventive measures enables failures to be avoided. • Faster root cause analysis: Analysis of field data helps to more quickly identify the causes of failures. • Increased customer satisfaction: Responding to explicit customer needs. 	<p>Challenges</p> <ul style="list-style-type: none"> • Data quality: Low quality of unstructured field data, e.g. customer feedback • Data integration: Integration of heterogeneous data formats from different sources • Data analysis: Complex interactions require expert knowledge (e.g. differentiating between correlation and causality). • Effort: The benefits of the field data analysis must be greater than the effort required to acquire and analyze the data.
<p>Procedure</p> <ul style="list-style-type: none"> • Definition of use case: Objectives, metrics (e.g., accuracy, ROI), applicability, scope, requirements, team and departments. 	

- **Data acquisition:** Identification of data sources, transmission and storage of the data. Determination of data formats / defining data standards.
- **Data preparation and modeling:** Defining data quality requirements and deriving methods for improving data quality. Deriving a data model for mapping the interdependencies.
- **Data analysis and validation:** Rule-based or AI-powered analysis of field data patterns and derivation of fault patterns, causes of faults or patterns in usage behavior. Validation of the identified patterns.
- **Derivation of measures:** Derivation and introduction of proactive or reactive measures for the prevention/elimination of faults in the product during the use phase. Derivation of potential for product optimization.

QM roles:

Employees in customer quality

6.7.1 Example

A quality team evaluates failures in the field that are reported via various customer portals on a weekly basis (e.g. warranty/goodwill claims, repair shop reports, return notifications). These reports are automatically transferred to a central field database and standardized there.

In practice, however, it is difficult to recognize at an early stage from the large number of reports which product and usage combinations are actually conspicuous and which fluctuations are "normal," for example due to seasonal effects. For example, failures of air conditioning components often occur more frequently in summer than in winter, without there necessarily being a new quality defect.

Instead of setting up each evaluation manually, the team uses an AI-powered analysis system that starts once a week as a batch run. The system automatically forms all relevant combinations such as customer, product, vehicle model and production plant and generates a time series for each combination, e.g. "Number of field failures per calendar week."

A forecast and expectation model is created for each of these time series, which takes into account both the long-term trend and recurring seasonal patterns. This allows the system to distinguish whether a current increase is expected (e.g. summer increase in air conditioning) or whether it exceeds the expected level. It also checks the last few weeks against the calculated expected range and flags outliers.

For example, the system can recognize that the combination "Customer A – air conditioning compressor variant 3 – platform X – plant 2" has been significantly above the expected trend for several weeks, even though seasonality has already been taken into account. At the same time, it shows that although other combinations also increase in summer, they remain within the expected seasonal framework. The system calculates a deviation value for the conspicuous combination and places it at the top of a weekly ranking.

The quality team uses this ranking list as a worklist. For the top cases, an expert plausibility check is first carried out (e.g. data completeness from the portals, reporting delays, duplicates, changed error codes). Subsequently, specific in-depth investigations are triggered, for example by production period, batch, software version or supplier. This can result in specific measures, such as a restriction to certain production weeks, an adjustment of the inspection characteristics, a root cause analysis with the plant or a targeted returns investigation.

The end result is a field data analysis that runs regularly and reproducibly, methodically takes seasonal effects into account and draws the team's attention to the combinations where an increased risk of failure is most likely. The evaluation of and decision on measures to be taken remains entirely with the specialist team. The system provides assistance through structured data preparation, trend identification and prioritization.

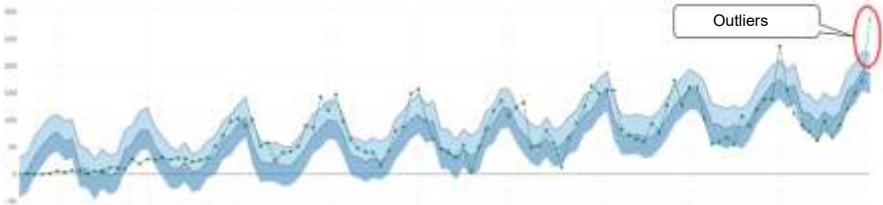


Figure 6-7: Seasonal outlier detection in field data analysis – failure time series of an air conditioning system with AI-generated expectation range and marked outlier

Dealing with changes	Interpretation and evaluation of the AI output
<ul style="list-style-type: none"> • Pre-evaluate and version changes to data sources. • Data monitoring: Regularly check whether the volume or quality of the data has changed significantly (e.g. missing values, new value ranges). If there are any noticeable changes, identify the cause. • Record label changes (e.g. new error codes, changed workshop coding); document re-labeling rules. • Ensure traceability: Each analysis must be reproducible (data version, filter, model version, threshold values). 	<ul style="list-style-type: none"> • Plausibility and consistency check: Are trends/anomalies consistent with known events? • Consider the result as an indication: AI results are an indication, not proof. Consult additional information before making decisions (e.g. repair shop reports, parts inspection, complaints). • Cross-check to be sure: Check random samples and compare with known cases. If possible: Compare with a simple evaluation (e.g. frequencies/trends) for validation. • Use clear decision categories, e.g. "observe," "investigate further," "act immediately." The decision is made by the responsible specialist department, not the AI.

6.8 Review of development work products

Description <p>During product development, quality engineers carry out a large number of reviews of work products against quality criteria, work instructions, guidelines or other applicable documents. This task is supported by an AI system and deviations are flagged. The AI can also check only certain aspects if there are other aspects that cannot be checked closely enough by the AI.</p>	
Framework conditions <ul style="list-style-type: none">• Quality criteria, work instructions, guidelines and other applicable documents for development work products are available in machine-readable formats• Quality criteria, work instructions, guidelines and other applicable documents are stored in the AI system to ensure a well-founded assessment of deviations.• Large documents are converted into stable sections/clauses through automated chunking and parsing.• Hybrid approach: Deterministic text segmentation combined with AI content analysis.• Rule-based assessment: The work product is classified as ready for release according to clear criteria based on the frequency and severity of the deviations found	
Added value <ul style="list-style-type: none">• Deviations are identified and classified if necessary• Efficiency and time gains through AI-powered review.• All deviations found appear in a clearly structured overview to serve as a basis for lessons learned.	Challenges <ul style="list-style-type: none">• Reliable detection of contextual deviations is challenging; misclassifications are possible.• Unstructured texts, scans, tables and inconsistent formatting impair segmentation and thus the quality of comparison.

	<ul style="list-style-type: none"> • Very large documents put a strain on semantic analysis; chunking improves the results. • Human validation/checking necessary.
<p>Procedure</p> <ul style="list-style-type: none"> • Define comparison documents (e.g. guidelines, other applicable documents). • Structured preprocessing: Parsing, classification, structuring into sections/clauses. • Generation of a structured deviation overview with evaluation and recommendations. • Review of the result by the specialist department/Quality. 	
<p>QM roles</p> <p>Employee in development quality, quality auditor, quality assessor</p>	

6.8.1 Example

Review of software requirements during the development of an control unit against quality criteria.

A complex control unit has a large number of requirements for the software. These must be checked against quality criteria (e.g. traceability, formulation, consistency, testability).

An AI system assists with this task and reports deviations from the quality criteria defined in the project.

In advance, the requirements documents must be broken down into chunks that can be handled by the AI, but which still reflect the relationships necessary for checking consistency, for example.

6.8.2 Example

Review of the implementation of control unit software against errata (list of known errors) in supplier software.

Modern control unit software integrates many components from suppliers (e.g. AUTOSAR stack) and uses supplier-provided tools (e.g. compilers). The manufacturers of these components regularly publish extensive collections of errata against which the implementation of the control unit software must be checked.

For this purpose, the corresponding errata are preprocessed and integrated into the AI system.

The fully integrated software with its relevant sources is now checked against these implemented errata and returns whether they were taken into account during implementation. This must also be broken down into chunks that can be processed by the AI but that still provide the required context.

Dealing with changes	Interpretation and evaluation of the AI output
<ul style="list-style-type: none">• Define change classes: Subdivide adjustments to the AI system into classes (e.g. A: purely linguistic/cosmetic, B: functionally relevant but within the same framework, C: fundamental change in behavior). Class C triggers a new approval process.• Minor changes (class A): New or amended quality criteria, work instructions, guidelines and other applicable documents can be incorporated if they do not	<ul style="list-style-type: none">• Measurement of the compliance rate between AI suggestion and expert review.• Versioning of the comparison reports with release note.• Use of reference documents with known changes for validation.• Ensure traceability of AI suggestions: For each reported deviation, the AI provides a justification

deviate significantly from the original in terms of scope and format. After a dual-control review and documentation, these adjustments can be activated without a new approval.

