

Q in KI: Anforderungen an Tools und Prozesse zur Absicherung

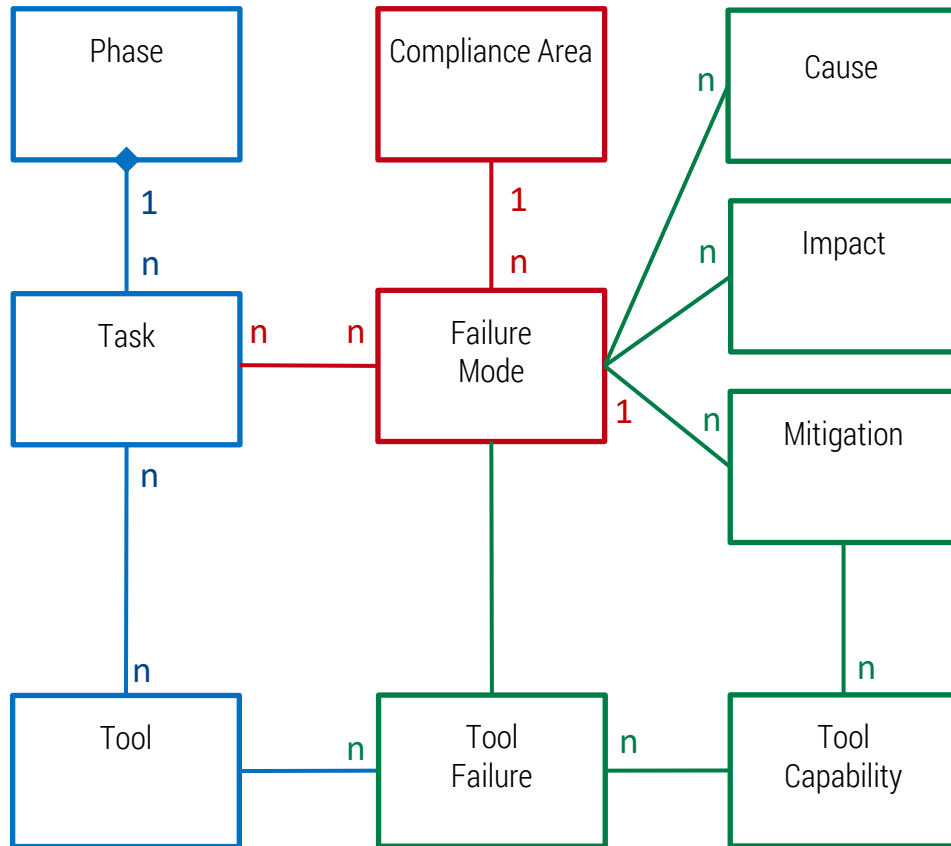
Anleitung zur Nutzung von AssessML

Dr. Björn Schünemann (bjoern.schuenemann@aqigmbh.de)

Dr. Jürgen Großmann (juergen.grossmann@fokus.fraunhofer.de)

Vorgehen im Projekt und verwendete Terminologie

Vom KI-Entwicklungszyklus zur Werkzeugqualifizierung



- Systematische Aufstellung der **Phasen (Phase)** und **Aufgaben (Task)** im KI-Entwicklungszyklus sowie beispielhafter **Werkzeuge (Tool)**.
- Identifikation der **Abweichungen / Fehlerzustände (Failure Mode)** entlang der **Aufgaben** für jedes **Konformitätsgebiet (Compliance Area)**.
- Identifikation von **Ursachen (Cause)**, **Wirkungen (Impact)** und möglicher **Gegenmaßnahmen (Mitigation)** für jede **Abweichung (Failure Mode)**.
- Identifikation von spezifischen **nicht-konformem Verhalten von Werkzeugen (Tool Failure)** und benötigter **Werkzeugfähigkeiten (ToolCapability)** für eine **Abweichung (Failure Mode)**.

Phase (Phase): Phase aus dem KI-Entwicklungszyklus

Task (Aufgabe): Aufgabe während des KI-Entwicklungszyklus, die von Werkzeugen unterstützt wird.

Tool (Werkzeug): Werkzeug, das eine Aufgabe unterstützt/ausführt.

Compliance Area (Konformitätsgebiet): Konformitätsgebiet (funktionale Sicherheit, Datenschutz, KI-Regulierung).

Failure Mode (Fehlerzustand): Abweichung, die zu Nichtkonformität führt.

Tool Failure (Werkzeugfehler): Nicht-konformes Verhalten von Werkzeugen als Ursache für eine Abweichung

Cause (Ursache): Ursachen für das Zustandekommen der Abweichung.

Impact (Auswirkung): Auswirkungen, die die Abweichung hat.

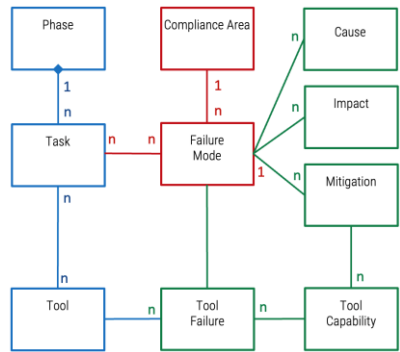
Mitigation (Minderung): Maßnahme zur Vermeidung der Abweichung.

Tool Capability (Werkzeugfähigkeiten): Benötigte Fähigkeiten eines Werkzeugs zur Vermeidung der Abweichung.

Systematische Ausarbeitung der Inhalte

Ergebnis ist eine Python-basierte Webanwendung

Terminologie



Inhalte der Excel-Tabelle werden gemäß verwendeter Terminologie erarbeitet

Compliance Area	Phase	Task	Failure Mode	Impact	Severity	Causes	Mitigation	Occurrence	Detectability	RPN	Potential Tool Failure	Recommended Tool Capability
Functional Safety	Training & Validation	Deep Learning & ML Frameworks	Lack of Real-Time Processing	Trained models and supporting frameworks are unable to deliver low-latency, real-time inference performance required for critical applications such as autonomous systems.	High	Model architectures are too large or complex for target real-time environments. Frameworks do not leverage hardware acceleration efficiently (e.g., GPU, TPU, Edge devices). Pipeline overhead (e.g., preprocessing, postprocessing) adds excessive latency during inference.	Apply model optimization techniques such as pruning, quantization, or knowledge distillation. Ensure deployment-ready models are benchmarked and optimized for specific hardware accelerators. Optimize full pipeline latency by batching operations and reducing unnecessary transformations.				May fail to guarantee deterministic training execution and runtime validation, resulting in trained models that do not meet real-time safety-critical performance requirements.	Framework support for model optimization (e.g., TensorRT, ONNX Runtime optimizations). Support for hardware-specific runtime optimizations and deployment targeting (e.g., EdgeTPU, CUDA kernels). Framework support for pipeline fusion, minimal preprocessing APIs, and asynchronous inference.
Functional Safety	Training & Validation	Deep Learning & ML Frameworks	Inadequate Error Handling & Logging	Frameworks and models lack robust error detection, recovery, and structured logging capabilities, making root cause analysis and system recovery difficult.	High	Training and inference exceptions are not properly caught or logged. Losses, gradients, or other training dynamics are not systematically monitored for anomalies. Training and deployment pipelines lack comprehensive logging and metadata tracking.	Integrate structured exception handling and mandatory logging for all critical training and inference operations. Implement runtime monitors for numerical instabilities (e.g., NaNs, divergence) and alert on thresholds. Embed pipeline-wide metadata capture and systematic experiment logging into the training framework.				May fail to provide comprehensive error logging and runtime monitoring, resulting in trained models with limited traceability for failure analysis in safety-critical systems.	Framework support for structured error reporting, custom callbacks, and logging integration. Built-in hooks for anomaly detection during training and validation (gradient checkers, divergence detectors). ML lifecycle management tools (e.g., MLflow, Weights & Biases) integration for metadata and error tracking.
Functional Safety	Training & Validation	Deep Learning & ML Frameworks	Model Instability & Poor Generalization	Inconsistent training leading to unreliable AI behaviour in critical applications.	High	Lack of deterministic/reproducible training settings due to parallel computation and floating-point arithmetic. Limited built-in validation or monitoring. No built-in numerical stability mechanisms.	Implement deterministic training pipelines with seed control and precision configuration to ensure repeatability. Integrate continuous performance monitoring during training with automatic validation checkpoints. Use frameworks that incorporate gradient clipping, normalization, and regularization for improved numerical robustness.				May fail to enforce deterministic training procedures and runtime validation, resulting in trained models with unpredictable behavior and poor generalization in safety-critical applications.	Explicit deterministic training support with reproducibility settings and controlled parallelism. Built-in validation dashboards and metric tracking APIs. Numerical stability modules with configurable regularization options.

Dashboard wurde basierend auf der Excel-Tabelle erstellt

Risikobasierte Werkzeugqualifizierung (Details)

Zweistufiges Verfahren zur Ermittlung von aufgabenbezogenen Risiken und einer Qualifizierung von Werkzeugen entlang der Risiken

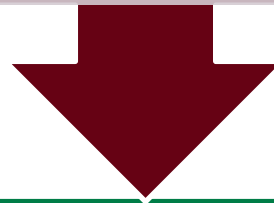
1. Ermittlung von Risiken in den Entwicklungsaufgaben (nach FMEA)

$$RPN = Severity * Occurrence * Detection$$

1.1 Auswahl der relevanten Aufgaben und Phasen

1.2 Auswahl der relevanten Konformitätsbereiche und Fehlermodi

1.3 Berechnung des RPN für ausgesuchte Fehlermodi in einer Aufgabe

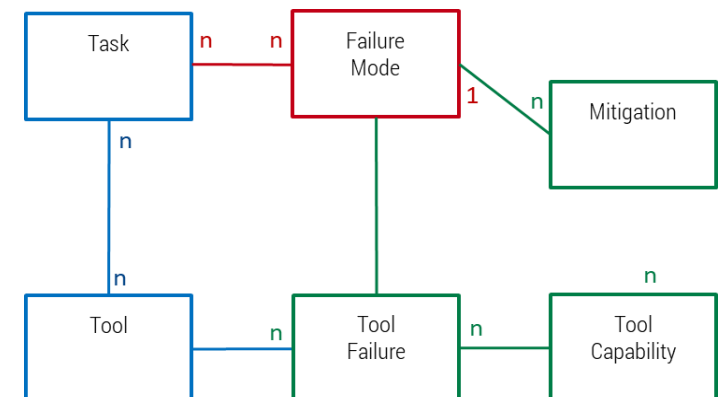
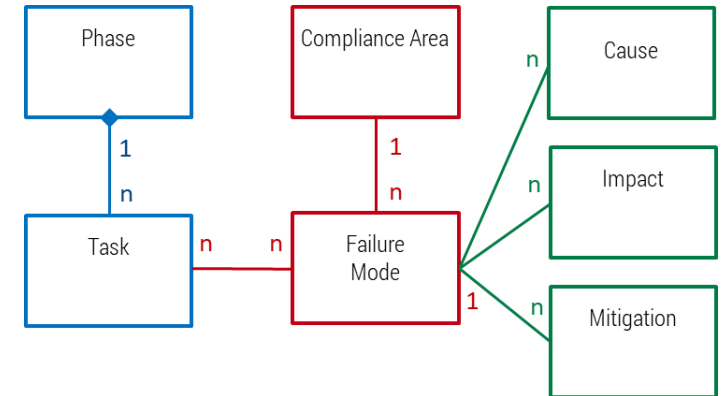


2. Ermittlung des Qualifizierungsbedarfs für KI-Werkzeuge

2.1 Einschätzung über die Bedeutung des Werkzeugs für eine Abweichung (Tool Impact)

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

2.3 Abschließende Werkzeugbewertung und Handlungsempfehlung auf Basis der RPN und des TCL



AssessML: Aufbau des Dashboardes

Startseite

AssessML: Assessment of development tools for ML

Choose Phase(s): Choose Task(s): Choose Tool(s):

Failure Mode Selection Task Assessment Tool Assessment Assessment Summary

Settings filename (e.g. settin|)

Failure Mode and Compliance Area Selection

Select	Failure Mode	Description
<input checked="" type="checkbox"/>	AI Regulation	
<input checked="" type="checkbox"/>	Lack of AI Transparency	Failure to provide explanations for AI decisions, violating AI Act transparency requirements.
<input checked="" type="checkbox"/>	Bias & Fairness Non-Compliance	High-risk AI models reinforcing biases, leading to discrimination and non-compliance with AI Act fairness guidelines.
<input checked="" type="checkbox"/>	Lack of Model Risk Assessment	Failure to categorize AI models under AI Act risk levels and document safety concerns.
<input checked="" type="checkbox"/>	Security Vulnerabilities in AI	Exposure to adversarial attacks compromising AI safety and trustworthiness.
<input checked="" type="checkbox"/>	Non-compliant Human Oversight Mechanisms	Failure to implement necessary human-in-the-loop controls in high-risk AI applications.
<input checked="" type="checkbox"/>	Insufficient AI Impact Assessment	Failure to conduct proper risk-benefit analysis for high-risk AI applications.
<input checked="" type="checkbox"/>	Lack of Ethical Safeguards in AI	Failure to ensure ethical principles (e.g., non-harm, fairness) in AI decision-making.
<input checked="" type="checkbox"/>	Inadequate Monitoring of Deployed AI Systems	Lack of continuous assessment of AI systems in real-world settings.
<input checked="" type="checkbox"/>	Lack of Auditability & Documentation	Inability to provide detailed logs of data usage, model training, validation etc. hindering compliance audits.
<input checked="" type="checkbox"/>	Data Protection	
<input checked="" type="checkbox"/>	Unauthorized Data Access	Lack of strict access control leading to potential data breaches and privacy violations.
<input checked="" type="checkbox"/>	Failure to Handle Right to Erasure	Non-compliance with GDPR's Right to be Forgotten, leading to legal risks.
<input checked="" type="checkbox"/>	Personal Data Leakage	ML models unintentionally exposing personal data during training or inference.
<input checked="" type="checkbox"/>	Lack of Data Minimization	Collecting and storing excessive personal data, violating GDPR principles.
<input checked="" type="checkbox"/>	Failure in Data Anonymization & Pseudonymization	Inadequate de-identification techniques lead to re-identifiable personal data.
<input checked="" type="checkbox"/>	Non-Transparent User Consent Management	Lack of clear user consent tracking for data collection and processing.
<input checked="" type="checkbox"/>	Cross-border Data Transfer Violations	Failure to comply with international data transfer rules under GDPR/Schrems II.
<input checked="" type="checkbox"/>	Functional Safety	
<input checked="" type="checkbox"/>	Data Integrity Failure	Corrupt or inconsistent data leading to incorrect model predictions, endangering safety-critical systems.
<input checked="" type="checkbox"/>	Lack of Real-Time Processing	Failure to meet real-time inference requirements in autonomous or safety-critical applications.
<input checked="" type="checkbox"/>	Failure Mode Undetection	Inability to detect and mitigate failure conditions in deployed AI models, leading to high-risk scenarios.
<input checked="" type="checkbox"/>	Inconsistent Model Reproducibility	Non-deterministic model outputs lead to unpredictable AI system behavior.
<input checked="" type="checkbox"/>	Insufficient Edge Case Handling	AI models fail to correctly handle rare but critical safety scenarios, leading to unreliable operation.
<input checked="" type="checkbox"/>	Undetected Model Performance Drift	Model performance degrades over time without triggering corrective actions, leading to unsafe decisions.
<input checked="" type="checkbox"/>	Insufficient Redundancy & Failover Mechanisms	Lack of backup AI models or failover strategies in critical systems.
<input checked="" type="checkbox"/>	Inadequate Error Handling & Logging	Poor logging and monitoring hinder failure analysis and root cause identification.
<input checked="" type="checkbox"/>	Model Instability & Poor Generalization	Model may perform unpredictably or fail to generalize to unseen conditions, undermining functional safety.