- Targeted testing of **functionally relevant changes (class B)**: Additions that more precisely specify the behavior but do not fundamentally change the role of the AI – e.g. new quality criteria categories, refined chunking rules, additional reference documents – are tested with a representative set of test requirement documents. Experts assess whether classification and deviation reports are still accurate and consistent.
- **Treat fundamental changes (class C)** as a new system version: In the case of new document formats (e.g. new structuring conventions, previously unsupported file formats), a full regression test must be carried out. Adjustments that change the risk profile of the system (e.g. automatic release of work products by the AI) require a

with reference to the specific quality criterion (e.g. "Requirement not testable – missing measurement or acceptance condition"). This makes it clear on which rule the deviation is based.

- **Measurement of compliance rate**: The agreement between the AI suggestion and the result of the manual expert review is measured and documented. Deviations between the AI assessment and the human assessment are analyzed and used to continuously improve the system.
- **Versioning of the comparison reports with release note**: Each deviation report generated is saved with a time stamp, the system version used and a release note from the quality engineer responsible. This means that the version of the quality criteria and the AI model used for the assessment can be traced at any time.

complete reassessment and approval.

- **Define guard rails:** Typical guardrails include: The AI only makes suggestions, not final release decisions; each reported deviation contains a reference to the violated quality criterion; the work product is released exclusively by the responsible quality engineer. If these guard rails remain unchanged, the application does not require a new approval.

- **Validation using reference documents:** Reference requirement documents with known, predefined deviations are used for quality assurance of the AI system. The AI outputs are checked against these expected values (e.g. number of detected deviations, false positives, unrecognized deviations).
- **Human review and re-release is mandatory:** The quality engineer sees all AI suggestions including classification and criteria reference. He decides which reported deviations to confirm, adjust or reject as false positives. The final classification of the work product as releasable is the sole responsibility of the human.

6.8.3 Notes

The use case described can also be applied to other work products.

6.9 VDA chatbot

Description

AI-powered chatbot for quick, qualified answers to questions relating to VDA regulations, e.g. methods such as 8D, FMEA and audits, inspection processes and, optionally, a company FAQ. The bot provides verified,

qualified answers with references (volume/chapter/section) and, where appropriate, step-by-step instructions. Knowledge is provided via RAG; free internet searches are prevented.

Framework conditions:

- Knowledge database available (relevant VDA volumes + company FAQ + related VDA/company standards.
- Creation of a question catalog for the VDA chatbot to improve answer quality
- Integration of user feedback (e.g. thumbs up/down) for continuous quality improvement.
- Definition of responsibilities, target groups and multi-level access.
- Change management for content and data quality.
- Option to minimize risk: Switch to rule-based responses (bot may only display verbatim excerpts and verified summaries from approved VDA sections; no free formulation beyond that)
- Citation requirement (RAG): Answers may only be generated if a valid source was found in the VDA volume.

Added value

Fast, consistent access to VDA/QM knowledge; reduced search and waiting times; uniform, verified answers across all locations.

Challenges

Ensuring data/content quality; validation so that only approved content is output.

Procedure

- Set up a chatbot that uses relevant VDA/QM documents as a RAG knowledge base. Prevent internet searches.
- Create a master prompt to specify how the chatbot should behave when responding (see chapter "Recommended actions" XX)
- Check answer quality via question catalog; evaluation by VDA/QM experts.

- Expand user group depending on response quality → scaling

Roles

Employees in supplier quality, employees in development quality, employees in production quality, employees in customer quality, employees in the quality management system, quality auditor/assessor, quality manager

6.9.1 Example

User logs into a web frontend and asks a question (e.g. "Which requirements apply for the release of production processes and products according to VDA 2?"). The backend uses a semantic vector database to search for the appropriate text passages from the VDA volume (retrieval augmented generation). These passages are transferred together with the question to the AI module, which generates an answer and adds a source reference with chapter/section number.

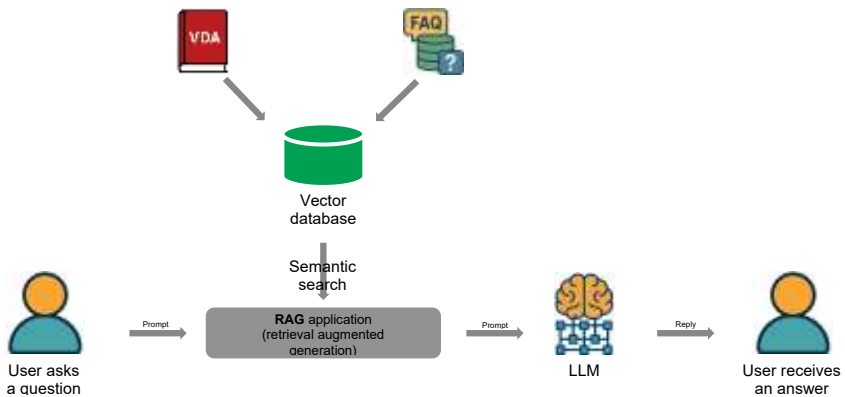


Figure 6-8: VDA chatbot – RAG architecture with semantic vector database search and LLM-supported answer generation

Example of a web frontend for the VDA chatbot. Sterilized web chatbot mask with login header, central question-answer area with mandatory source reference as well as feedback functions and quick access to specific topics.

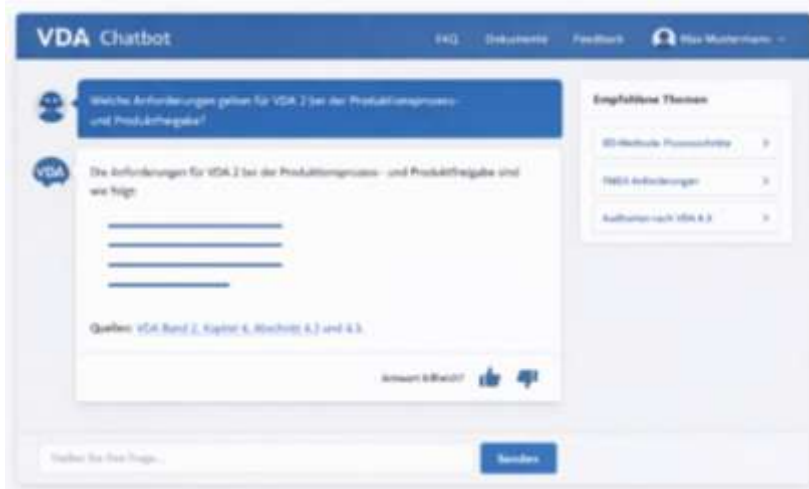


Figure 6-9: Example chatbot interface with AI-generated answer and source reference

Dealing with changes	Interpretation and evaluation of the AI output
<ul style="list-style-type: none"> • The configuration (permitted sources, answer rules, citation format) is maintained on an ongoing basis. • Changes undergo a functional and technical dual-control review, including testing, before being rolled out. 	<ul style="list-style-type: none"> • Measurement of the abandonment rate (responses without a source). • Logging of every interaction (question, answer, source ID, rule version). • Regular comparison against a defined list of questions with known correct answers ("golden set").

- User feedback is evaluated on a weekly basis; critical errors are rectified immediately as hotfixes.

6.9.2 Notes

A comparable system can also implement the enhancements proposed here. For example, the same VDA chatbot architecture could be used to implement standards and compliance content, a QM FAQ bot for suppliers or lessons learned. Thanks to the modular structure of the knowledge database and the use of retrieval augmented generation (RAG), the concept can be flexibly transferred to other quality areas.

6.10 Speech mining for work instructions

Description

Employees use their voice to describe the steps they are doing as they are actually doing them. The audio data is transcribed using speech-to-text and then structured via NLP (natural language processing) into activities, preconditions, material/tools and safety and quality features. On this basis, the AI generates a standardized work instruction that can be transferred to the management system and released.

Framework conditions

- Clearly define the start scenario and process boundaries; clarify critical steps in advance
- Define roles: Process owners/subject-area experts for functional reviews.
- Hardware: Smartphone or suitable recording device; if required, headset or hands-free solution to avoid distraction during safety-critical activities.

- Data protection and transparency: Consent to voice recording, purpose limitation, deletion concept, access controls; observe management system specifications.
- Visual supplements (photos/frames) and clear step-by-step lists make the work instructions easier to understand and follow.
- Automatic rejection if security or test features are missing.
- Visual confirmation: Addition of photos/videos to validate the AI text.

Added value

- Accelerated, practice-oriented documentation of the process and standardization based directly on the execution.
- Prompt updating of instructions in the event of process changes; fewer subsequent interviews/corrections.
- Reduction of subsequent corrections
- Collaborative creation: Multiple people can provide input, which is then consolidated.
- Multiple people can be involved in the recording.

Challenges

- Dialects or individual ways of speaking.
- Specialist terminology and consistency.
- Acoustics/recording situation (e.g. production noise, overlapping speakers).
- Users must learn/accept "voice work." Details such as parameters or test criteria must not be forgotten.
- Data protection: Consent, purpose limitation, deletion concept and access controls.

Procedure

- Clarify data protection: Define legal basis, consent, storage and deletion rules; inform data subjects.
- Define scope: Specify concrete process and target format (template of the work instruction).
- Provide setup: Recording device/smartphone; if required, hands-free hardware and noise-capable microphone for each speaker.

- Describe clearly during the activity: What? With what? Why? Inspection criteria? Safety?
- Transcription and structuring: Speech-to-text with time stamps.
- Define prompts/extraction rules: Define specifications for NLP/generation
- Switch off internet search so that entered data is not influenced by external information.
- Subsequently, generated text must be checked by an expert.
- Copy text into template and format accordingly.
- Copy text into the work instruction template, add visual elements.

Roles

Employees in supplier quality, employees in development quality, employees in production quality, employees in the quality management system

6.10.1 Example

When assembling a ballpoint pen, an experienced employee describes his work process orally as he does the activity. The recording is automatically transcribed and then analyzed using NLP (language processing). The AI recognizes relevant activity blocks, components used (e.g. refill, barrel parts, spring) and typical quality features (e.g. clean running of the mechanism) and structures this information according to the company-wide template for work instructions. In addition, the AI can automatically generate a graphical process flow from the extracted activities, which clearly illustrates the assembly process. A complete, standardized work instruction is thus created from the spoken description, which is then checked for accuracy by the process owners and transferred to the management system.



Figure 6-10: Example AI-generated process flow for work instructions – ballpoint pen assembly, derived from spoken employee description

Dealing with changes	Interpretation and evaluation of the AI output
<ul style="list-style-type: none"> • Templates and extraction rules are versioned. • Small adjustments can be activated immediately after a dual-control review. • Data protection (consent, storage and deletion periods) remains up to date. • The recording setup is standardized. 	<ul style="list-style-type: none"> • Completeness check of mandatory fields. • Release by an expert is mandatory (human-in-the-loop). • Logging of versioning.

6.10.2 Note

The approach described can be transferred to any manual assembly process and scales well in environments with frequent variant changes or short product life cycles.

The quality of the results depends heavily on acoustics, speech discipline and consistency of the specialist language. Expert validation remains mandatory.

6.11 Comparing documents

<p>Description</p> <p>Customer specifications, contracts or standards are compared using a hybrid architecture consisting of deterministic components and AI-powered analysis processes. The documents are first preprocessed in a structured manner (parsing, section recognition, clause mapping). AI then evaluates the differences in content semantically and assigns them to meaning categories. No changes to words or characters are displayed; instead, contextualized change objects are created with a brief description, relevance, risk and impact assessment and recommendations.</p>	
<p>Framework conditions</p> <ul style="list-style-type: none"> • Documents are available in machine-readable formats • Company guidelines are stored in the AI application to provide a sound basis for assessing deviations and making recommendations. • A fixed preprocessing logic ensures that segmentation, mapping and sequence remain reproducible – pure AI agents are not sufficient for this. • Large documents are converted into stable sections/clauses through automated chunking and parsing. • Hybrid approach: Deterministic text segmentation combined with AI content analysis. • Rule-based assessment: Risk and obligation are classified using fixed keywords (e.g. "must", "must not"). 	
<p>Added value</p> <ul style="list-style-type: none"> • Changes are presented according to significance and 	<p>Challenges</p>

<p>context; this speeds up the expert review.</p> <ul style="list-style-type: none"> Automated risk and impact assessment per change with clear recommendations. Efficiency and time gains through AI-assisted comparison and structured processing of results. All relevant changes appear in a clearly structured overview. 	<ul style="list-style-type: none"> Reliably recognizing contextual changes in meaning is challenging; misclassifications are possible. Unstructured texts, scans, tables and inconsistent formatting impair segmentation and thus the quality of comparison. Very large documents put a strain on semantic analysis; chunking improves the results. Observe the maximum number of tokens. Human validation/checking necessary.
--	--

Procedure

- Define comparison documents.
- Structured preprocessing: Parsing, classification, structuring into sections/clauses.
- Semantic analysis of changes using embeddings and similarity metrics.
- Comparison with standards and guidelines; classification according to risk types/levels and description of the effects.
- Creation of a structured change overview with evaluation and recommendations.
- Review of the document comparison by the specialist department.

QM roles

Employees in supplier quality, employees in development quality, employees in production quality, employees in customer quality, employees in the quality management system, quality auditors, quality assessors

6.11.1 Example

The process describes how AI is used to compare the content of documents (e.g. automated comparison of two versions of a customer specification; detection of changes in meaning and suggestion of risk categories).

Semantic segmentation avoids the lost-in-the-middle problem – an effect in which large language models process information from the middle of long texts more poorly or overlook it.

This means that the context of the content is fully preserved and changes can be evaluated precisely and traceably.

Process steps:

1. Upload documents – provision and versioning of the documents to be compared.
2. Semantic segmentation – dividing the documents into logically related sections.
3. Semantic comparison – analysis of content differences based on embeddings and similarity metrics.
4. Interpretation & evaluation – generative summary and evaluation of the changes according to relevance, reference to standards and risk.
5. Consolidation & report – consolidation of all results into a traceable report that can be released.

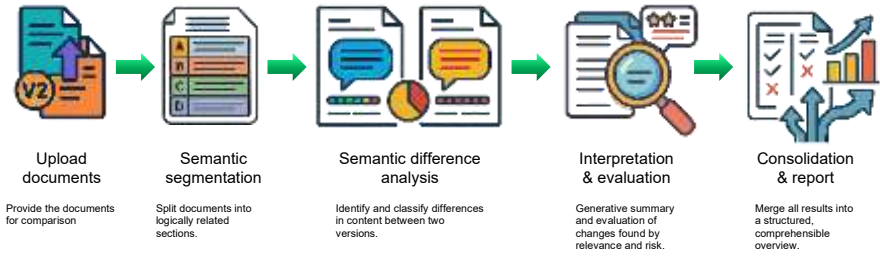


Figure 6-11: AI-powered document comparison – semantic segmentation and difference analysis enable precise detection and risk assessment of content changes between two document versions

Segment-based context window management helps to avoid lost-in-the-middle errors.

Example display:

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-KUM-02 angepasst werden muss
6.3	Planung von Änderungen	Abschnitt erweitert um Bewertung der Auswirkung auf Lieferanten	Bedeutungsänderung	Mittel	Rücksprache mit Einkauf empfohlen
8.2.1	Lenkung fehlerhafter Produkte	Absatz um KI-gestützte Prüfprozesse ergänzt	Neue Technologiebezug	Hoch	Bewertung durch Fachbereich QS erforderlich

Dealing with changes	Interpretation and evaluation of the AI output
<ul style="list-style-type: none"> Rules for mandatory and optional clauses and risk categories are maintained as a configuration. Regression tests are carried out for new document formats; mapping is adapted. 	<ul style="list-style-type: none"> Measurement of the compliance rate between AI suggestion and expert review. The compliance rates are to be defined by the respective company.

- New standards undergo an expert review; configuration changes can then be made quickly.

- Versioning of the comparison reports with release note.
- Use of reference documents with known changes for validation.

6.11.2 Notes

The use case described can also be used for the automated analysis and comparison of standards or process documents. NLP-supported processes can be used to automatically identify new or amended normative requirements and compare their content with existing QMS processes.

This results in a structured gap analysis that presents changes in a traceable manner and derives recommendations for necessary adjustments to the process.

Technical note

The comparison is not based on pure AI agents. Stability, reproducibility and auditability are only achieved through a hybrid pipeline of deterministic logic and AI-powered evaluation, similar to established industrial solutions that use a combination of structured parsing, embedding procedures and a downstream evaluation step.

6.12 Interactive learning

Description

Use of an AI-powered, virtual "QM teacher" that explains complex quality management topics to employees simply, interactively and at any time. The AI answers questions about standards, processes, test methods, failure patterns or QA tools in natural language and adapts the explanations to the user's level of knowledge. The virtual teacher is used both for quick questions in day-to-day work and for structured training (e.g. VDA standards, test processes, 8D, machine capabilities, test equipment capability, etc.)

Framework conditions

Complex requirements, standards and process standards necessitate that the knowledge conveyed is very accurate. Employees in production, QM, engineering and maintenance have very different levels of prior knowledge, but need quick access to reliable information on a day-to-day basis. QM documentation is extensive, frequently updated and difficult to access for many users, which leads to gaps in knowledge and recurring queries. At the same time, requirements are increasing due to internal audits, VDA standards and customer-specific guidelines, making consistent training increasingly important. Many companies work in multiple languages, which requires multilingual knowledge transfer.