Die Startseite ist in drei Sektoren untergliedert:

[A: Auswahl für Phasen und Aufgaben](#)

Auswahltabs für zu betrachtende Phasen, Aufgaben oder Werkzeuge.

[B: Auswahl der Fehlerzustände und Assessment-durchführung](#)

Dient zur Auswahl der betrachteten Fehlerzustände, sowie zur Ermittlung von Risiken in den Entwicklungsaufgaben und des Qualifizierungsbedarfs für KI-Werkzeuge.

[C: Speichern und Laden](#)

Voreingestellte oder bereits durchgeführte Assessments können gespeichert und abgerufen werden.

1. Ermittlung von Risiken in den Entwicklungsaufgaben

1.1 Auswahl der relevanten Aufgaben und Phasen

1. Ermittlung von Risiken in den Entwicklungsaufgaben (nach FMEA)

$$\text{RPN} = \text{Severity} * \text{Occurrence} * \text{Detection}$$

1.1 Auswahl der relevanten Aufgaben und Phasen

1.2. Auswahl der relevanten Konformitätsbereiche und Fehlerzustände

1.3. Berechnung des RPN für ausgesuchte Fehlerzustände in einer Aufgabe

AssessML: Auswahlbereich

Wozu dient der Auswahlbereich?

Das Dashboard beinhaltet 25 Fehlerzustände, die von Start an ausgewählt sind.

Dadurch werden sämtliche Zustände automatisch in das Task und Tool Assessment sowie in die Assessment Summary übernommen, wodurch das Board schnell unübersichtlich werden kann.

AssessML: Assessment of development tools for ML

Choose Phases(s): Choose Task(s): Choose Tool(s):

Failure Mode and Compliance Area Selection

Select	Failure Mode	Description
<input checked="" type="checkbox"/>	AI Regulation	
<input checked="" type="checkbox"/>	Lack of AI Transparency	Failure to provide explanations for AI decisions, violating AI Act transparency requirements.
<input checked="" type="checkbox"/>	Bias & Fairness Non-Compliance	High-risk AI models reinforcing biases, leading to discrimination and non-compliance with AI Act fairness guidelines.
<input checked="" type="checkbox"/>	Lack of Model Risk Assessment	Failure to categorize AI models under AI Act risk levels and document safety concerns.
<input checked="" type="checkbox"/>	Security Vulnerabilities in AI	Exposure to adversarial attacks compromising AI safety and trustworthiness.
<input checked="" type="checkbox"/>	Non-compliant Human Oversight Mechanisms	Failure to implement necessary human-in-the-loop controls in high-risk AI applications.
<input checked="" type="checkbox"/>	Insufficient AI Impact Assessment	Failure to conduct proper risk-benefit analysis for high-risk AI applications.
<input checked="" type="checkbox"/>	Lack of Ethical Safeguards in AI	Failure to ensure ethical principles (e.g., non-harm, fairness) in AI decision-making.
<input checked="" type="checkbox"/>	Inadequate Monitoring of Deployed AI Systems	Lack of continuous assessment of AI systems in real-world settings.
<input checked="" type="checkbox"/>	Lack of Auditability & Documentation	Inability to provide detailed logs of data usage, model training, validation etc. hindering compliance audits.
<input checked="" type="checkbox"/>	Data Protection	
<input checked="" type="checkbox"/>	Unauthorized Data Access	Lack of strict access control leading to potential data breaches and privacy violations.
<input checked="" type="checkbox"/>	Failure to Handle Right to Erasure	Non-compliance with GDPR's Right to be Forgotten, leading to legal risks.
<input checked="" type="checkbox"/>	Personal Data Leakage	ML models unintentionally exposing personal data during training or inference.
<input checked="" type="checkbox"/>	Lack of Data Minimization	Collecting and storing excessive personal data, violating GDPR principles.
<input checked="" type="checkbox"/>	Failure in Data Anonymization & Pseudonymization	Inadequate de-identification techniques lead to re-identifiable personal data.
<input checked="" type="checkbox"/>	Non-Transparent User Consent Management	Lack of clear user consent tracking for data collection and processing.
<input checked="" type="checkbox"/>	Cross-border Data Transfer Violations	Failure to comply with international data transfer rules under GDPR/Schrems II.
<input checked="" type="checkbox"/>	Functional Safety	
<input checked="" type="checkbox"/>	Data Integrity Failure	Corrupt or inconsistent data leading to incorrect model predictions, endangering safety-critical systems.
<input checked="" type="checkbox"/>	Lack of Real-Time Processing	Failure to meet real-time inference requirements in autonomous or safety-critical applications.
<input checked="" type="checkbox"/>	Failure Mode Undetection	Inability to detect and mitigate failure conditions in deployed AI models, leading to high-risk scenarios.
<input checked="" type="checkbox"/>	Inconsistent Model Reproducibility	Non-deterministic model outputs lead to unpredictable AI system behavior.
<input checked="" type="checkbox"/>	Insufficient Edge Case Handling	AI models fail to correctly handle rare but critical safety scenarios, leading to unreliable operation.
<input checked="" type="checkbox"/>	Undetected Model Performance Drift	Model performance degrades over time without triggering corrective actions, leading to unsafe decisions.
<input checked="" type="checkbox"/>	Insufficient Redundancy & Failover Mechanisms	Lack of backup AI models or failover strategies in critical systems.
<input checked="" type="checkbox"/>	Inadequate Error Handling & Logging	Poor logging and monitoring hinder failure analysis and root cause identification.
<input checked="" type="checkbox"/>	Model Instability & Poor Generalization	Model may perform unpredictably or fail to generalize to unseen conditions, undermining functional safety.

Task Assessment

Tool Assessment

Assessment Summary

AssessML: Auswahlbereich

Wozu dient der Auswahlbereich?

Um eine gezieltere und effizientere Bearbeitung zu ermöglichen, bietet das Dashboard die Möglichkeit, über den **Auswahlbereich** zu filtern.

Jede getroffene Auswahl reduziert die angezeigten Fehlerzustände und erleichtert damit die Bearbeitung der folgenden Schritte in den Assessments und in dem Assessments Summary.

The screenshot shows a dashboard with a red box highlighting the filter section. It contains three dropdown menus: 'Choose Phase(s):', 'Choose Task(s):', and 'Choose Tool(s):'. Below these are four tabs: 'Failure Mode Selection', 'Task Assessment', 'Tool Assessment', and 'Assessment Summary'. A second row of dropdown menus is also visible below the tabs.

Failure Mode and Compliance Area Selection

Select	Failure Mode	Description
<input checked="" type="checkbox"/>	AI Regulation	
<input checked="" type="checkbox"/>	Lack of AI Transparency	Failure to provide explanations for AI decisions, violating AI Act transparency requirements.
<input checked="" type="checkbox"/>	Bias & Fairness Non-Compliance	High-risk AI models reinforcing biases, leading to discrimination and non-compliance with AI Act fairness guidelines.
<input checked="" type="checkbox"/>	Lack of Model Risk Assessment	Failure to categorize AI models under AI Act risk levels and document safety concerns.
<input checked="" type="checkbox"/>	Security Vulnerabilities in AI	Exposure to adversarial attacks compromising AI safety and trustworthiness.
<input checked="" type="checkbox"/>	Non-compliant Human Oversight Mechanisms	Failure to implement necessary human-in-the-loop controls in high-risk AI applications.
<input checked="" type="checkbox"/>	Inadequate AI Impact Assessment	Failure to conduct proper risk-benefit analysis for high-risk AI applications.
<input checked="" type="checkbox"/>	Lack of Ethical Safeguards in AI	Failure to ensure ethical principles (e.g., non-harm, fairness) in AI decision-making.
<input checked="" type="checkbox"/>	Inadequate Monitoring of Deployed AI Systems	Lack of continuous assessment of AI systems in real-world settings.
<input checked="" type="checkbox"/>	Lack of Auditability & Documentation	Inability to provide detailed logs of data usage, model training, validation etc. hindering compliance audits.
<input checked="" type="checkbox"/>	Data Protection	
<input checked="" type="checkbox"/>	Unauthorized Data Access	Lack of strict access control leading to potential data breaches and privacy violations.
<input checked="" type="checkbox"/>	Failure to Handle Right to Erasure	Non-compliance with GDPR's Right to be Forgotten, leading to legal risks.
<input checked="" type="checkbox"/>	Personal Data Leakage	ML models unintentionally exposing personal data during training or inference.
<input checked="" type="checkbox"/>	Lack of Data Minimization	Collecting and storing excessive personal data, violating GDPR principles.
<input checked="" type="checkbox"/>	Failures in Data Anonymization & Pseudonymization	Inadequate de-identification techniques lead to re-identifiable personal data.
<input checked="" type="checkbox"/>	Non-Transparent User Consent Management	Lack of clear user consent tracking for data collection and processing.
<input checked="" type="checkbox"/>	Cross-border Data Transfer Violations	Failure to comply with international data transfer rules under GDPR/Schrems II.
<input checked="" type="checkbox"/>	Functional Safety	
<input checked="" type="checkbox"/>	Data Integrity Failure	Corrupt or inconsistent data leading to incorrect model predictions, endangering safety-critical systems.
<input checked="" type="checkbox"/>	Lack of Real-Time Processing	Failure to meet real-time inference requirements in autonomous or safety-critical applications.
<input checked="" type="checkbox"/>	Failure Mode Undetection	Inability to detect and mitigate failure conditions in deployed AI models, leading to high-risk scenarios.
<input checked="" type="checkbox"/>	Inconsistent Model Reproducibility	Non-deterministic model outputs lead to unpredictable AI system behavior.
<input checked="" type="checkbox"/>	Inadequate Edge Case Handling	AI models fail to correctly handle rare but critical safety scenarios, leading to unreliable operation.
<input checked="" type="checkbox"/>	Undetected Model Performance Drift	Model performance degrades over time without triggering corrective actions, leading to unsafe decisions.
<input checked="" type="checkbox"/>	Inadequate Redundancy & Failover Mechanisms	Lack of backup AI models or failover strategies in critical systems.
<input checked="" type="checkbox"/>	Inadequate Error Handling & Logging	Poor logging and monitoring hinder failure analysis and root cause identification.
<input checked="" type="checkbox"/>	Model Intuitibility & Poor Generalization	Model may perform unpredictably or fail to generalize to unseen conditions, undermining functional safety.

[Settings \(Name: n.g. with\)](#) [Save Settings](#) [Load Settings](#)

The screenshot shows four panels of the dashboard: 'Failure Mode Overview', 'Task Assessment', 'Tool Assessment', and 'Assessment Summary'. Each panel displays a list of failure modes with their respective risk levels and status.

Anleitung für Assessments folgt ab Folie 18

AssessML: Auswahlbereich

1. Auswahlbereich

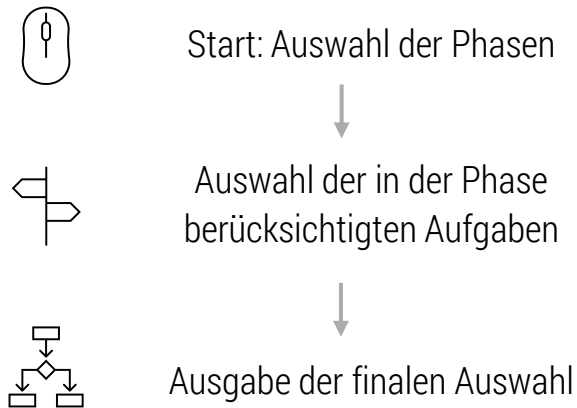
Choose Phase(s):

Choose Task(s):

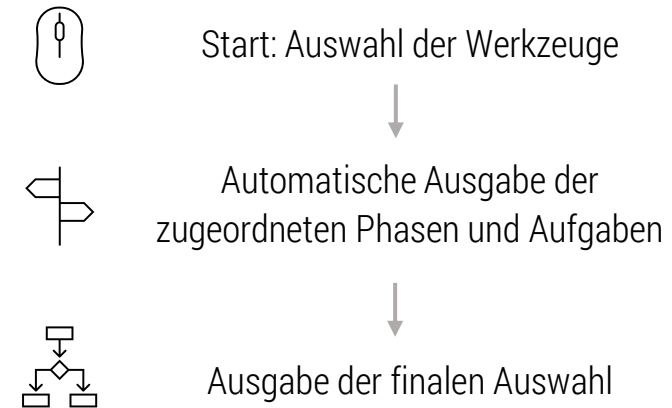
Choose Tool(s):

Der **Auswahlbereich** unterteilt sich in die drei **Suchtabs** *Phase, Aufgabe und Werkzeuge*, wodurch dem Nutzer ermöglicht wird, die Untersuchung auf die präferierten Phasen und Aufgaben zu beschränken. Dabei gibt es **zwei** Arten Phase und Aufgabe zu bestimmen:

1. Weg: Phasen und Aufgaben



2. Weg: Werkzeuge



AssessML: Auswahlbereich

1. Weg: Phasen und Aufgaben

Start: Auswahl der betrachteten Phasen

Choose Phase(s): Choose Task(s): Choose Tool(s):

	Task Assessment	Tool Assessment	Assessment Summary
	Description		
<input checked="" type="checkbox"/> All Phases			
<input type="checkbox"/> Data Curation			
<input type="checkbox"/> Deployment			
<input type="checkbox"/> Experimentation			
<input type="checkbox"/> Inference & Monitoring			
<input type="checkbox"/> Training & Validation			
<input checked="" type="checkbox"/> Lack of AI Transparency			Failure to provide explanations for AI decisions, violating AI Act transparency requirements.
<input checked="" type="checkbox"/> Bias & Fairness Non-Compliance			High-risk AI models reinforcing biases, leading to discrimination and non-compliance with AI Act fairness guidelines.
<input checked="" type="checkbox"/> Lack of Model Risk Assessment			Failure to categorize AI models under AI Act risk levels and document safety concerns.
<input checked="" type="checkbox"/> Security Vulnerabilities in AI			Exposure to adversarial attacks compromising AI safety and trustworthiness.
<input checked="" type="checkbox"/> Non-compliant Human Oversight Mechanisms			Failure to implement necessary human-in-the-loop controls in high-risk AI applications.
<input checked="" type="checkbox"/> Insufficient AI Impact Assessment			Failure to conduct proper risk-benefit analysis for high-risk AI applications.
<input checked="" type="checkbox"/> Lack of Ethical Safeguards in AI			Failure to ensure ethical principles (e.g., non-harm, fairness) in AI decision-making.
<input checked="" type="checkbox"/> Inadequate Monitoring of Deployed AI Systems			Lack of continuous assessment of AI systems in real-world settings.
<input checked="" type="checkbox"/> Lack of Auditability & Documentation			Inability to provide detailed logs of data usage, model training, validation etc. hindering compliance audits.
<input checked="" type="checkbox"/> Data Protection			
<input checked="" type="checkbox"/> Unauthorized Data Access			Lack of strict access control leading to potential data breaches and privacy violations.
<input checked="" type="checkbox"/> Failure to Handle Right to Erasure			Non-compliance with GDPR's Right to be Forgotten, leading to legal risks.
<input checked="" type="checkbox"/> Personal Data Leakage			ML models unintentionally exposing personal data during training or inference.
<input checked="" type="checkbox"/> Lack of Data Minimization			Collecting and storing excessive personal data, violating GDPR principles.
<input checked="" type="checkbox"/> Failure in Data Anonymization & Pseudonymization			Inadequate de-identification techniques lead to re-identifiable personal data.
<input checked="" type="checkbox"/> Non-Transparent User Consent Management			Lack of clear user consent tracking for data collection and processing.
<input checked="" type="checkbox"/> Cross-border Data Transfer Violations			Failure to comply with international data transfer rules under GDPR/Schrems II.
<input checked="" type="checkbox"/> Functional Safety			
<input checked="" type="checkbox"/> Data Integrity Failure			Corrupt or inconsistent data leading to incorrect model predictions, endangering safety-critical systems.
<input checked="" type="checkbox"/> Lack of Real-Time Processing			Failure to meet real-time inference requirements in autonomous or safety-critical applications.
<input checked="" type="checkbox"/> Failure Mode Undetection			Inability to detect and mitigate failure conditions in deployed AI models, leading to high-risk scenarios.
<input checked="" type="checkbox"/> Inconsistent Model Reproducibility			Non-deterministic model outputs lead to unpredictable AI system behavior.
<input checked="" type="checkbox"/> Insufficient Edge Case Handling			AI models fail to correctly handle rare but critical safety scenarios, leading to unreliable operation.
<input checked="" type="checkbox"/> Undetected Model Performance Drift			Model performance degrades over time without triggering corrective actions, leading to unsafe decisions.
<input checked="" type="checkbox"/> Insufficient Redundancy & Failover Mechanisms			Lack of backup AI models or failover strategies in critical systems.
<input checked="" type="checkbox"/> Inadequate Error Handling & Logging			Poor logging and monitoring hinder failure analysis and root cause identification.
<input checked="" type="checkbox"/> Model Instability & Poor Generalization			Model may perform unpredictably or fail to generalize to unseen conditions, undermining functional safety.

Settings filename (e.g. sett1)

AssessML: Auswahlbereich

1. Weg: Phasen und Aufgaben

Auswahl der in der Phase berücksichtigten Aufgaben

Choose Phase(s):
× Training & Validation ×

Choose Task(s):
Select...
Bias & Fairness Audits
Data Versioning & Management
Deep Learning & ML Frameworks
Distributed Training & Optimization
Model Evaluation & Explainability
Performance Metrics & Validation

Choose Tool(s):
Select...
Assessment
Assessment Summary

Failure Mode Selection

Failure Mode and Compliance Area Selection

Select	Failure Mode
<input checked="" type="checkbox"/>	AI Regulation
<input checked="" type="checkbox"/>	Lack of AI Transparency Failure to provide explanations for AI decisions, violating AI Act transparency requirements.
<input checked="" type="checkbox"/>	Bias & Fairness Non-Compliance High-risk AI models reinforcing biases, leading to discrimination and non-compliance with AI Act fairness guidelines.
<input checked="" type="checkbox"/>	Lack of Ethical Safeguards in AI Failure to ensure ethical principles (e.g., non-harm, fairness) in AI decision-making.
<input checked="" type="checkbox"/>	Lack of Auditability & Documentation Inability to provide detailed logs of data usage, model training, validation etc. hindering compliance audits.
<input checked="" type="checkbox"/>	Data Protection
<input checked="" type="checkbox"/>	Personal Data Leakage ML models unintentionally exposing personal data during training or inference.
<input checked="" type="checkbox"/>	Functional Safety
<input checked="" type="checkbox"/>	Lack of Real-Time Processing Failure to meet real-time inference requirements in autonomous or safety-critical applications.
<input checked="" type="checkbox"/>	Failure Mode Undetection Inability to detect and mitigate failure conditions in deployed AI models, leading to high-risk scenarios.
<input checked="" type="checkbox"/>	Inconsistent Model Reproducibility Non-deterministic model outputs lead to unpredictable AI system behavior.
<input checked="" type="checkbox"/>	Inadequate Error Handling & Logging Poor logging and monitoring hinder failure analysis and root cause identification.
<input checked="" type="checkbox"/>	Model Instability & Poor Generalization Model may perform unpredictably or fail to generalize to unseen conditions, undermining functional safety.

Settings filename (e.g. setti) Save Settings Load Settings

Hinweis:

Durch Auswahl der Phasen erfolgt eine automatische Filterung der Fehlerzustände.

AssessML: Auswahlbereich

1. Weg: Phasen und Aufgaben

Ausgabe der finalen Auswahl

Choose Phase(s): Choose Task(s): Choose Tool(s):

Failure Mode Selection	Task Assessment	Tool Assessment	Assessment Summary
------------------------	-----------------	-----------------	--------------------

Failure Mode and Compliance Area Selection

Select	Failure Mode	Description
<input checked="" type="checkbox"/>	AI Regulation	
<input checked="" type="checkbox"/>	Lack of AI Transparency	Failure to provide explanations for AI decisions, violating AI Act transparency requirements.
<input checked="" type="checkbox"/>	Bias & Fairness Non-Compliance	High-risk AI models reinforcing biases, leading to discrimination and non-compliance with AI Act fairness guidelines.
<input checked="" type="checkbox"/>	Data Protection	
<input checked="" type="checkbox"/>	Personal Data Leakage	ML models unintentionally exposing personal data during training or inference.
<input checked="" type="checkbox"/>	Functional Safety	
<input checked="" type="checkbox"/>	Lack of Real-Time Processing	Failure to meet real-time inference requirements in autonomous or safety-critical applications.
<input checked="" type="checkbox"/>	Inadequate Error Handling & Logging	Poor logging and monitoring hinder failure analysis and root cause identification.
<input checked="" type="checkbox"/>	Model Instability & Poor Generalization	Model may perform unpredictably or fail to generalize to unseen conditions, undermining functional safety.

Settings filename (e.g. settii)

Nachdem die Tabs „Phasen“ und „Aufgaben“ von links nach rechts durchgegangen sind, zeigt das Dashboard alle potenziellen Fehlerzustände an, die durch die Filterung übrig geblieben sind. → Es erfolgt eine Einschränkung der möglichen Fehlerzustände!

AssessML: Auswahlbereich

2. Weg: Werkzeuge

Start: Auswahl der verwendeten Werkzeuge

Choose Phase(s): Choose Task(s): Choose Tool(s):

Failure Mode Selection Task Assessment Tool Assessment

Failure Mode and Compliance Area Selection

Select	Failure Mode	Description
<input checked="" type="checkbox"/>	AI Regulation	
<input checked="" type="checkbox"/>	Lack of AI Transparency	Failure to provide explanations for AI decisions, violating AI Act transparency requirements.
<input checked="" type="checkbox"/>	Bias & Fairness Non-Compliance	High-risk AI models reinforcing biases, leading to discrimination and non-compliance with AI Act fairness guidelines.
<input checked="" type="checkbox"/>	Lack of Model Risk Assessment	Failure to categorize AI models under AI Act risk levels and document safety concerns.
<input checked="" type="checkbox"/>	Security Vulnerabilities in AI	Exposure to adversarial attacks compromising AI safety and trustworthiness.
<input checked="" type="checkbox"/>	Non-compliant Human Oversight Mechanisms	Failure to implement necessary human-in-the-loop controls in high-risk AI applications.
<input checked="" type="checkbox"/>	Insufficient AI Impact Assessment	Failure to conduct proper risk-benefit analysis for high-risk AI applications.
<input checked="" type="checkbox"/>	Lack of Ethical Safeguards in AI	Failure to ensure ethical principles (e.g., non-harm, fairness) in AI decision-making.
<input checked="" type="checkbox"/>	Inadequate Monitoring of Deployed AI Systems	Lack of continuous assessment of AI systems in real-world settings.
<input checked="" type="checkbox"/>	Lack of Auditability & Documentation	Inability to provide detailed logs of data usage, model training, validation etc. hindering compliance audits.
<input checked="" type="checkbox"/>	Data Protection	
<input checked="" type="checkbox"/>	Unauthorized Data Access	Lack of strict access control leading to potential data breaches and privacy violations.
<input checked="" type="checkbox"/>	Failure to Handle Right to Erasure	Non-compliance with GDPR's Right to be Forgotten, leading to legal risks.
<input checked="" type="checkbox"/>	Personal Data Leakage	ML models unintentionally exposing personal data during training or inference.
<input checked="" type="checkbox"/>	Lack of Data Minimization	Collecting and storing excessive personal data, violating GDPR principles.
<input checked="" type="checkbox"/>	Failure in Data Anonymization & Pseudonymization	Inadequate de-identification techniques lead to re-identifiable personal data.
<input checked="" type="checkbox"/>	Non-Transparent User Consent Management	Lack of clear user consent tracking for data collection and processing.
<input checked="" type="checkbox"/>	Cross-border Data Transfer Violations	Failure to comply with international data transfer rules under GDPR/Schrems II.
<input checked="" type="checkbox"/>	Functional Safety	
<input checked="" type="checkbox"/>	Data Integrity Failure	Corrupt or inconsistent data leading to incorrect model predictions, endangering safety-critical systems.
<input checked="" type="checkbox"/>	Lack of Real-Time Processing	Failure to meet real-time inference requirements in autonomous or safety-critical applications.
<input checked="" type="checkbox"/>	Failure Mode Undetection	Inability to detect and mitigate failure conditions in deployed AI models, leading to high-risk scenarios.
<input checked="" type="checkbox"/>	Inconsistent Model Reproducibility	Non-deterministic model outputs lead to unpredictable AI system behavior.
<input checked="" type="checkbox"/>	Insufficient Edge Case Handling	AI models fail to correctly handle rare but critical safety scenarios, leading to unreliable operation.
<input checked="" type="checkbox"/>	Undetected Model Performance Drift	Model performance degrades over time without triggering corrective actions, leading to unsafe decisions.
<input checked="" type="checkbox"/>	Insufficient Redundancy & Failover Mechanisms	Lack of backup AI models or failover strategies in critical systems.
<input checked="" type="checkbox"/>	Inadequate Error Handling & Logging	Poor logging and monitoring hinder failure analysis and root cause identification.
<input checked="" type="checkbox"/>	Model Instability & Poor Generalization	Model may perform unpredictably or fail to generalize to unseen conditions, undermining functional safety.