Against this background, the virtual AI teacher is designed to provide accurate and approved information at all times. Data protection and IT security requirements must be taken into account during implementation, as QM data is often sensitive. At the same time, QM experts must be involved in order to release and continuously maintain content. Overall, these framework conditions are the result of the struggle between stringent training requirements, limited resources, regulatory requirements and the desire to make knowledge quickly available on the shop floor.

Added value

- **Faster clarification of QM questions on a day-to-day basis** without having to consult experts or spend a long time searching through documents.
- **Uniform understanding** of standards, processes and methods across locations.
- **Better decisions** thanks to consistent, verified answers.
- **Efficient introduction of new employees** to QM processes,

Challenges

- **Up-to-date QM content** in the event of changes to standards (VDA, IATF, internal specifications).
- **Validation of responses:** The AI may only output approved QM information.
- **Complexity of specialist language** and site-specific terms must be correctly understood.
- **Confidentiality of quality data**, audit findings or production problems.

inspection plans, standards and tools.

- **Fewer repetitive questions** to QM experts and therefore more focus on value-adding tasks.
- **Interactive explanations of failure patterns**, test methods or root cause methods (e.g. 5Why, Ishikawa).
- **Rapid updating** in the event of changes to VDA regulations, internal specifications or audit requirements.

- **Acceptance by auditors and QM teams** who expect exact and regulation-compliant formulations.
- **Integration with existing knowledge material** (process descriptions, inspection plans, lessons learned).

Procedure

- **Define scope** → Narrow down topics: e.g. 8D, test equipment capability, machine capability, VDA 6.3; define error catalogs and target groups: Production, QM, Maintenance, Engineering.
- **Build knowledge base** → Gather relevant QM documents (VDA standards, internal specifications, inspection plans, FAQs) and structure, version and validate content for the AI.
- **Select technical platform** → Decision: on-premises AI, cloud AI with secure company access or integration into the company learning platform.
- **Develop a pilot** → Start with a clear QM area (e.g. "Virtual teacher for test equipment capability"). Define typical user questions (from audits, shop floor, engineering).
- **Test & validate** → QM experts check the quality of the answers. Improve content iteratively and fill any gaps.
- **Rollout & training** → Introduction on the shop floor and in QM. Short tutorials, live demos, "Ask the virtual QM teacher" sessions.

- **Operation & continuous improvement** → Define responsibilities (content owner, QM experts). Monitor usage, analyze frequent questions, continuously update content.

QM roles:

Employees in supplier quality, employees in development quality, employees in production quality, employees in customer quality, employees in the quality management system, quality auditors, quality assessors, quality managers AI Q-Data Engineer, AI Q-Data Analyst, AI Q-Data Scientist, AI Q-Data Manager

7 Excursion: Risk-based assessment of AI development tools

This section presents a methodology that enables a risk-based assessment of AI development tools. AI development tools are tools that are used in the AI life cycle to develop AI applications, AI components of applications or complete AI systems. A risk-based approach is recommended for the operational assessment and, where necessary, qualification of AI development tools. The aim is to be able to manage the assessment and safeguarding effort required in proportion to the risk. The approach is based on established practices in safety domains (including ISO 26262, DO-330/DO-178C) and uses their core variables tool impact (TI)⁸ and tool error detection (TD)⁹, while also addressing other regulatory areas such as data protection and AI regulation (EU AI Act) with its requirements for transparency, traceability and human monitoring.

The methodology consists of two analysis processes that can be carried out independently of each other.

- Analysis of the process risk for selected development tasks in the AI development life cycle.
- Assessment of one or more development tools in the context of its task in the development life cycle, taking into account the previously determined process risks.

The separation into a process risk analysis and a tool-specific assessment increases the flexibility of the approach. Thanks to this modular structure,

⁸ The tool impact describes the influence of the tool on the safety of the product and can be divided into TI-0 (tool has no influence on safety) and TI-1 (tool can jeopardize the safety of the product).

⁹ Tool error detection describes the probability of detecting an error in the tool. Tool error detection can be defined in three levels: TD-0 (high probability of detection), TD-1 (a tool error is likely detected) and TD-2 (a tool error is likely not detected).

process risks can initially be identified and assessed independently of specific development tools. This creates a stable, referenceable risk basis along the AI development life cycle.

The assessment of individual development tools is then carried out contextually based on the previously determined process risks. Changes in the tool chain, for example due to the replacing, updating or adding of new tools, therefore do not require a complete reassessment of the overall process, but merely a new tool-specific analysis in the respective application context.

In addition, the decoupling of the two analysis steps allows the development of standardized templates or reference models for typical development processes. Such predefined process risk profiles can be reused across the organization and adapted to specific projects, which supports both efficiency gains and a consistent basis for assessment.

7.1 Explanation of the basic concepts

The assessment process described here is based on a formalized description of the AI life cycle, the development tasks that take place within it, the tools used and the associated risks. The aim is to systematically identify risks, to make the context in which they arise transparent, and to describe the significance of tools as a risk factor in a structured manner.

The necessary basis for this is a consistent information model that explicitly describes the relevant concepts and their dependencies. This enables the traceable and reusable documentation of factors that influence risk, their cause chains and possible countermeasures. The model supports both the process risk analysis and the downstream tool-specific assessment. The model is shown in Figure 7-1.

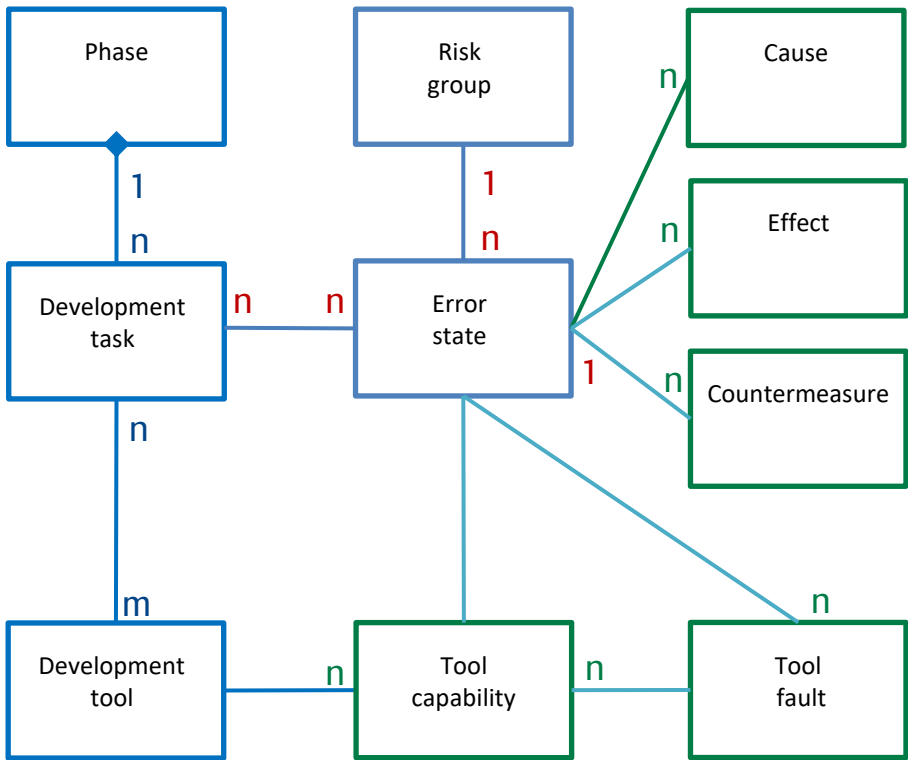


Figure 7-1: Information model

The central concepts of the model shown in Figure 7-1 are explained below.

- **Phases:** The AI life cycle is divided into structured phases (see also process phases in section 5.2.1). Each phase represents a delimited context in which specific regulatory, safety-related and data protection requirements come into effect. The phase structure serves as a framework for allocating development tasks, identifying typical error states and systematically localizing risks.
- **Development tasks:** Specific development tasks are defined within each phase. Examples include tasks such as "*data preprocessing and transformation*," "*feature engineering and extraction*" in the "*data preparation*" phase or "*experiment tracking & management*"

in the "*AI modeling*" phase. The development tasks are the central point of reference for the risk analysis. For each task, it is examined which errors can occur and which regulatory or safety-relevant requirements would be violated as a result. Development tasks thus represent the link between the abstract process phase and concrete operational activities.

- Error states: Possible error states are identified for each development task. An error state describes a specific undesirable state of a process or system that can occur in the context of a development task and, if it occurs, leads to non-compliance with regulatory, data protection or safety-critical requirements. Examples of error states include:
 - Loss of data integrity, lack of real-time processing, inconsistent model reproducibility or inadequate handling of edge cases, but also
 - Errors in the anonymization and pseudonymization of data, non-transparent management of user consents, violations in cross-border data transfers or non-compliance with bias and fairness requirements.

Each error state is assigned to a risk group. This makes it transparent which regulatory or safety-related dimension is affected.

- Risk groups: For structuring purposes, error states are assigned to thematic risk groups (see section 5.1). This grouping enables systematic compliance with regulatory requirements (e.g., the EU AI Act) and the structured derivation of testing and validation measures.
- Cause, effect and countermeasures: Based on the error states and risk groups, causes (e.g. errors in task processing, inadequate safeguards and controls), effects (e.g. incorrect or missing result arti-

facts) and countermeasures (e.g. procedural measures) are identified for each error state. This creates a traceable cause-and-effect chain that enables targeted risk reduction measures to be taken.

- **Development tools:** Development tasks are typically supported by specific tools, such as ML frameworks, training environments, version management systems or monitoring platforms. Development tools must always be considered in context. They are not assessed in isolation, but rather in relation to the respective development task and the associated risks.
- **Tool capabilities:** For each development task, the tool capabilities used to implement a development task are analyzed. Their existence, maturity and correct configuration are central assessment criteria for tool qualification.
- **Tool errors:** A tool error is a condition in which a tool malfunctions while performing its functions, thereby causing an error. Causes include incorrect implementation, incorrect configuration or insufficient validation. Tool errors can cause the aforementioned error states either directly or, due to reduced tool capability, indirectly.

The central benefit of the information model is that it provides a clearly structured organization and analysis framework that enables the systematic identification and analysis of risks while explicitly taking into account dependencies between development tools, their tasks and errors as well as the development tasks and the associated process risks.

This makes interactions and cause-effect chains transparent and prevents isolated, individual considerations.

7.2 Performing the risk-based assessment of AI development tools

The risk-based assessment of AI development tools follows a structured, two-stage procedure that systematically considers both the process perspective and the tool level. The starting point is the consideration of specific development tasks within defined phases of the AI life cycle. For these

tasks, potential error states, their causes and effects as well as suitable countermeasures are identified and evaluated in terms of their criticality. This results in a traceable risk assessment for the implementation of selected development activities.

The tools used in these tasks are then evaluated on this basis. This involves analyzing the extent to which tool-specific properties or potential errors can contribute to the emergence of identified risks. The focus is therefore on the question of how a tool affects the risk associated with a given development task and what safeguarding or qualification requirements this entails.

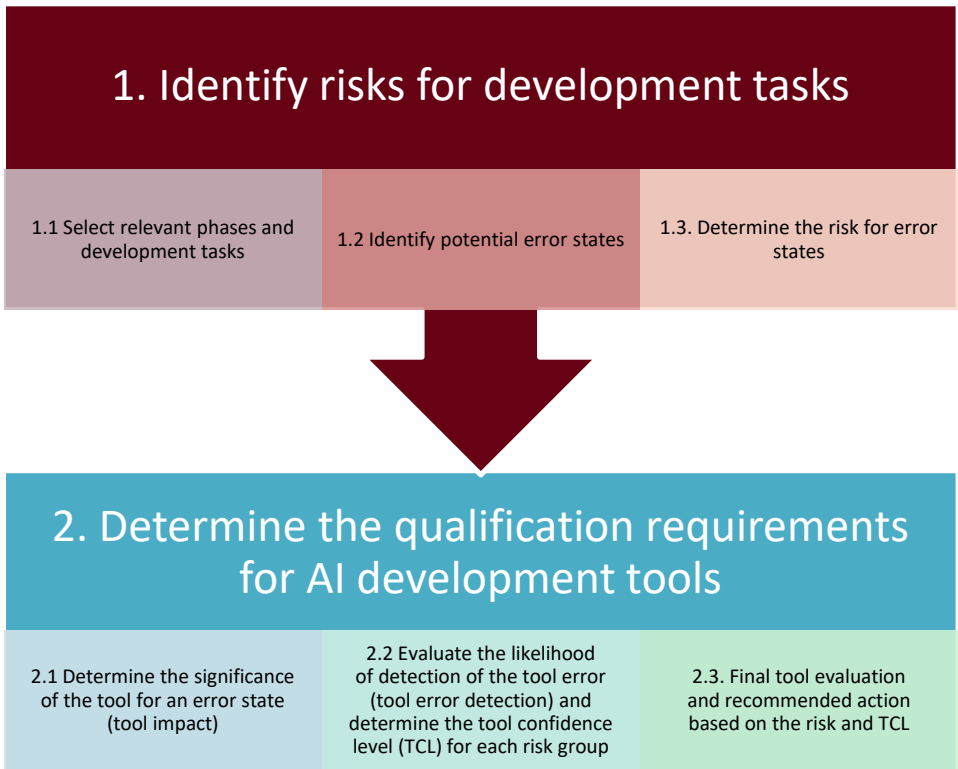


Figure 7-2: Procedure for determining the qualification requirements for AI tools

This combined procedure enables consistent qualification requirements to be derived. Risks are first determined in the subject-area-specific and regulatory context of the task and then considered in relation to the tools used. This supports a transparent, context-specific and methodologically sound assessment of AI development tools along the entire life cycle.

7.2.1 Step 1: Identify risks for selected development tasks

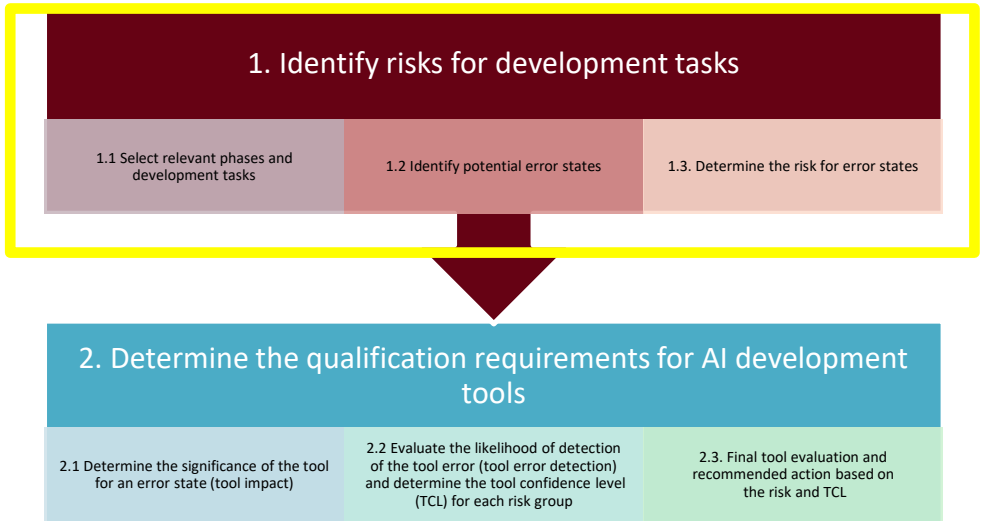


Figure 7-3: Identifying risks for development tasks

The aim of the first step is to systematically identify and evaluate process risks for selected development tasks within defined phases of the AI development life cycle. The focus is on the question of which error states can occur in a specific task, what effects these have and how critical they are to be assessed in a regulatory or safety-relevant context. The identification of risks can be based on established procedures such as failure mode and effects analysis (FMEA). The procedure is carried out in several structured sub-steps.

7.2.1.1 Sub-step 1.1: Select relevant phases and development tasks



Figure 7-4: Selecting relevant phases and development tasks

First, the phases of the AI life cycle to be assessed are determined. Within these phases, the specific development tasks are identified and analyzed with regard to possible risks. This grouping into a defined phase model ensures that the analysis is carried out consistently, completely and in a comparable manner. At the same time, the targeted selection of individual tasks allows for a focused examination of particularly critical process steps.

7.2.1.2 Sub-step 1.2: Identify potential error states

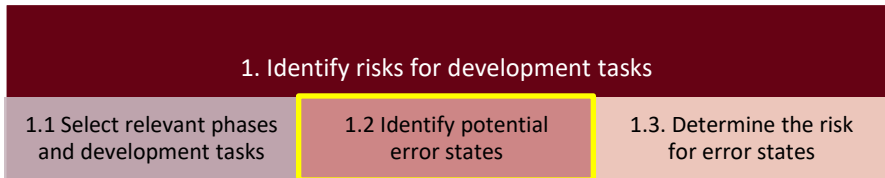


Figure 7-5: Identifying potential error states

In the next step, potential error states are identified for the selected development tasks. These are derived taking into account relevant risk groups and compliance areas (e.g., safety, data protection, transparency, robustness).

Not every risk group is equally relevant in every context. Depending on the application domain, system criticality or regulatory framework, certain error states can be included or excluded. In addition, possible causes, potential effects and possible countermeasures are systematically recorded and documented for each identified error state.