TensorFlow Serving
TensorRT
Weights & Biases
WhyLabs
XGBoost
tsfresh

Settings filename (e.g. setti)



AssessML: Auswahlbereich

2. Weg: Werkzeuge

Automatische Ausgabe der zugeordneten Phasen und Aufgaben

Choose Phase(s): Choose Task(s): Choose Tool(s):

Failure Mode Selection	Task Assessment	Tool Assessment	Assessment Summary
------------------------	-----------------	-----------------	--------------------

Failure Mode and Compliance Area Selection

Select	Failure Mode	Description
<input checked="" type="checkbox"/>	AI Regulation	
<input checked="" type="checkbox"/>	Lack of AI Transparency	Failure to provide explanations for AI decisions, violating AI Act transparency requirements.
<input checked="" type="checkbox"/>	Bias & Fairness Non-Compliance	High-risk AI models reinforcing biases, leading to discrimination and non-compliance with AI Act fairness guidelines.
<input checked="" type="checkbox"/>	Data Protection	
<input checked="" type="checkbox"/>	Personal Data Leakage	ML models unintentionally exposing personal data during training or inference.
<input checked="" type="checkbox"/>	Functional Safety	
<input checked="" type="checkbox"/>	Lack of Real-Time Processing	Failure to meet real-time inference requirements in autonomous or safety-critical applications.
<input checked="" type="checkbox"/>	Inadequate Error Handling & Logging	Poor logging and monitoring hinder failure analysis and root cause identification.
<input checked="" type="checkbox"/>	Model Instability & Poor Generalization	Model may perform unpredictably or fail to generalize to unseen conditions, undermining functional safety.

Settings filename (e.g. setti)

Nachdem der Tab „Werkzeuge“ durchgegangen wird, zeigt das Dashboard alle Fehlerzustände an, die dem gewählten Werkzeug zugeordnet sind. → Es erfolgt eine spezifische Einschränkung der Fehlerzustände!

AssessML: Auswahlbereich

Hinweis

Auch eine Mehrfachauswahl ist möglich, dies gilt für **beide** Wege!

1. Weg: Phasen und Aufgaben

Werden mehrere Phasen und Aufgaben ausgewählt, werden die Fehlerzustände angezeigt, die für alle ausgewählten Phasen und Aufgaben relevant sind.

The screenshot shows the 'AssessML' interface with the following configuration:

- Choose Phase(s):** Data Curation, Deployment
- Choose Task(s):** Data Labeling & Annotation, Cloud Deployment, Data Preprocessing & Transformation

The interface displays a table with columns: Failure Mode Selection, Task Assessment, Tool Assessment, and Assessment Summary. Below the table is a 'Failure Mode and Compliance Area Selection' section with a table of failure modes:

Select	Failure Mode	Description
<input checked="" type="checkbox"/>	AI Regulation	
<input checked="" type="checkbox"/>	Lack of AI Transparency	Failure to provide explanations for AI decisions, violating AI Act transparency requirements.
<input checked="" type="checkbox"/>	Bias & Fairness Non-Compliance	High-risk AI models reinforcing biases, leading to discrimination and non-compliance with AI Act fairness guidelines.
<input checked="" type="checkbox"/>	Security Vulnerabilities in AI	Exposure to adversarial attacks compromising AI safety and trustworthiness.
<input checked="" type="checkbox"/>	Data Protection	
<input checked="" type="checkbox"/>	Unauthorized Data Access	Lack of strict access control leading to potential data breaches and privacy violations.
<input checked="" type="checkbox"/>	Personal Data Leakage	ML models unintentionally exposing personal data during training or inference.
<input checked="" type="checkbox"/>	Failure in Data Anonymization & Pseudonymization	Inadequate de-identification techniques lead to re-identifiable personal data.
<input checked="" type="checkbox"/>	Cross-border Data Transfer Violations	Failure to comply with international data transfer rules under GDPR, Schrems II.
<input checked="" type="checkbox"/>	Functional Safety	
<input checked="" type="checkbox"/>	Data Integrity Failure	Corrupt or inconsistent data leading to incorrect model predictions, endangering safety-critical systems.
<input checked="" type="checkbox"/>	Failure Mode Undetection	Inability to detect and mitigate failure conditions in deployed AI models, leading to high-risk scenarios.
<input checked="" type="checkbox"/>	Inconsistent Edge Case Handling	AI models fail to correctly handle rare but critical safety scenarios, leading to unreliable operation.
<input checked="" type="checkbox"/>	Inefficient Redundancy & Failover Mechanisms	Lack of backup AI models or failover strategies in critical systems.
<input checked="" type="checkbox"/>	Model Instability & Poor Generalization	Model may perform unpredictably or fail to generalize to unseen conditions, undermining functional safety.

Settings filename (e.g. test) | Save Settings | Load Settings

Es erfolgt eine Einschränkung der möglichen Fehlerzustände!

2. Weg: Werkzeuge

Werden mehrere Werkzeuge ausgewählt, werden die Fehlerzustände angezeigt, die für alle ausgewählten Werkzeuge relevant sind.

The screenshot shows the 'AssessML' interface with the following configuration:

- Choose Phase(s):** Data Curation, Training & Validation
- Choose Task(s):** Data Collection & Integration, Data Labeling & Annotation, Deep Learning & ML Frameworks
- Choose Tool(s):** XGBoost, Airbyte, Amazon SageMaker Ground Truth

The interface displays a table with columns: Failure Mode Selection, Task Assessment, Tool Assessment, and Assessment Summary. Below the table is a 'Failure Mode and Compliance Area Selection' section with a table of failure modes:

Select	Failure Mode	Description
<input checked="" type="checkbox"/>	AI Regulation	
<input checked="" type="checkbox"/>	Lack of AI Transparency	Failure to provide explanations for AI decisions, violating AI Act transparency requirements.
<input checked="" type="checkbox"/>	Bias & Fairness Non-Compliance	High-risk AI models reinforcing biases, leading to discrimination and non-compliance with AI Act fairness guidelines.
<input checked="" type="checkbox"/>	Data Protection	
<input checked="" type="checkbox"/>	Unauthorized Data Access	Lack of strict access control leading to potential data breaches and privacy violations.
<input checked="" type="checkbox"/>	Personal Data Leakage	ML models unintentionally exposing personal data during training or inference.
<input checked="" type="checkbox"/>	Lack of Data Minimization	Collecting and storing excessive personal data, violating GDPR principles.
<input checked="" type="checkbox"/>	Non-Transparent User Consent Management	Lack of clear user consent tracking for data collection and processing.
<input checked="" type="checkbox"/>	Functional Safety	
<input checked="" type="checkbox"/>	Lack of Real-Time Processing	Failure to meet real-time inference requirements in autonomous or safety-critical applications.
<input checked="" type="checkbox"/>	Failure Mode Undetection	Inability to detect and mitigate failure conditions in deployed AI models, leading to high-risk scenarios.
<input checked="" type="checkbox"/>	Inconsistent Model Reproducibility	Non-deterministic model outputs lead to unpredictable AI system behavior.
<input checked="" type="checkbox"/>	Insufficient Edge Case Handling	AI models fail to correctly handle rare but critical safety scenarios, leading to unreliable operation.
<input checked="" type="checkbox"/>	Inadequate Error Handling & Logging	Poor logging and monitoring hinder failure analysis and root cause identification.
<input checked="" type="checkbox"/>	Model Instability & Poor Generalization	Model may perform unpredictably or fail to generalize to unseen conditions, undermining functional safety.

Settings filename (e.g. test) | Save Settings | Load Settings

Es erfolgt eine spezifische Einschränkung der Fehlerzustände!

1. Ermittlung von Risiken in den Entwicklungsaufgaben

1.2 Auswahl der relevanten Konformitätsbereiche und Fehlerzustände

1. Ermittlung von Risiken in den Entwicklungsaufgaben (nach FMEA)

$$\text{RPN} = \text{Severity} * \text{Occurrence} * \text{Detection}$$

1.1 Auswahl der relevanten Aufgaben und Phasen

1.2 Auswahl der relevanten Konformitätsbereiche und Fehlerzustände

1.3 Berechnung des RPN für ausgesuchte Fehlerzustände in einer Aufgabe

AssessML: Auswahl der Fehlerzustände

Aufbau



Failure Mode and Compliance Area Selection

Select	Failure Mode	Description
<input checked="" type="checkbox"/>	AI Regulation	
<input checked="" type="checkbox"/>	Lack of AI Transparency	Failure to provide explanations for AI decisions, violating AI Act transparency requirements.
<input checked="" type="checkbox"/>	Bias & Fairness Non-Compliance	High-risk AI models reinforcing biases, leading to discrimination and non-compliance with AI Act fairness guidelines.
<input checked="" type="checkbox"/>	Lack of Model Risk Assessment	Failure to categorize AI models under AI Act risk levels and document safety concerns.
<input checked="" type="checkbox"/>	Security Vulnerabilities in AI	Exposure to adversarial attacks compromising AI safety and trustworthiness.
<input checked="" type="checkbox"/>	Non-compliant Human Oversight Mechanisms	Failure to implement necessary human-in-the-loop controls in high-risk AI applications.
<input checked="" type="checkbox"/>	Insufficient AI Impact Assessment	Failure to conduct proper risk-benefit analysis for high-risk AI applications.
<input checked="" type="checkbox"/>	Lack of Ethical Safeguards in AI	Failure to ensure ethical principles (e.g., non-harm, fairness) in AI decision-making.
<input checked="" type="checkbox"/>	Inadequate Monitoring of Deployed AI Systems	Lack of continuous assessment of AI systems in real-world settings.
<input checked="" type="checkbox"/>	Lack of Auditability & Documentation	Inability to provide detailed logs of data usage, model training, validation etc. hindering compliance audits.
<input checked="" type="checkbox"/>	Data Protection	
<input checked="" type="checkbox"/>	Unauthorized Data Access	Lack of strict access control leading to potential data breaches and privacy violations.
<input checked="" type="checkbox"/>	Failure to Handle Right to Erasure	Non-compliance with GDPR's Right to be Forgotten, leading to legal risks.
<input checked="" type="checkbox"/>	Personal Data Leakage	ML models unintentionally exposing personal data during training or inference.
<input checked="" type="checkbox"/>	Lack of Data Minimization	Collecting and storing excessive personal data, violating GDPR principles.
<input checked="" type="checkbox"/>	Failure in Data Anonymization & Pseudonymization	Inadequate de-identification techniques lead to re-identifiable personal data.
<input checked="" type="checkbox"/>	Non-Transparent User Consent Management	Lack of clear user consent tracking for data collection and processing.
<input checked="" type="checkbox"/>	Cross-border Data Transfer Violations	Failure to comply with international data transfer rules under GDPR/Schrems II.
<input checked="" type="checkbox"/>	Functional Safety	
<input checked="" type="checkbox"/>	Data Integrity Failure	Corrupt or inconsistent data leading to incorrect model predictions, endangering safety-critical systems.
<input checked="" type="checkbox"/>	Lack of Real-Time Processing	Failure to meet real-time inference requirements in autonomous or safety-critical applications.
<input checked="" type="checkbox"/>	Failure Mode Undetection	Inability to detect and mitigate failure conditions in deployed AI models, leading to high-risk scenarios.
<input checked="" type="checkbox"/>	Inconsistent Model Reproducibility	Non-deterministic model outputs lead to unpredictable AI system behavior.
<input checked="" type="checkbox"/>	Insufficient Edge Case Handling	AI models fail to correctly handle rare but critical safety scenarios, leading to unreliable operation.
<input checked="" type="checkbox"/>	Undetected Model Performance Drift	Model performance degrades over time without triggering corrective actions, leading to unsafe decisions.
<input checked="" type="checkbox"/>	Insufficient Redundancy & Failover Mechanisms	Lack of backup AI models or failover strategies in critical systems.
<input checked="" type="checkbox"/>	Inadequate Error Handling & Logging	Poor logging and monitoring hinder failure analysis and root cause identification.
<input checked="" type="checkbox"/>	Model Instability & Poor Generalization	Model may perform unpredictably or fail to generalize to unseen conditions, undermining functional safety.

Der Bereich **Bewertung** unterteilt sich in die vier Slides:

- **Failure Mode Selection** (Auflistung der Fehlerzustände),
- **Task Assessment** (Ermittlung von Risiken in den Entwicklungsaufgaben),
- **Tool Assessment** (Qualifizierungsbedarfs für KI-Werkzeuge) und
- **Assessment Summary** (Qualifizierungszusammenfassung auf Basis des RPN und des TCL)

AssessML: Auswahl der Fehlerzustände

Feinauswahl - Failure Mode Overview

Failure Mode Selection	Task Assessment	Tool Assessment	Assessment Summary
------------------------	-----------------	-----------------	--------------------

Failure Mode and Compliance Area Selection

Select	Failure Mode	Description
<input checked="" type="checkbox"/>	AI Regulation	
<input checked="" type="checkbox"/>	Lack of AI Transparency	Failure to provide explanations for AI decisions, violating AI Act transparency requirements.
<input checked="" type="checkbox"/>	Bias & Fairness Non-Compliance	High-risk AI models reinforcing biases, leading to discrimination and non-compliance with AI Act fairness guidelines.
<input checked="" type="checkbox"/>	Lack of Model Risk Assessment	Failure to categorize AI models under AI Act risk levels and document safety concerns.
<input checked="" type="checkbox"/>	Security Vulnerabilities in AI	Exposure to adversarial attacks compromising AI safety and trustworthiness.
<input checked="" type="checkbox"/>	Non-compliant Human Oversight Mechanisms	Failure to implement necessary human-in-the-loop controls in high-risk AI applications.
<input checked="" type="checkbox"/>	Insufficient AI Impact Assessment	Failure to conduct proper risk-benefit analysis for high-risk AI applications.
<input checked="" type="checkbox"/>	Lack of Ethical Safeguards in AI	Failure to ensure ethical principles (e.g., non-harm, fairness) in AI decision-making.
<input checked="" type="checkbox"/>	Inadequate Monitoring of Deployed AI Systems	Lack of continuous assessment of AI systems in real-world settings.
<input checked="" type="checkbox"/>	Lack of Auditability & Documentation	Inability to provide detailed logs of data usage, model training, validation etc. hindering compliance audits.
<input checked="" type="checkbox"/>	Data Protection	
<input checked="" type="checkbox"/>	Unauthorized Data Access	Lack of strict access control leading to potential data breaches and privacy violations.
<input checked="" type="checkbox"/>	Failure to Handle Right to Erasure	Non-compliance with GDPR's Right to be Forgotten, leading to legal risks.
<input checked="" type="checkbox"/>	Personal Data Leakage	ML models unintentionally exposing personal data during training or inference.
<input checked="" type="checkbox"/>	Lack of Data Minimization	Collecting and storing excessive personal data, violating GDPR principles.
<input checked="" type="checkbox"/>	Failure in Data Anonymization & Pseudonymization	Inadequate de-identification techniques lead to re-identifiable personal data.
<input checked="" type="checkbox"/>	Non-Transparent User Consent Management	Lack of clear user consent tracking for data collection and processing.
<input checked="" type="checkbox"/>	Cross-border Data Transfer Violations	Failure to comply with international data transfer rules under GDPR/Schrems II.
<input checked="" type="checkbox"/>	Functional Safety	
<input checked="" type="checkbox"/>	Data Integrity Failure	Corrupt or inconsistent data leading to incorrect model predictions, endangering safety-critical systems.
<input checked="" type="checkbox"/>	Lack of Real-Time Processing	Failure to meet real-time inference requirements in autonomous or safety-critical applications.
<input checked="" type="checkbox"/>	Failure Mode Undetection	Inability to detect and mitigate failure conditions in deployed AI models, leading to high-risk scenarios.
<input checked="" type="checkbox"/>	Inconsistent Model Reproducibility	Non-deterministic model outputs lead to unpredictable AI system behavior.
<input checked="" type="checkbox"/>	Insufficient Edge Case Handling	AI models fail to correctly handle rare but critical safety scenarios, leading to unreliable operation.
<input checked="" type="checkbox"/>	Undetected Model Performance Drift	Model performance degrades over time without triggering corrective actions, leading to unsafe decisions.
<input checked="" type="checkbox"/>	Insufficient Redundancy & Failover Mechanisms	Lack of backup AI models or failover strategies in critical systems.
<input checked="" type="checkbox"/>	Inadequate Error Handling & Logging	Poor logging and monitoring hinder failure analysis and root cause identification.
<input checked="" type="checkbox"/>	Model Instability & Poor Generalization	Model may perform unpredictably or fail to generalize to unseen conditions, undermining functional safety.