7.2.1.3 Sub-step 1.3: Determine the risk for the error states



Figure 7-6: Determining the risk for error states

The identified error states form the basis for the subsequent risk assessment. As part of an FMEA, for example, the assessment is carried out using the established factors:

- Severity (S) – severity of the impact in the event of a failure (scale 1–10)
- Occurrence (O) – probability of occurrence (scale 1–10)
- Detection (D) – probability of detection before it causes a problem (scale 1–10)

These three factors can then be used to calculate either a risk priority number (RPN) or action priority (AP).

While the $RPN = S \times O \times D$ is determined multiplicatively, the AP is a priority-based classification according to a defined evaluation matrix. In particular, the severity factor is weighted more heavily here so that higher potential harm can lead to a high priority regardless of the low probability of occurrence.

The result of this step is a structured, evaluated risk list for the development tasks under consideration. This forms the subject-area-specific and regulatory basis for the subsequent tool-specific evaluation in the second step.

7.2.2 Step 2: Tool evaluation and determination of the qualification requirements

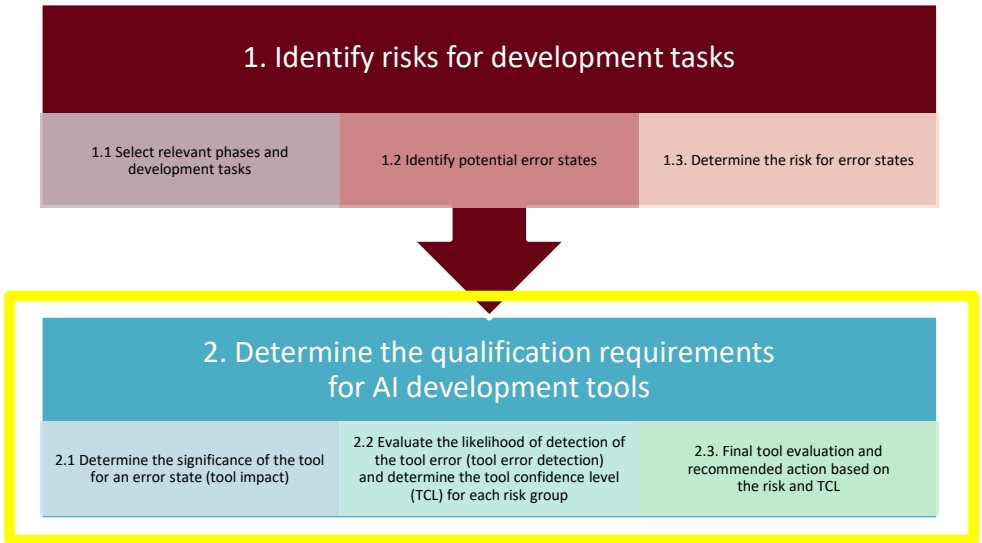


Figure 7-7: Determining the qualification requirements for AI development tools

Building on the previously described process, the second step involves evaluating the tools used. The aim is to systematically record and evaluate the potential influence of the development tools on the occurrence or avoidance of identified error states. The methodological framework is based on the principles of tool evaluation from ISO 26262, which serves as a guide for the structured evaluation and preparation of a potential tool qualification.

7.2.2.1 Sub-step 2.1: Determine the significance of the tool for an error state (tool impact)

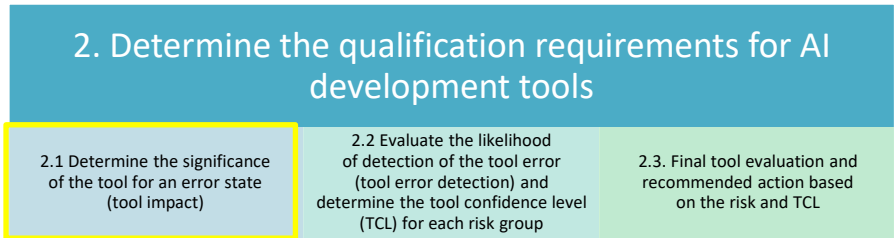


Figure 7-8: Determining the impact of the tool on an error state

In the first step, the impact that a tool error could have on the occurrence or amplification of a process error is estimated for each relevant error state. This assessment is referred to as the tool impact (TI). The tool impact is high if an error in the tool can lead directly or indirectly to safety-critical, data protection-relevant or regulatory non-conformities. A low tool impact means that the impact of the tool on the error state under consideration is marginal or easily compensated for.

7.2.2.2 Sub-step 2.2: Evaluate the likelihood of detection of the tool error (tool error detection) and determine the tool confidence level (TCL) for each risk group

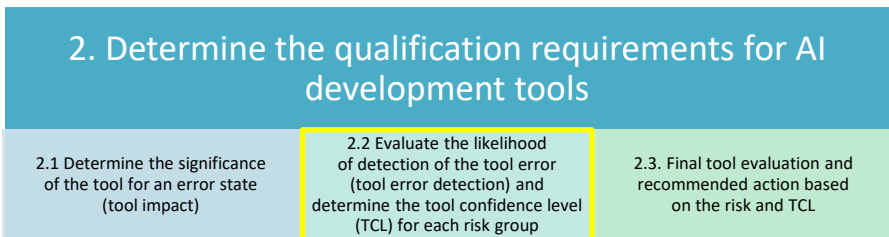


Figure 7-9: Evaluating the likelihood of detection of the tool error and determining the TCL

An assessment is then made of how likely it is that an error caused by the tool will be detected before it leads to incorrect behavior or non-conformity. This assessment is referred to as tool error detection (TD).

A high TD value indicates that tool errors are difficult to detect and additional inspection mechanisms or process measures may be necessary. A low TD value, on the other hand, indicates that errors in the tool can typically be detected quickly or identified by downstream checks.

The combination of TI and TD results in a qualitative confidence measure known as the tool confidence level (TCL). This measure is not used for formal classification in the sense of ISO 26262, but rather as a guide for assessing the required confidence in a tool:

1. TCL 1: Tool can be used without taking special measures; risks are low or adequately controlled by processes.
2. TCL 2 – 3: Higher confidence in the tool is required; additional checks or accompanying measures are recommended.

TI and TD values can be assigned for each combination of task, error state and tool. The resulting TCL is calculated from the values determined and linked to the previously determined risk priority number (RPN) from the FMEA.

7.2.2.3 Sub-step 2.3: Final tool evaluation and recommended action based on the risk and TCL

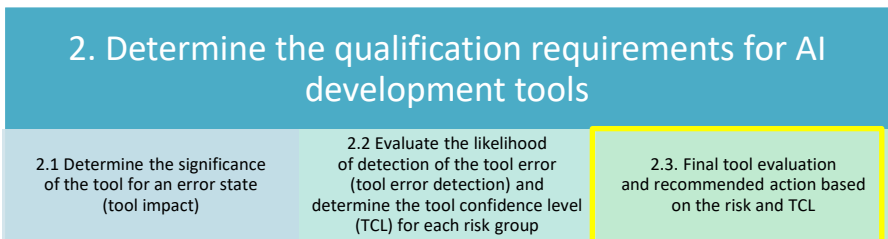


Figure 7-10: Final tool evaluation and recommended action

The combination of the process-related risk and the TCL (tool confidence level) creates a prioritized overview of the most critical tools in the respective development context.

The two assessment values can be filtered using thresholds so that only those combinations are taken into account that have both a higher process risk and a high tool confidence level.

This integrated view can be used to derive targeted measures for process improvement and tool validation.

Two types of measures are entered for each prioritized combination of task, error state and tool evaluation:

- Possible process mitigations: organizational or procedural countermeasures that contribute to reducing the identified risk (e.g. additional validation steps, manual reviews, release processes).
- Testable tool capabilities: functional requirements for the tool that enable technical validation (e.g. deterministic versioning, audit logging, automatic error detection)

Process mitigations and testable tool capabilities can be selected and documented individually. The result is a traceable, task-specific compilation of all identified risks, their assessments (RPN, TCL) and the derived countermeasures and tool requirements.