Der Bereich **Failure Mode Selection** zeigt:

- Auflistung aller Fehlerzustände,
- Erklärung für jeden Fehlerzustand

und dient zur Auswahl der gesuchten/ benötigten Fehlerzustände der drei Kompatibilitätsgebiete.

AssessML: Auswahl der Fehlerzustände

Feinauswahl - Failure Mode Overview

Zur detaillierten Suche kann der Nutzer mit Hilfe der „Select-Kästchen“ weitere Filterungen vornehmen

- Nimmt man eine Auswahl für die einzelnen Fehlerzustände vor, werden die nicht weiter berücksichtigten Modi **weiß** hinterlegt.
- Nimmt man eine Auswahl für einen gesamten Kompatibilitätsgebiet vor, wird das nicht weiter berücksichtigte Gebiet **grau** hinterlegt.

Auswahl der Fehlerzustände über an-/abklicken der „Select-Kästchen“ je Fehlerzustände und/oder je Kompatibilitätsgebiete

Select	Failure Mode	Select	Failure Mode
<input checked="" type="checkbox"/>	AI Regulation	<input type="checkbox"/>	AI Regulation
<input checked="" type="checkbox"/>	Lack of AI Transparency	<input checked="" type="checkbox"/>	Lack of AI Transparency
<input checked="" type="checkbox"/>	Bias & Fairness Non-Compliance	<input checked="" type="checkbox"/>	Bias & Fairness Non-Compliance
<input checked="" type="checkbox"/>	Lack of Model Risk Assessment	<input checked="" type="checkbox"/>	Lack of Model Risk Assessment
<input checked="" type="checkbox"/>	Security Vulnerabilities in AI	<input checked="" type="checkbox"/>	Security Vulnerabilities in AI
<input type="checkbox"/>	Non-compliant Human Oversight Mechanism	<input checked="" type="checkbox"/>	Non-compliant Human Oversight Mechanisms
<input type="checkbox"/>	Insufficient AI Impact Assessment	<input checked="" type="checkbox"/>	Insufficient AI Impact Assessment
<input checked="" type="checkbox"/>	Lack of Ethical Safeguards in AI	<input checked="" type="checkbox"/>	Lack of Ethical Safeguards in AI
<input checked="" type="checkbox"/>	Inadequate Monitoring of Deployed AI System	<input checked="" type="checkbox"/>	Inadequate Monitoring of Deployed AI Systems
<input type="checkbox"/>	Lack of Auditability & Documentation	<input checked="" type="checkbox"/>	Lack of Auditability & Documentation
<input checked="" type="checkbox"/>	Data Protection	<input checked="" type="checkbox"/>	Data Protection
<input checked="" type="checkbox"/>	Unauthorized Data Access	<input checked="" type="checkbox"/>	Unauthorized Data Access
<input type="checkbox"/>	Failure to Handle Right to Erasure	<input checked="" type="checkbox"/>	Failure to Handle Right to Erasure
<input checked="" type="checkbox"/>	Personal Data Leakage	<input checked="" type="checkbox"/>	Personal Data Leakage
<input checked="" type="checkbox"/>	Lack of Data Minimization	<input checked="" type="checkbox"/>	Lack of Data Minimization
<input type="checkbox"/>	Failure in Data Anonymization & Pseudonym	<input checked="" type="checkbox"/>	Failure in Data Anonymization & Pseudonymization
<input type="checkbox"/>	Non-Transparent User Consent Management	<input checked="" type="checkbox"/>	Non-Transparent User Consent Management
<input checked="" type="checkbox"/>	Cross-border Data Transfer Violations	<input checked="" type="checkbox"/>	Cross-border Data Transfer Violations
<input checked="" type="checkbox"/>	Functional Safety	<input type="checkbox"/>	Functional Safety
<input checked="" type="checkbox"/>	Data Integrity Failure	<input checked="" type="checkbox"/>	Data Integrity Failure
<input checked="" type="checkbox"/>	Lack of Real-Time Processing	<input checked="" type="checkbox"/>	Lack of Real-Time Processing
<input type="checkbox"/>	Failure Mode Undetection	<input checked="" type="checkbox"/>	Failure Mode Undetection
<input checked="" type="checkbox"/>	Inconsistent Model Reproducibility	<input checked="" type="checkbox"/>	Inconsistent Model Reproducibility

1. Ermittlung von Risiken in den Entwicklungsaufgaben

1.3 Berechnung des RPN für ausgesuchte Fehlerzustände in einer Aufgabe

1. Ermittlung von Risiken in den Entwicklungsaufgaben (nach FMEA)

$$\text{RPN} = \text{Severity} * \text{Occurrence} * \text{Detection}$$

1.2 Auswahl der relevanten Aufgaben und Phasen

1.2 Auswahl der relevanten Konformitätsbereiche und Fehlerzustände

1.3 Berechnung des RPN für ausgesuchte Fehlerzustände in einer Aufgabe

AssessML: Bewertung der Fehlerzustände

Task Assessment - Berechnung des RPN für ausgesuchte Fehlerzustände

Der Bereich **Task Assessment** dient zur Berechnung der Risikoprioritätskennzahl (RPN) je ausgewähltem Fehlerzustand und/oder Kompatibilitätsgebiet.



Failure Mode Analysis for Individual ML Task

Phase: Training & Validation

Task: Deep Learning & ML Frameworks

Optimizing model parameters based on training data.

Artifact: Trained model

AI Regulation

Failure mode: Lack of AI Transparency

Impact: Failure to provide interpretability and documentation of model decisions.

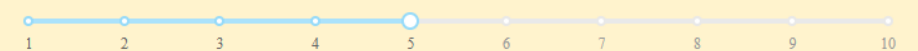
Potential Causes:

- No built-in interpretability or explainability methods.
- Poor model decision-tracking and logging support.
- Limited compatibility with external explainability libraries

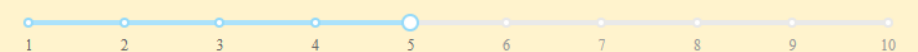
Possible Mitigations:

- Utilize ML frameworks that offer native support for model interpretation techniques like SHAP and LIME.
- Establish structured logging of intermediate decisions and model outputs for traceability.
- Ensure framework compatibility with transparency libraries via plugin architecture or open APIs.

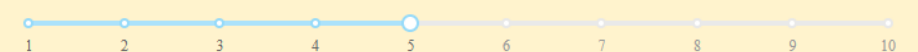
Severity of failure mode



Occurrence of failure mode



Detection of failure mode



RPN:

AssessML: Bewertung der Fehlerzustände

Task Assessment - Berechnung des RPN für ausgesuchte Fehlerzustände

Der **RPN** bildet die Summe der drei Kennziffern:

- **Severity** (Bedeutung der Fehlerfolge),
- **Occurrence** (Auftrittswahrscheinlichkeit der Fehlerursache) und
- **Detection** (Entdeckungswahrscheinlichkeit des Fehlers oder seiner Ursache)

$$\text{RPN} = \text{Severity} * \text{Occurrence} * \text{Detection}$$

Phase: Training & Validation

Task: Deep Learning & ML Frameworks

Optimizing model parameters based on training data.

Artifact: Trained model

Risk of Deep Learning & ML Frameworks in AI Regulation

Failure mode: Lack of AI Transparency: Failure to provide interpretability and documentation of model decisions.

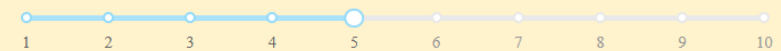
Potential Causes:

- No built-in interpretability or explainability methods.
- Poor model decision-tracking and logging support.
- Limited compatibility with external explainability libraries

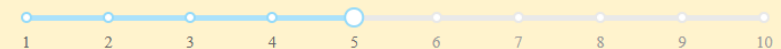
Possible Mitigations:

- Utilize ML frameworks that offer native support for model interpretation techniques like SHAP and LIME.
- Establish structured logging of intermediate decisions and model outputs for traceability.
- Ensure framework compatibility with transparency libraries via plugin architecture or open APIs.

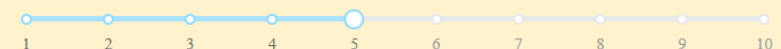
Severity of failure mode



Occurrence of failure mode



Detection of failure mode



RPN:

AssessML: Bewertung der Fehlerzustände

Bedeutung der Fehlerfolge (Severity)

Aktion:

Bewertung der **Schweregrades** für die Failure-Modes durchführen, die in dem Tool den ausgewählten ML-Aufgaben zugeordnet sind.

Phase: Training & Validation

Task: Deep Learning & ML Frameworks

Optimizing model parameters based on training data.

Artifact: Trained model

Risk of Deep Learning & ML Frameworks in AI Regulation

Failure mode: Lack of AI Transparency: Failure to provide interpretability and documentation of model decisions.

Potential Causes:

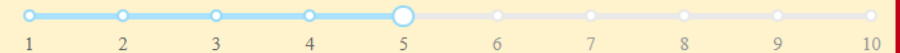
- No built-in interpretability or explainability methods.
- Poor model decision-tracking and logging support.
- Limited compatibility with external explainability libraries

Possible Mitigations:

- Utilize ML frameworks that offer native support for model interpretation techniques like SHAP and LIME.
- Establish structured logging of intermediate decisions and model outputs for traceability.
- Ensure framework compatibility with transparency libraries via plugin architecture or open APIs.

Durch das bewegen des Reglers, können die Werte 1 – 3; 4 – 6; 7 – 10 bestimmt werden.

Severity of failure mode



Occurrence of failure mode



Detection of failure mode



RPN:

AssessML: Bewertung der Fehlerzustände

Auftretenswahrscheinlichkeit der Fehlerursache (Occurence)

Aktion:

Bewertung der **Auftrittswahrscheinlichkeit** für die Failure-Modes durchführen, die in dem Tool den ausgewählten ML-Aufgaben zugeordnet sind.

Phase: Training & Validation

Task: Deep Learning & ML Frameworks

Optimizing model parameters based on training data.

Artifact: Trained model

Risk of Deep Learning & ML Frameworks in AI Regulation

Failure mode: Lack of AI Transparency: Failure to provide interpretability and documentation of model decisions.

Potential Causes:

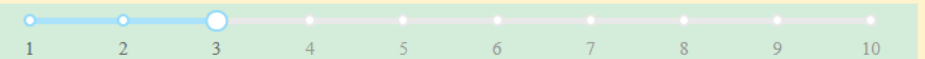
- No built-in interpretability or explainability methods.
- Poor model decision-tracking and logging support.
- Limited compatibility with external explainability libraries

Possible Mitigations:

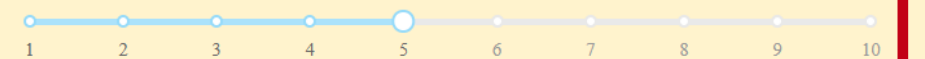
- Utilize ML frameworks that offer native support for model interpretation techniques like SHAP and LIME.
- Establish structured logging of intermediate decisions and model outputs for traceability.
- Ensure framework compatibility with transparency libraries via plugin architecture or open APIs.

Durch das bewegen des Reglers, können die Werte 1 – 3; 4 – 6; 7 – 10 bestimmt werden.

Severity of failure mode



Occurence of failure mode



Detection of failure mode



RPN:

AssessML: Bewertung der Fehlerzustände

Entdeckungswahrscheinlichkeit des Fehlers oder seiner Ursache (Detectability)

Aktion:

Bewertung der **Erkennbarkeit** für die Failure-Modes durchführen, die in dem Tool den ausgewählten ML-Aufgaben zugeordnet sind.

Phase: Training & Validation

Task: Deep Learning & ML Frameworks

Optimizing model parameters based on training data.

Artifact: Trained model

Risk of Deep Learning & ML Frameworks in AI Regulation

Failure mode: Lack of AI Transparency: Failure to provide interpretability and documentation of model decisions.