7.2.3 List of potential error states (example)

Table 7-1: Assignment of error states to risk groups and development tasks

Risk group (main group + others)	Error state	Development tasks
Product characteristics (regulatory, financial risk)	Loss of data integrity: Corrupted or inconsistent data leads to incorrect model predictions and jeopardizes safety-critical systems.	Data storage & management; data pre-processing & transformation

Product characteristics (financial risk, reputational risk)	Lack of real-time processing: Non-compliance with real-time inference requirements in autonomous or safety-critical applications.	Deep learning & ML frameworks; real-time & batch inference
Regulatory (product characteristics, financial risk)	Inconsistent model reproducibility: Non-deterministic or undocumented training processes lead to unpredictable model behavior.	Experiment tracking & management; model evaluation & explainability; data versioning & management; data collection & integration
Product characteristics (bias & fairness, reputational risk)	Insufficient treatment of edge cases: AI models do not handle rare but safety-critical scenarios correctly.	Data labeling & annotation; feature engineering & extraction; synthetic data generation
Product characteristics (regulatory, financial risk)	Unrecognized model performance drift: Model performance deteriorates over time without corrective action.	Experiment tracking & management; model monitoring & drift detection; real-time & batch inference; anomaly detection
Product characteristics (financial risk, reputational risk)	Insufficient redundancy & fail-over mechanisms: Lack of backup models or failover strategies in critical systems.	CI/CD for ML (MLOps); cloud deployment; edge & IoT deployment
Regulation (transparency, financial risk)	Insufficient error handling & logging: Insufficient logging makes error analysis and root cause identification difficult.	Deep learning & ML frameworks; model evaluation & explainability
Product characteristics (financial risk, reputational risk)	Model instability & poor generalization: Model behaves unpredictably or does not generalize reliably.	Data preprocessing & transformation; feature engineering & extraction; synthetic data generation; hyperparameter tuning; algorithm prototyping & development; deep learning & ML frameworks;

		distributed training & optimization
Data protection (regulatory, reputational risk, financial risk)	Unauthorized data access: Lack of access controls leads to data breaches.	Data storage & management; data collection & integration; cloud deployment
Data protection (regulatory, reputational risk)	Inadequate implementation of the right to erasure: Non-compliance with the GDPR right to be forgotten.	Data storage & management
Data protection (reputational risk, financial risk)	Disclosure of personal data: ML models unintentionally disclose personal data.	Deep learning & ML frameworks; anomaly detection; cloud deployment
Data protection (regulatory, financial risk)	Lack of data minimization: Disproportionate collection and storage of personal data.	Data collection & integration; data quality & bias detection
Data protection (regulatory, reputational risk)	Errors in anonymization & pseudonymization: Insufficient de-identification leads to re-identifiable data.	Feature engineering & extraction; data storage & management; data preprocessing & transformation; synthetic data generation
Data protection (transparency, regulation)	Non-transparent consent management: Lack of transparent tracking of user consent.	Data collection & integration; regulatory compliance monitoring
Data protection (regulatory, financial risk)	Breaches in cross-border data transfer: Non-compliance with international data transfer rules.	Cloud deployment; data storage & management
Transparency (regulation, reputational risk)	Lack of transparency of AI systems: Lack of explainability of AI decisions.	Feature engineering & extraction; model evaluation & explainability; data labeling & annotation; synthetic data generation; deep learning & ML frameworks; edge & IoT deployment

Bias & fairness (regulatory, reputational risk, financial risk)	Bias & fairness non-compliance: Reinforcement of discrimination and breaches of fairness.	Bias & fairness audits; deep learning & ML frameworks; data labeling & annotation; data quality & bias detection; feature engineering & extraction; distributed training & optimization
Regulation (transparency, reputational risk)	Lack of model risk assessment: Lack of classification according to AI Act risk classes.	Regulatory compliance monitoring; risk management
Product features (regulatory, reputational risk, financial risk)	Security vulnerabilities in AI systems: Vulnerability to typical AI attack vectors (adversarial attacks, data poisoning, etc.)	Data storage & management; data pre-processing & transformation; deep learning & ML frameworks; cloud deployment; edge & IoT deployment
Regulation (transparency, reputational risk)	Non-compliant human oversight mechanisms: Lack of human-in-the-loop controls.	Regulatory compliance monitoring; observability & performance tracking
Bias & fairness (reputational risk, regulatory risk)	Lack of ethical safeguards: Ethical principles not sufficiently considered.	Bias & fairness audits; model evaluation & explainability
Regulatory (financial risk, reputational risk)	Insufficient AI impact assessment: Lack of risk-benefit analysis.	Bias & fairness audits, model evaluation & explainability

7.3 Example application of the method

To better illustrate the method, an example is shown below of using the method to evaluate the fictitious tool *MLtoolExample*. This example focuses on the development of an AI-powered evaluation of SPC data in quality management.

7.3.1 Context of the AI-powered SPC evaluation in quality management

The application under consideration is used for the automated analysis of process KPIs in automotive series production. The aim is to detect deviations, trends and anomalies at an early stage in order to ensure process stability and product conformity. It is used within a quality management system, such as IATF 16949 and ISO 9001, and is integrated into APQP/RGA processes.

Incorrect or untraceable model evaluations can lead to incorrect process approvals, unnecessary production interruptions or the release of non-compliant components. In particular, this would affect the requirements related to documented information, change management, validation, requalification, and auditability. Machine learning models extend classic SPC methods by recognizing complex multivariate correlations and gradual process changes. Their reproducibility and continuous monitoring are therefore crucial for compliance with quality and conformity requirements.

7.3.2 Context of the fictitious development tool *MLtoolExample*

The fictitious development tool *MLtoolExample* is an open source platform to support the ML life cycle. The functions for experiment tracking, model versioning and management of training artifacts are particularly relevant for the tool evaluation. The tool enables the structured recording of parameters, metrics, model versions and environmental information and thus creates the technical basis for traceability, reproducibility and auditable model releases.

7.3.3 Step 1: Identify risks for development tasks

In the first step of the risk assessment, potential error states are systematically identified for selected development tasks and evaluated with regard to their impact on quality, conformity and product capability in order to create a reliable basis for the subsequent tool-specific analysis.

7.3.3.1 Sub-step 1.1: Select relevant development tasks

The risk assessment focuses on those development tasks in which the fictitious tool *MLtool/Example* has a direct influence on traceability and quality assurance. These include, in particular:

- Experiment tracking & management
- Data versioning & management
- CI/CD for ML (MLOps)

In this example, the analysis focuses on the development tasks "Experiment tracking & management" and "CI/CD for ML," as these lay the groundwork for reproducible model versions and documented model releases.

7.3.3.2 Sub-step 1.2: Select relevant error states

Two relevant error states are considered for the selected development task.

Table 7-2: Selected error states for the development task "Experiment tracking & management"

Risk group (main group + others)	Error state	Development tasks
Regulatory (product characteristics, financial risk)	Inconsistent model reproducibility: Non-deterministic or undocumented training processes lead to unpredictable model behavior.	Experiment tracking & management; model evaluation & explainability; data versioning & management; data collection & integration
Product characteristics (financial risk, reputational risk)	Insufficient redundancy & fail-over mechanisms: Lack of backup models or failover strategies in critical systems.	CI/CD for ML (MLOps); cloud deployment; edge & IoT deployment

The first error state addresses the lack of reproducibility of training states and model versions due to incomplete or inconsistent entering of training parameters, data statuses or environmental information. As a result,

the model versions in the development of automated analysis applications cannot be clearly reconstructed, validated or formally released. The error state thus directly violates normative requirements for traceability, change release and process validation from the main risk group "Regulatory." Secondly, there is a link to the "product characteristics" risk group, as model versions that cannot be clearly reconstructed can lead to incorrect process evaluations. If, for example, an unvalidated model version is used in live operation, this can influence the assessment of process capability and thus indirectly product conformity.

The second error state addresses inadequate redundancy and failover mechanisms in the operation of the AI-powered analysis application. It describes the case where there is no defined backup strategy or fallback solution in the event of a failure in a model, deployment instance or infrastructure component. In such a scenario, automated process monitoring may be interrupted or outdated or non-validated model versions may be used.

The main risk group is "product characteristics," as the continuous assessment of process capability is part of quality-relevant production monitoring. If the system fails without a defined fallback strategy, the ability to detect process deviations at an early stage may be limited. As a result, there is a risk that non-compliant products will be produced or released.

The "financial risk" risk group is also involved, as system failures can lead to production interruptions, increased testing costs or delayed error detection. There is also a link to the "reputational risk" risk group, particularly if delivery deadlines or quality indicators are affected and this has an impact on customer ratings or supplier status.

7.3.3.3 Sub-step 1.3: Evaluate error states

Only tools whose development is traceable, controlled and fully documented may be used for the automated analysis of process KPIs. The quality of the analysis application depends not only on the subsequent model behavior, but also to a large extent on the rigor with which training processes, model versions and associated parameters are recorded and versioned in the development process.

If model versions, training parameters or environmental conditions are not consistently recorded, it is not possible to reliably prove the conditions under which a certain analysis state was created. This means that there is no formal proof that the analysis tool was developed, validated and released with the necessary methodological care. In the event of errors, such as audit deviations, internal quality analyses or customer complaints, it is therefore not possible to prove that the analysis application was developed in accordance with the requirements for documented information, traceability and change control. This example is shown in the second row of the table in *Figure 12*. For this error state, the severity factor is rated as 9, as the regulatory conformity of the application can be directly called into question in the event of an error. The occurrence factor is rated as 5, as inconsistencies often result from incomplete logging or missing versioning. The detection factor is rated as 7, as deficits typically only become apparent during reproduction attempts, release testing or audits. This results in an RPN of 315 ($RPN = severity \times occurrence \times detection = 9 \times 5 \times 7 = 315$).