Potential Causes:

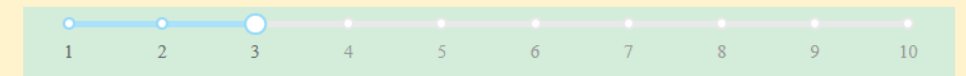
- No built-in interpretability or explainability methods.
- Poor model decision-tracking and logging support.
- Limited compatibility with external explainability libraries

Possible Mitigations:

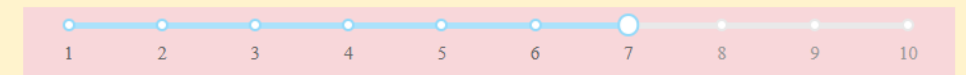
- Utilize ML frameworks that offer native support for model interpretation techniques like SHAP and LIME.
- Establish structured logging of intermediate decisions and model outputs for traceability.
- Ensure framework compatibility with transparency libraries via plugin architecture or open APIs.

Durch das bewegen des Reglers, können die Werte 1 – 3; 4 – 6; 7 – 10 bestimmt werden.

Severity of failure mode



Occurrence of failure mode



Detection of failure mode



RPN:

AssessML: Bewertung der Fehlerzustände

Auswertung des RPN

Aktion:

Berechnung des **RPN** für die Failure-Modes, die in dem Tool den ausgewählten ML-Aufgaben zugeordnet sind.

Phase: Training & Validation

Task: Deep Learning & ML Frameworks

Optimizing model parameters based on training data.

Artifact: Trained model

Risk of Deep Learning & ML Frameworks in AI Regulation

Failure mode: Lack of AI Transparency: Failure to provide interpretability and documentation of model decisions.

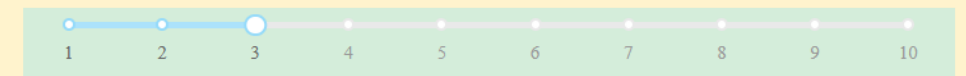
Potential Causes:

- No built-in interpretability or explainability methods.
- Poor model decision-tracking and logging support.
- Limited compatibility with external explainability libraries

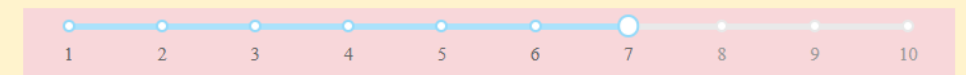
Possible Mitigations:

- Utilize ML frameworks that offer native support for model interpretation techniques like SHAP and LIME.
- Establish structured logging of intermediate decisions and model outputs for traceability.
- Ensure framework compatibility with transparency libraries via plugin architecture or open APIs.

Severity of failure mode



Occurrence of failure mode



Detection of failure mode



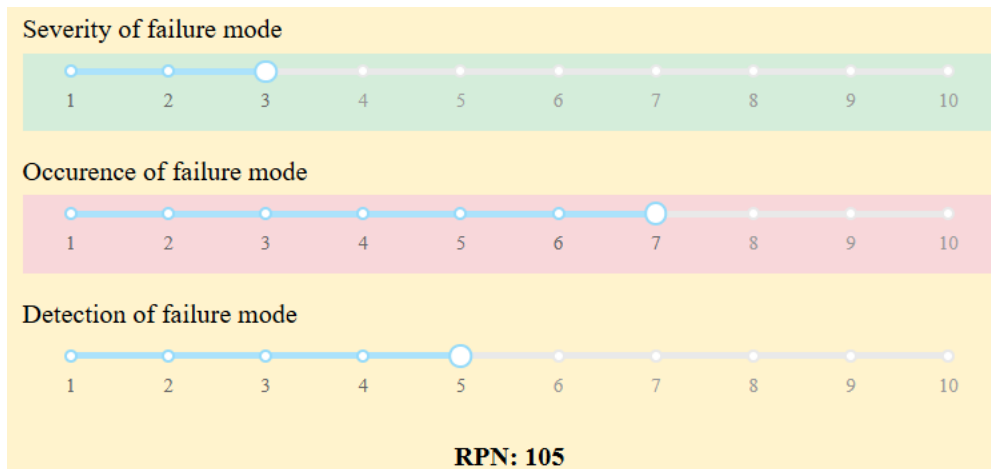
RPN: 105

Berechnung des RPN (1 – 29; 30 – 299; 300 -1000).

AssessML: Bewertung der Fehlerzustände

Task Assessment - Berechnung des RPN für ausgesuchte Fehlerzustände in einer Aufgabe

$$\text{RPN} = \text{Severity} * \text{Occurrence} * \text{Detection}$$



Auswertung der **Risikoprioritätskennzahl** (RPN):

- Niedrige RPN-Werte (RPN 1–30) weisen auf eine geringe Kritikalität hin und erfordern keine Maßnahmen.
- Mittlere RPN-Werte (RPN 31–299) weisen auf eine mittlere Kritikalität hin und erfordern daher Maßnahmen.
- Hohe Werte (RPN 300–1000) weisen auf eine hohe Kritikalität hin und erfordern umfangreichere Maßnahmen.

→ Die Ermittlung von Risiken (RPN) in den Entwicklungsaufgaben (nach FMEA) fokussiert sich auf die Ermittlung je Use-Case

2. Ermittlung des Qualifizierungsbedarfs für KI- Werkzeuge

2.1 Einschätzung über die Bedeutung des Werkzeugs für einen Fehler (Tool Impact)

2. Ermittlung des Qualifizierungsbedarfs für KI-Werkzeuge

2.1 Einschätzung über die Bedeutung des Werkzeugs für einen Fehler (Tool Impact)

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

2.3 Abschließende Werkzeugbewertung und Handlungsempfehlung auf Basis der RPN und des TCL

AssessML: Auswirkung eines Werkzeugfehlers

Tool Assessment - Ermittlung des Qualifizierungsbedarfs für KI-Werkzeuge

Der Bereich **Tool Assessment** dient zur Ermittlung des Qualifizierungsbedarfs für KI-Werkzeuge je ausgewähltem Fehlerzustand und/oder Kompatibilitätsgebiet.



Tool Impact and Tool Error Detection Analysis

Phase: Training & Validation

Task: Deep Learning & ML Frameworks

Optimizing model parameters based on training data.

Artifact: Trained model

AI Regulation

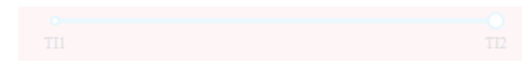
Tool Confidence Level:

Failure mode: Lack of AI Transparency

RPN: 105

Potential tool failure: May fail to generate or link detailed training documentation and validation reports, resulting in trained models lacking required transparency and conformity evidence.

Tool Impact Value



Tool Error Detection



AssessML: Auswirkung eines Werkzeugfehlers

Auswirkungsgrad des Werkzeuges



Tool Impact and Tool Error Detection Analysis

Phase: Training & Validation

Task: Deep Learning & ML Frameworks

Optimizing model parameters based on training data.

Artifact: Trained model

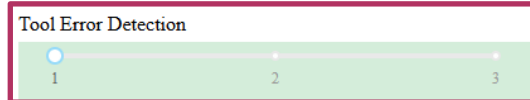
AI Regulation

Tool Confidence Level:

Failure mode: Lack of AI Transparency

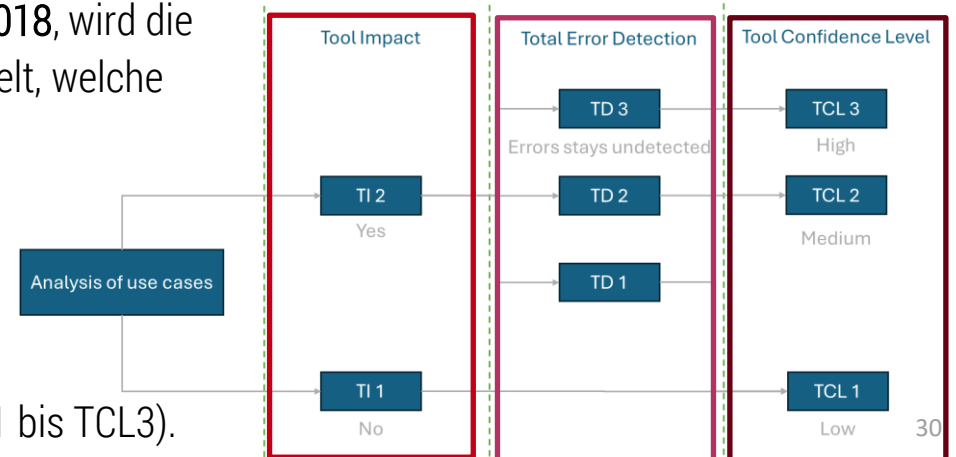
RPN: 105

Potential tool failure: May fail to generate or link detailed training documentation and validation reports, resulting in trained models lacking required transparency and conformity evidence.



In Anlehnung an das Kapitel 11.4.5.2 der ISO 26262:2018, wird die Bestimmung des **Tool Confidence Levels (TCL)** ermittelt, welche auf den drei folgenden Teilen basiert:

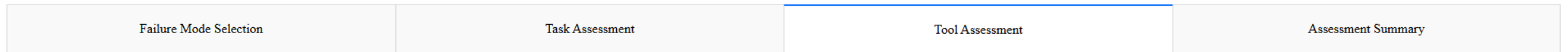
- **Tool Impact (TI)**, beurteilt, ob ein Fehler des Tools sicherheitsrelevant sein kann,
- **Tool Error Detection (TD)**, beurteilt die Wahrscheinlichkeit, dass ein Fehler entdeckt wird, und resultiert im
- **Tool Confidence Level (TCL)**, dem erforderlichen Maß an Vertrauen in das Tool (TCL1 bis TCL3).



AssessML: Auswirkung eines Werkzeugfehlers

Einschätzung über die Bedeutung des Werkzeugs für einen Fehler (Tool Impact)

Aktion: Hat ein potenzieller Fehler des Tools Einfluss auf die Funktionalität sicherheitsrelevanter Systeme?



Tool Impact and Tool Error Detection Analysis

Phase: Training & Validation

Task: Deep Learning & ML Frameworks

Optimizing model parameters based on training data.

Artifact: Trained model

AI Regulation

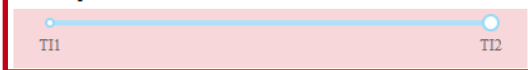
Tool Confidence Level:

Failure mode: Lack of AI Transparency

RPN: 105

Potential tool failure: May fail to generate or link detailed training documentation and conformity evidence.

Tool Impact Value



Tool Error Detection



Durch das Bewegen des Reglers erfolgt
Einschätzung über das Vorhandensein
von sicherheitsrelevanten Auswirkungen
(1 (nein) , 2 (ja))

2. Ermittlung des Qualifizierungsbedarfs für KI- Werkzeuge

2.2 Bewertung der Erkennbarkeit des Werkzeugfehlers (Tool Error Detection) und TCL

2. Ermittlung des Qualifizierungsbedarfs für KI-Werkzeuge

2.1 Einschätzung über die Bedeutung des Werkzeugs für einen Fehler (Tool Impact)

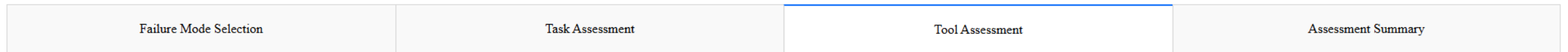
2.2 Bewertung der Erkennbarkeit des Werkzeugfehlers (Tool Error Detection) und Bestimmung des Tool Confidence Level (TCL) pro Konformitätsbereich

2.3 Abschließende Werkzeugbewertung und Handlungsempfehlung auf Basis der RPN und des TCL

AssessML: Erkennbarkeit des Werkzeugfehlers

Bewertung der Erkennbarkeit des Werkzeugfehlers (Tool Error Detection)

Aktion: Wie gut können Fehler, die durch das Tool entstehen, im weiteren Entwicklungsprozess erkannt werden?



Tool Impact and Tool Error Detection Analysis

Phase: Training & Validation

Task: Deep Learning & ML Frameworks

Optimizing model parameters based on training data.

Artifact: Trained model

AI Regulation

Tool Confidence Level:

Failure mode: Lack of AI Transparency

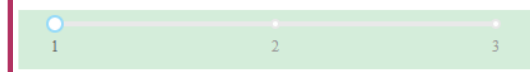
RPN: 105

Potential tool failure: May fail to generate or link detailed training documentation and validation reports, resulting in trained models lacking required transparency and conformity evidence.

Tool Impact Value



Tool Error Detection



Durch das bewegen des Reglers erfolgt Einschätzung über das Vorhandensein von sicherheitsrelevanten Auswirkungen
(1 = wird erkannt, 2 = könnten unentdeckt bleiben, 3 = bleiben unentdeckt)

Hintergrund: Berechnung TCL

Vertrauensbedarf für die Werkzeugnutzung

Ausgangslage: Wie hoch muss das Vertrauen in das Tool sein (bzw. ist eine Toolqualifikation notwendig)?

Table 3 — Determination of the Tool Confidence Level (TCL)

		Tool error detection		
		TD1	TD2	TD3
Tool impact	TI1	TCL1	TCL1	TCL1
	TI2	TCL1	TCL2	TCL3

Quelle: ISO 26262:2018, Kap. 11.4.5.4



Aspekt	Bild	Textbeschreibung
Ziel	<i>Zuordnung des passenden Tool Confidence Levels (TCL) auf Basis von Tool Impact (TI) und Tool Error Detection (TD)</i>	<i>Festlegen, wie hoch das Vertrauen in das Tool sein muss (bzw. ob eine Toolqualifikation notwendig ist)</i>
TCL1 (niedrig)	<ul style="list-style-type: none"> Tool hat keinen sicherheitsrelevanten Einfluss (TI1) oder Fehler werden sicher erkannt (TI2 + TD1) → Keine Toolqualifikation notwendig 	<ul style="list-style-type: none"> Geringe Bedeutung für Produktqualität Keine Toolqualifikation notwendig → Vertrauen in Toolverhalten aus Sicht ISO 26262 nicht kritisch
TCL2 (mittel)	<ul style="list-style-type: none"> Tool hat sicherheitsrelevanten Einfluss (TI2) Fehler nicht sicher erkennbar (TD2) 	<ul style="list-style-type: none"> Tool wichtig für Produktqualität Toolqualifikation erforderlich Anforderungen abhängig vom ASIL-Level
TCL3 (hoch)	<ul style="list-style-type: none"> Tool hat sicherheitsrelevanten Einfluss (TI2) Fehler bleiben unentdeckt (TD3) 	<ul style="list-style-type: none"> Hohe Bedeutung für Produktqualität Toolqualifikation zwingend nötig, hohe Anforderungen Unterschiede zwischen TCL2 und TCL3 sind methodisch relevant, aber nicht dramatisch (je nach ASIL)

Hintergrund: Berechnung TCL

Vertrauensbedarf für die Werkzeugnutzung

Ausgangslage: Wie hoch muss das Vertrauen in das Tool sein (bzw. ist eine Toolqualifikation notwendig)?