The second error state concerns inadequate redundancy and failover mechanisms in the operation of the AI-powered analysis application. In particular, it should be noted that the fictitious *MLtoolExample* tool acts as part of the CI/CD and deployment pipeline and manages and versions model versions and makes them available for live environments. If the fictitious *MLtoolExample* tool is not integrated into a robust backup or rollback strategy or if there are no defined fallback mechanisms for model or infrastructure components, a validated replacement model version cannot be provided in the event of a failure. In such a scenario, continuous process monitoring is interrupted, or unapproved or outdated model versions are used. This example is shown in the first row of the table in *Figure 12*. For this error state, the severity factor is given a value of 8, as there may be an impact on product quality and delivery capability. The occurrence factor is rated as 6, as infrastructure or deployment misconfigurations cannot be ruled out in an industrial environment. The detection factor is rated as 6, as weaknesses in backup or rollback strategies often only become apparent in the event of an

incident or during stress tests. This results in an RPN of 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

Figure 7-11: Evaluation of the error states in the example

7.3.4 Step 2: Tool evaluation and determination of the qualification requirements

In the second step, based on the previously assessed error states, we examine the role played by the development tool in causing these errors and determine the level of confidence required in the tool to adequately manage the identified risks.

7.3.4.1 Sub-step 2.1: Determine the significance of the tool for the error states (tool impact)

With regard to the lack of reproducibility of training and model versions, the fictitious tool *MLtoolExample* has a high tool impact (TI=2), as the tool is directly responsible for the structured logging, versioning and storage of parameters, model artifacts and environmental information. If logging or versioning mechanisms are configured incorrectly or not used in full, deficits arise in the development documentation. These initially have a technical effect on the reconstructability of individual training runs, but in a later step lead to a documentation problem at the application level. As the provision of

evidence for internal and external audits is largely based on the completeness of this information, a tool error can directly affect the regulatory compliance of the entire analysis tool. The tool impact must therefore be classified as high.

The impact of the tool on the error state of insufficient redundancy and failover mechanisms must be considered separately. Although the fictitious tool *MLtoolExample* is part of the CI/CD and deployment pipeline and supports the management of model versions and their provision for different operating environments, it only contributes indirectly to the overall robustness of failover. The actual mechanisms for ensuring system availability, e.g. redundancy, container orchestration, monitoring and automated rollback or fallback processes, are usually provided by the underlying deployment and infrastructure platform. In this context, the fictitious *MLtoolExample* tool merely provides the organizational basis for managing the model versions without implementing the operational failover logic itself. Since there are therefore several independent technical and organizational protection mechanisms that ensure the robustness of the backup and recovery strategy, the direct influence of a tool error on the occurrence of the error state under consideration is limited. The tool impact is therefore classified as low (TI=1), as the fictitious tool *MLtoolExample* contributes to the traceability and provision of model versions, but does not primarily determine the stability or availability of the underlying system architecture.

7.3.4.2 Sub-step 2.2: Evaluate the likelihood of detection of the tool error and determine the tool confidence level (TCL)



Figure 7-12: Deriving the tool confidence level (TCL)

Tool errors relating to reproducibility are often not immediately visible in practice. Incomplete logging, incorrect versioning or inconsistent metadata typically only become apparent during targeted reproduction attempts, as part of model comparisons or during internal and external audits. As long as there is no concrete reason to follow up, a deficit in the development documentation can remain undetected. Since the fictitious tool *MLtoolExample* plays a central role in ensuring compliant development documentation and the tool impact was rated as high (TI=2), the limited immediate detectability of tool errors (TD=3) leads to an increased tool confidence level (TCL=3) in the main risk group "regulatory" (see the second row in the table in Figure 7-13). The required level of confidence in the tool is correspondingly high, as a deficit in documentation can jeopardize the regulatory conformity of the entire analysis application.

In conjunction with inadequate redundancy and failover mechanisms, the detectability of tool errors is limited. Weaknesses in rollback configurations, model identifiers or backup paths often only become visible in the

event of an actual failure or as part of targeted failure tests. As long as no corresponding tests are carried out, deficits in the deployment and recovery logic can remain undetected. The detectability of possible tool errors should be classified as TD = 2 (medium), as configuration or integration problems typically only become visible when interacting with other components in the deployment pipeline. The combination of low tool impact (TI=1) and medium error detectability (TD=2) results in a correspondingly low tool confidence level (TCL=1) for the main risk group "product features" (see the first row in the table in Figure 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

Figure 7-13: Determining the tool confidence level in the example

7.3.4.3 Sub-step 2.3: Final tool evaluation and recommended action based on the RPN and the TCL

The final evaluation of the tool results from the combination of the risk priority indicators (RPN) determined in the first step and the tool confidence level (TCL) determined in the second step. While the RPN describes the inherent criticality of the error state in the process context, the TCL indicates the level of confidence and protection required for the tool.

For the development task "Experiment tracking & management", the error state "Inconsistent model reproducibility" was assessed with a high RPN of 315. In combination with a high tool impact (TI=2) and a high detectability (TD=3) of tool errors, this results in a high TCL (TCL=3) in the main risk group "regulatory" (see the second row in the tables in Figure 7-11 and Figure 7-13). This means that the tool must provide functions that ensure the complete recording and versioning of all experiment-relevant information. This includes, in particular, the systematic logging of random seeds (starting values for random number generators for the reproducibility of training results) and environmental parameters, the versioning of model artifacts and hyperparameters as well as the consistent recording of dependencies such as library versions. In addition, organizational measures are required, such as binding specifications for seed fixing and clear version labels for all experiments. The tool must provide technical support for these requirements by enabling automatic artifact versioning and seamless logging.

For the development task "CI/CD for ML (MLOps)," the error state "Insufficient redundancy and failover mechanisms" was assessed with an RPN of 288. In combination with a low tool impact (TI=1) and a medium detectability of configuration errors (TD=2), this results in a low TCL (TCL=1) in the main risk group "product characteristics" (see first row in the tables in Figure 7-11 and Figure 7-13). This leads to the requirement that the tool must support the controlled provision of multiple model versions, defined rollback mechanisms and validated fallback strategies. In particular, automated validation checks before going live, clearly defined deployment scenarios and the simulation of failure and stress scenarios are recommended. The fictitious tool *MLtoolExample* should therefore be integrated into a CI/CD infrastructure in such a way that conditional deployments, rollback triggers and the parallel validation of multiple model versions are supported. This ensures that a tested and approved replacement model version is available in the event of a failure.

The combined consideration of RPN and TCL shows that, despite different risk perspectives, the two error states place different demands on the safeguarding of the tool. In the first error state, the focus is on the verifiability of the development rigor, as deficits in logging and versioning can directly affect the regulatory conformity of the analysis application. This results in an increased need to safeguard the tool with regard to traceability and documentation.

In the second error state, on the other hand, the focus is on the resilience and controllability of the operating state. As the direct influence of the tool on system availability is limited and additional infrastructural safety mechanisms exist, there is a lower tool-specific safeguarding requirement here. The recommended action is therefore to specifically qualify the tool in the first case, for example through defined test cases, reference datasets or documented validation of the tool results, and to safeguard it through binding configuration and documentation guidelines. In the second case, correct integration into the existing CI/CD and deployment infrastructure and regular testing of the failover mechanisms must be ensured.

7.4 Overall classification in the context of tool qualification

The method enables a holistic consideration of process and tool risks along the AI life cycle. It supports developers in not only recognizing risks, but also specifically relating them to the reliability and trustworthiness of the tools used.

The example shown here illustrates the logic of risk-based tool evaluation in quality management. The starting point is not the tool itself, but rather the identification of specific error states in the context of defined development tasks. Only on this basis can we analyze the extent to which the tool influences the occurrence or control of these defects and how effectively potential tool defects can be detected.

The two considered error states demonstrate that different risk dimensions are addressed, even though the same development task and the same tool are being examined. In both cases, the high tool impact and the limited direct detectability of errors result in an increased tool confidence level.

This makes it clear that the need for qualification is not derived abstractly from the type of tool, but rather from the systematic linking of error state, risk group, process context and detectability of tool errors.

Quality Management in the Automotive Industry

The current versions of the VDA publications covering quality management in the automotive industry can be found on the internet under <https://www.vda-qmc.de>.

You may also order via this homepage.

Reference:

Verband der Automobilindustrie e. V. (VDA)
Qualitäts Management Center (QMC)

10117 Berlin, Behrenstr. 35

Phone +49 (0) 30 89 78 42-235, Fax +49 (0) 30 89 78 42-605

Email: info@vda-qmc.de, Internet: www.vda-qmc.de

VDA QMC

German Association of the Automotive Industry
Quality Management Center