Table 3 — Determination of the Tool Confidence Level (TCL)

		Tool error detection		
		TD1	TD2	TD3
Tool impact	T11	TCL1	TCL1	TCL1
	T12	TCL1	TCL2	TCL3

Quelle: ISO 26262:2018, Kap. 11.4.5.4



Aspekt	Bild	Textbeschreibung
Ziel	<i>Zuordnung des passenden Tool Confidence Levels (TCL) auf Basis von Tool Impact (TI) und Tool Error Detection (TD)</i>	<i>Festlegen, wie hoch das Vertrauen in das Tool sein muss (bzw. ob eine Toolqualifikation notwendig ist)</i>
TCL1 (niedrig)	<ul style="list-style-type: none"> Tool hat keinen sicherheitsrelevanten Einfluss (T11) oder Fehler werden sicher erkannt (T12 + TD1) <p>→ Keine Toolqualifikation notwendig</p>	<ul style="list-style-type: none"> Geringe Bedeutung für Produktqualität Keine Toolqualifikation notwendig <p>→ Vertrauen in Toolverhalten aus Sicht ISO 26262 nicht kritisch</p>
TCL2 (mittel)	<ul style="list-style-type: none"> Tool hat sicherheitsrelevanten Einfluss (T12) Fehler nicht sicher erkennbar (TD2) 	<ul style="list-style-type: none"> Tool wichtig für Produktqualität Toolqualifikation erforderlich Anforderungen abhängig vom ASIL-Level
TCL3 (hoch)	<ul style="list-style-type: none"> Tool hat sicherheitsrelevanten Einfluss (T12) Fehler bleiben unentdeckt (TD3) 	<ul style="list-style-type: none"> Hohe Bedeutung für Produktqualität Toolqualifikation zwingend nötig, hohe Anforderungen Unterschiede zwischen TCL2 und TCL3 sind methodisch relevant, aber nicht dramatisch (je nach ASIL)

Hintergrund: Berechnung TCL

Vertrauensbedarf für die Werkzeugnutzung

Ausgangslage: Wie hoch muss das Vertrauen in das Tool sein (bzw. ist eine Toolqualifikation notwendig)?

Table 3 — Determination of the Tool Confidence Level (TCL)

		Tool error detection		
		TD1	TD2	TD3
Tool impact	TI1	TCL1	TCL1	TCL1
	TI2	TCL1	TCL2	TCL3

Quelle: ISO 26262:2018, Kap. 11.4.5.4



Aspekt	Bild	Textbeschreibung
Ziel	<i>Zuordnung des passenden Tool Confidence Levels (TCL) auf Basis von Tool Impact (TI) und Tool Error Detection (TD)</i>	<i>Festlegen, wie hoch das Vertrauen in das Tool sein muss (bzw. ob eine Toolqualifikation notwendig ist)</i>
TCL1 (niedrig)	<ul style="list-style-type: none"> Tool hat keinen sicherheitsrelevanten Einfluss (TI1) oder Fehler werden sicher erkannt (TI2 + TD1) → Keine Toolqualifikation notwendig	<ul style="list-style-type: none"> Geringe Bedeutung für Produktqualität Keine Toolqualifikation notwendig → Vertrauen in Toolverhalten aus Sicht ISO 26262 nicht kritisch
TCL2 (mittel)	<ul style="list-style-type: none"> Tool hat sicherheitsrelevanten Einfluss (TI2) Fehler nicht sicher erkennbar (TD2) 	<ul style="list-style-type: none"> Tool wichtig für Produktqualität Toolqualifikation erforderlich Anforderungen abhängig vom ASIL-Level
TCL3 (hoch)	<ul style="list-style-type: none"> Tool hat sicherheitsrelevanten Einfluss (TI2) Fehler bleiben unentdeckt (TD3) 	<ul style="list-style-type: none"> Hohe Bedeutung für Produktqualität Toolqualifikation zwingend nötig, hohe Anforderungen Unterschiede zwischen TCL2 und TCL3 sind methodisch relevant, aber nicht dramatisch (je nach ASIL)

Hintergrund: Berechnung TCL

Vertrauensbedarf für die Werkzeugnutzung

Ausgangslage: Wie hoch muss das Vertrauen in das Tool sein (bzw. ist eine Toolqualifikation notwendig)?

Table 3 — Determination of the Tool Confidence Level (TCL)

		Tool error detection		
		TD1	TD2	TD3
Tool impact	TI1	TCL1	TCL1	TCL1
	TI2	TCL1	TCL2	TCL3

Quelle: ISO 26262:2018, Kap. 11.4.5.4

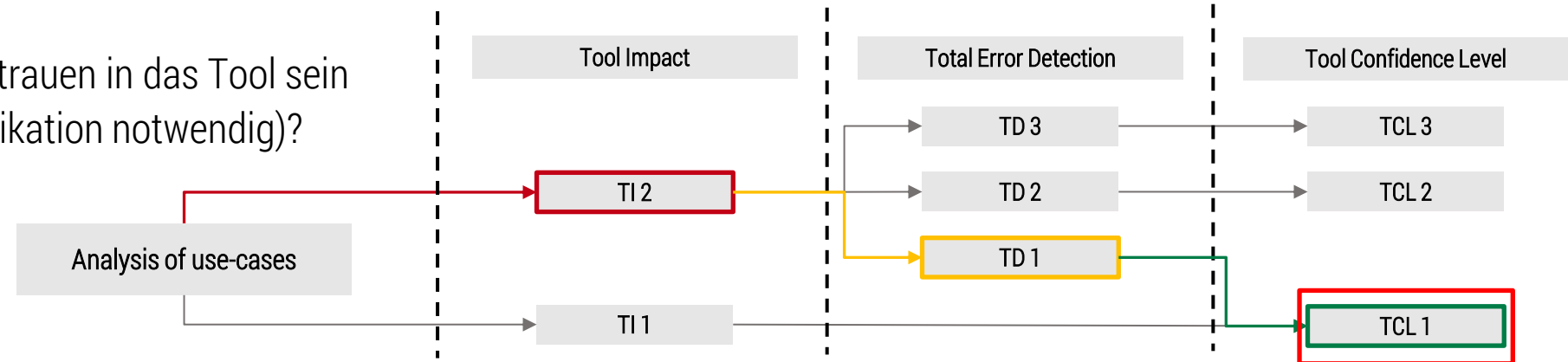


Aspekt	Bild	Textbeschreibung
Ziel	<i>Zuordnung des passenden Tool Confidence Levels (TCL) auf Basis von Tool Impact (TI) und Tool Error Detection (TD)</i>	<i>Festlegen, wie hoch das Vertrauen in das Tool sein muss (bzw. ob eine Toolqualifikation notwendig ist)</i>
TCL1 (niedrig)	<ul style="list-style-type: none"> Tool hat keinen sicherheitsrelevanten Einfluss (TI1) oder Fehler werden sicher erkannt (TI2 + TD1) <p>→ Keine Toolqualifikation notwendig</p>	<ul style="list-style-type: none"> Geringe Bedeutung für Produktqualität Keine Toolqualifikation notwendig <p>→ Vertrauen in Toolverhalten aus Sicht ISO 26262 nicht kritisch</p>
TCL2 (mittel)	<ul style="list-style-type: none"> Tool hat sicherheitsrelevanten Einfluss (TI2) Fehler nicht sicher erkennbar (TD2) 	<ul style="list-style-type: none"> Tool wichtig für Produktqualität Toolqualifikation erforderlich Anforderungen abhängig vom ASIL-Level
TCL3 (hoch)	<ul style="list-style-type: none"> Tool hat sicherheitsrelevanten Einfluss (TI2) Fehler bleiben unentdeckt (TD3) 	<ul style="list-style-type: none"> Hohe Bedeutung für Produktqualität Toolqualifikation zwingend nötig, hohe Anforderungen Unterschiede zwischen TCL2 und TCL3 sind methodisch relevant, aber nicht dramatisch (je nach ASIL)

AssessML: Berechnung TCL

Bestimmung des Tool Confidence Level (TCL)

Aktion: Wie hoch muss das Vertrauen in das Tool sein (bzw. ist eine Toolqualifikation notwendig)?



Tool Impact and Tool Error Detection Analysis

Phase: Training & Validation

Task: Deep Learning & ML Frameworks

Optimizing model parameters based on training data.

Artifact: Trained model

AI Regulation

Tool Confidence Level: 1

Failure mode: Lack of AI Transparency

RPN: 105

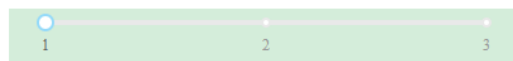
Potential tool failure: May fail to generate or link detailed training documentation and validation reports, resulting in trained models lacking required transparency and conformity evidence.

Berechnung des Vertrauenslevel
(1 = niedrig, 2 = mittel, 3 = hoch)

Tool Impact Value



Tool Error Detection



2. Ermittlung des Qualifizierungsbedarfs für KI- Werkzeuge

2.3 Werkzeugbewertung und Handlungsempfehlung auf Basis der RPN und des TCL

2. Ermittlung des Qualifizierungsbedarfs für KI-Werkzeuge

2.1 Einschätzung über die Bedeutung des Werkzeugs für einen Fehler (Tool Impact)

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

2.3 Abschließende Werkzeugbewertung und Handlungsempfehlung auf Basis der RPN und des TCL

AssessML: Bewertung und Handlungsempfehlung

Assessment Summery - Qualifizierungszusammenfassung

Der Bereich **Assessment Summery** dient zur Bewertung und Priorisierung der Qualifizierungszusammenfassung auf Basis des RPN und des TCL für ausgewählte Fehlerzustände.

Failure Mode Selection	Task Assessment	Tool Assessment	Assessment Summary
------------------------	-----------------	-----------------	---------------------------

Recommendations for Tool Qualification and Process Improvement

RPN threshold TCL threshold

Phase: Training & Validation

Task: Deep Learning & ML Frameworks

Optimizing model parameters based on training data.

Artifact: Trained model

AI Regulation

Tool Confidence Level: 1

Potential tool failure: Lack of AI Transparency

RPN: 105

Potential tool failure: May fail to generate or link detailed training documentation and validation reports, resulting in trained models lacking required transparency and conformity evidence.

Possible Process Mitigations:

- Utilize ML frameworks that offer native support for model interpretation techniques like SHAP and LIME.
- Establish structured logging of intermediate decisions and model outputs for traceability.
- Ensure framework compatibility with transparency libraries via plugin architecture or open APIs.

Testable Tool Capabilities:

- Integrated explainability methods (e.g., SHAP, LIME) within model training environments
- Comprehensive model decision logging and training metadata tracking
- Easy integration with external explainability libraries and visualization tools

AssessML: Bewertung und Handlungsempfehlung

Übersicht über die bestimmte Risikoprioritätskennzahl und das Werkzeugvertrauenslevel

Die Fehlerzustände können mit Hilfe der Eingaben für **Thresholds** (Schwellenwerte) gefiltert werden. Dabei können zwei verschiedene Angaben gewichtet werden:

- **RPN threshold** (Grenzwert für die Risikoprioritätskennzahl)
- **TCL threshold** (Grenzwert für das Werkzeugvertrauenslevel)

Failure Mode Selection	Task Assessment	Tool Assessment	Assessment Summary
------------------------	-----------------	-----------------	--------------------

Recommendations for Tool Qualification and Process Improvement

RPN threshold TCL threshold

Phase: Training & Validation

Task: Deep Learning & ML Frameworks

Optimizing model parameters based on training data.

Artifact: Trained model

AI Regulation

Tool Confidence Level: 1

Potential tool failure: Lack of AI Transparency

RPN: 105

Potential tool failure: May fail to generate or link detailed training documentation and validation reports, resulting in trained models lacking required transparency and conformity evidence.

Possible Process Mitigations:

- Utilize ML frameworks that offer native support for model interpretation techniques like SHAP and LIME.
- Establish structured logging of intermediate decisions and model outputs for traceability.
- Ensure framework compatibility with transparency libraries via plugin architecture or open APIs.

Testable Tool Capabilities:

- Integrated explainability methods (e.g., SHAP, LIME) within model training environments
- Comprehensive model decision logging and training metadata tracking
- Easy integration with external explainability libraries and visualization tools

AssessML: Bewertung und Handlungsempfehlung

Wozu dienen die Schwellenwerte (thresholds)?

Sämtliche Zustände werden automatisch in das Assessment Summary übernommen, wodurch das Board schnell unübersichtlich werden kann.

Mittels der Schwellenwerte können **kritische Fehlerzustände** oder **unsichere Tools** gezielt gefiltert werden, indem man diese passend einstellt. Durch das Einstellen der Filterung wird die Analyse übersichtlicher.

RPN threshold 1 TCL threshold 1

Failure Mode Selection Task Assessment Tool Assessment Assessment Summary

Recommendations for Tool Qualification and Process Improvement

RPN threshold 1 TCL threshold 1

Phase: Training & Validation

Task: Deep Learning & ML Frameworks
Optimizing model parameters based on training data.
Artifact: Trained model

Functional Safety

Tool Confidence Level: 2

Potential tool failure: Lack of Real-Time Processing
RPN: 544
Potential tool failure: May fail to guarantee deterministic training execution and runtime validation, resulting in trained models that do not meet real-time safety-critical performance requirements.

Possible Process Mitigations:

- Apply model optimization techniques such as pruning, quantization, or knowledge distillation.
- Ensure deployment-ready models are benchmarked and optimized for specific hardware accelerators.
- Optimize full pipeline latency by batching operations and reducing unnecessary transformations.

Testable Tool Capabilities:

- Framework support for model optimization (e.g., TensorRT, ONNX Runtime optimizations).
- Support for hardware-specific runtime optimizations and deployment targeting (e.g., EdgeTPU, CUDA kernels).
- Framework support for pipeline fusion, minimal preprocessing APIs, and asynchronous inference.

Tool Confidence Level: 3

Potential tool failure: Inadequate Error Handling & Logging
RPN: 30
Potential tool failure: May fail to provide comprehensive error logging and runtime monitoring, resulting in trained models with limited traceability for failure analysis in safety-critical systems.

Possible Process Mitigations:

- Integrate structured exception handling and mandatory logging for all critical training and inference operations.
- Implement custom monitors for numerical instabilities (e.g., NaN, divergence) and alert on thresholds.
- Embed pipeline-wide metadata capture and systematic experiment logging into the training framework.

Testable Tool Capabilities:

- Framework support for structured error reporting, custom callbacks, and logging integration.
- Built-in hooks for anomaly detection during training and validation (gradient checkers, divergence detectors).
- ML lifecycle management tools (e.g., MLFlow, Weights & Biases) integration for metadata and error tracking.

Tool Confidence Level: 1

Potential tool failure: Model Instability & Poor Generalization
RPN: 60
Potential tool failure: May fail to enforce deterministic training procedures and runtime validation, resulting in trained models with unpredictable behavior and poor generalization in safety-critical applications.

Possible Process Mitigations:

- Implement deterministic training pipelines with seed control and precision configuration to ensure repeatability.
- Integrate continuous performance monitoring during training with automatic validation checkpoints.
- Use frameworks that incorporate gradient clipping, normalization, and regularization for improved numerical robustness.

Testable Tool Capabilities:

- Explicit deterministic training support with reproducibility settings and controlled parallelism.
- Built-in validation dashboards and metric tracking APIs.
- Numerical stability modules with configurable regularization options.



RPN threshold 100 TCL threshold 2

Failure Mode Selection Task Assessment Tool Assessment Assessment Summary

Recommendations for Tool Qualification and Process Improvement

RPN threshold 100 TCL threshold 2

Phase: Training & Validation

Task: Deep Learning & ML Frameworks
Optimizing model parameters based on training data.
Artifact: Trained model

Functional Safety

Tool Confidence Level: 2

Potential tool failure: Lack of Real-Time Processing
RPN: 544
Potential tool failure: May fail to guarantee deterministic training execution and runtime validation, resulting in trained models that do not meet real-time safety-critical performance requirements.

Possible Process Mitigations:

- Apply model optimization techniques such as pruning, quantization, or knowledge distillation.
- Ensure deployment-ready models are benchmarked and optimized for specific hardware accelerators.
- Optimize full pipeline latency by batching operations and reducing unnecessary transformations.

Testable Tool Capabilities:

- Framework support for model optimization (e.g., TensorRT, ONNX Runtime optimizations).
- Support for hardware-specific runtime optimizations and deployment targeting (e.g., EdgeTPU, CUDA kernels).
- Framework support for pipeline fusion, minimal preprocessing APIs, and asynchronous inference.

Tool Confidence Level: 3

Potential tool failure: Inadequate Error Handling & Logging
RPN: 30
Potential tool failure: May fail to provide comprehensive error logging and runtime monitoring, resulting in trained models with limited traceability for failure analysis in safety-critical systems.

Possible Process Mitigations:

- Integrate structured exception handling and mandatory logging for all critical training and inference operations.
- Implement custom monitors for numerical instabilities (e.g., NaN, divergence) and alert on thresholds.
- Embed pipeline-wide metadata capture and systematic experiment logging into the training framework.

Testable Tool Capabilities:

- Framework support for structured error reporting, custom callbacks, and logging integration.
- Built-in hooks for anomaly detection during training and validation (gradient checkers, divergence detectors).
- ML lifecycle management tools (e.g., MLFlow, Weights & Biases) integration for metadata and error tracking.

Tool Confidence Level: 1

Potential tool failure: Model Instability & Poor Generalization
RPN: 60
Potential tool failure: May fail to enforce deterministic training procedures and runtime validation, resulting in trained models with unpredictable behavior and poor generalization in safety-critical applications.

Possible Process Mitigations:

- Implement deterministic training pipelines with seed control and precision configuration to ensure repeatability.
- Integrate continuous performance monitoring during training with automatic validation checkpoints.
- Use frameworks that incorporate gradient clipping, normalization, and regularization for improved numerical robustness.

Testable Tool Capabilities:

- Explicit deterministic training support with reproducibility settings and controlled parallelism.
- Built-in validation dashboards and metric tracking APIs.
- Numerical stability modules with configurable regularization options.

Settings (Rename (e.g. self)) Save Settings Load Settings

AssessML: Bewertung und Handlungsempfehlung

Übersicht über die bestimmte Risikoprioritätskennzahl und das Werkzeugvertrauenslevel

Die unten aufgeführten **Select-Kästchen (Checkboxes)** dienen dazu, konkrete Tool-Fähigkeiten oder mögliche Prozess-Maßnahmen auszuwählen:

Recommendations for Tool Qualification and Process Improvement

RPN threshold TCL threshold

Phase: Training & Validation

Task: Deep Learning & ML Frameworks

Optimizing model parameters based on training data.

Artifact: Trained model

AI Regulation

Tool Confidence Level: 1

Potential tool failure: Lack of AI Transparency

RPN: 105

Potential tool failure: May fail to generate or link detailed training documentation and validation reports, resulting in trained models lacking required transparency and conformity evidence.

Possible Process Mitigations:

- Utilize ML frameworks that offer native support for model interpretation techniques like SHAP and LIME.
- Establish structured logging of intermediate decisions and model outputs for traceability.
- Ensure framework compatibility with transparency libraries via plugin architecture or open APIs.

Testable Tool Capabilities:

- Integrated explainability methods (e.g., SHAP, LIME) within model training environments
- Comprehensive model decision logging and training metadata tracking
- Easy integration with external explainability libraries and visualization tools

Possible Process Mitigations (Mögliche Prozess-Maßnahmen):

Hier können Methoden oder Vorgehensweisen ausgewählt werden, die das Risiko eines bestimmten Fehlerzustands verringern sollen.

Testable Tool Capabilities (Überprüfbare Tool-Fähigkeiten):

Hier können konkrete Eigenschaften oder Funktionen von Tools markiert werden, die im Assessment getestet oder überprüft werden sollen.

AssessML: Export und Druck

Die Ergebnisse im Assessment Summary können durch drücken des Knopfes „Export PDF“ als PDF exportiert und ausgedruckt werden.

The screenshot displays the 'Assessment Summary' section of the AssessML tool. At the top, there are tabs for 'Failure Mode Selection', 'Task Assessment', 'Tool Assessment', and 'Assessment Summary'. Below the tabs, there are input fields for 'miflow' and 'Save Settings'. A red box highlights the 'Export PDF' button, with a red arrow pointing to the 'Drucken' (Print) dialog box. The 'Drucken' dialog shows the target 'prt-6058', 'Kopien' (Copies) set to 1, 'Ausrichtung' (Orientation) set to 'Hochformat' (Portrait), and 'Seiten' (Pages) set to 'Alle' (All). The 'Skalierung' (Scaling) is set to 'An Seitenbreite anpassen' (Fit to page width). The 'Drucken' button is highlighted in blue.

Recommendations for Tool Qualification and

Export PDF

RPN threshold 100 TCL threshold 1

Phase: Experimentation

Task: Experiment Tracking & Management
Organizing and maintaining data access, availability, and structure.

Artifact: Data storage system

Functional Safety

Tool Confidence Level: 2

Potential tool failure: Inconsistent Model Reproducibility
RPN: 315

Tool may fail to comprehensively capture or version all relevant experiment configurations and datasets, resulting in experiment tracking systems that hinder full reproducibility and traceability for safety-critical validations.

Possible Process Mitigations:

- Mandate seed fixing across frameworks and environment metadata capture for every run.
- Store all experiment artifacts, parameters, and outputs with unique, traceable version identifiers.
- Capture and lock dependency environments during each experiment run.

Testable Tool Capabilities:

- Experiment tracking platforms with random seed logging and environment snapshotting.
- Support for automatic artifact versioning and hyperparameter logging (e.g., Miflow, WandB).
- Environment capture and dependency management tools integrated with experiment tracking.

AI Regulation

Tool Confidence Level: 1

Potential tool failure: RPN: 210

Tool may fail to maintain categorizations, resulting in tracking systems that do not support required auditability and conformity assessments for high-risk AI systems.

Environment capture and dependency management tools integrated with experiment tracking.

Weitere Funktionen der AssessML

AssessML: Speichern und Laden

Warum soll man Speichern und Laden?

Mit der *Save and Load Settings* Option können getroffene Einstellungen und bereits durchgeführte Assessment-einstellungen gespeichert und bei Bedarf wieder geladen werden.

The screenshot displays the 'Initial Failure Mode and Compliance Area Selection' screen. It features a table with columns for 'Select', 'Failure Mode', and 'Description'. A settings overlay is positioned over the table, containing a 'Settings filename (e.g. setti)' input field, a 'Save Settings' button, and a 'Load Settings' button. The 'Model Instability & Poor Generalization' row is highlighted in green, and its 'Select' checkbox is checked.

Select	Failure Mode	Description
<input checked="" type="checkbox"/>	AI Regulation	
<input checked="" type="checkbox"/>	Lack of AI Transparency	Failure to provide explanations for AI decisions, violating AI Act transparency requirements.
<input checked="" type="checkbox"/>	Bias & Fairness Non-Compliance	High-risk AI models reinforcing biases, leading to discrimination and non-compliance with AI Act fairness guidelines.
<input checked="" type="checkbox"/>	Lack of Model Risk Assessment	Failure to categorize AI models under AI Act risk levels and document safety concerns.
<input checked="" type="checkbox"/>	Security Vulnerabilities in AI	Exposure to adversarial attacks compromising AI safety and trustworthiness.
<input checked="" type="checkbox"/>	Non-compliant Human Oversight Mechanisms	Failure to implement necessary human-in-the-loop controls in high-risk AI applications.
<input checked="" type="checkbox"/>	Insufficient AI Impact Assessment	Failure to conduct proper risk-benefit analysis for high-risk AI applications.
<input checked="" type="checkbox"/>	Lack of Ethical Safeguards in AI	Failure to ensure ethical principles (e.g., non-harm, fairness) in AI decision-making.
<input checked="" type="checkbox"/>	Inadequate Monitoring of Deployed AI Systems	Lack of continuous assessment of AI systems in real-world settings.
<input checked="" type="checkbox"/>	Lack of Auditability & Documentation	Inability to provide detailed logs of data usage, model training, validation etc. hindering compliance audits.
<input checked="" type="checkbox"/>	Data Protection	
<input checked="" type="checkbox"/>	Unauthorized Data Access	Lack of strict access control leading to potential data breaches and privacy violations.
<input checked="" type="checkbox"/>	Failure to Handle Right to Erasure	Non-compliance with GDPR's Right to be Forgotten, leading to legal risks.
<input checked="" type="checkbox"/>	Personal Data Leakage	ML models unintentionally exposing personal data during training or inference.
<input checked="" type="checkbox"/>	Lack of	
<input checked="" type="checkbox"/>	Failure	
<input checked="" type="checkbox"/>	Non-Tr	
<input checked="" type="checkbox"/>	Cross-b	
<input checked="" type="checkbox"/>	Function	
<input checked="" type="checkbox"/>	Data Int	
<input checked="" type="checkbox"/>	Lack of	
<input checked="" type="checkbox"/>	Failure Mode Undetection	Inability to detect and mitigate failure conditions in deployed AI models, leading to high-risk scenarios.
<input checked="" type="checkbox"/>	Inconsistent Model Reproducibility	Non-deterministic model outputs lead to unpredictable AI system behavior.
<input checked="" type="checkbox"/>	Insufficient Edge Case Handling	AI models fail to correctly handle rare but critical safety scenarios, leading to unreliable operation.
<input checked="" type="checkbox"/>	Undetected Model Performance Drift	Model performance degrades over time without triggering corrective actions, leading to unsafe decisions.
<input checked="" type="checkbox"/>	Insufficient Redundancy & Failover Mechanisms	Lack of backup AI models or failover strategies in critical systems.
<input checked="" type="checkbox"/>	Inadequate Error Handling & Logging	Poor logging and monitoring under failure analysis and root cause identification.
<input checked="" type="checkbox"/>	Model Instability & Poor Generalization	Model may perform unpredictably or fail to generalize to unseen conditions, undermining functional safety.

Settings filename (e.g. setti) Save Settings Load Settings

AssessML: Speichern und Laden

Erklärung des Speichervorgangs

Choose Phase(s): Choose Task(s): Choose Tool(s):

Failure Mode Selection	Task Assessment	Tool Assessment	Assessment Summary
------------------------	-----------------	-----------------	--------------------

Failure Mode and Compliance Area Selection

Select	Failure Mode	Description
<input checked="" type="checkbox"/>	AI Regulation	
<input checked="" type="checkbox"/>	Lack of AI Transparency	Failure to provide explanations for AI decisions, violating AI Act transparency requirements.
<input checked="" type="checkbox"/>	Bias & Fairness Non-Compliance	High-risk AI models reinforcing biases, leading to discrimination and non-compliance with AI Act fairness guidelines.
<input type="checkbox"/>	Data Protection	
<input checked="" type="checkbox"/>	Personal Data Leakage	ML models unintentionally exposing personal data during training or inference.
<input type="checkbox"/>	Functional Safety	
<input checked="" type="checkbox"/>	Lack of Real-Time Processing	Failure to meet real-time inference requirements in autonomous or safety-critical applications.
<input checked="" type="checkbox"/>	Inadequate Error Handling & Logging	Poor logging and monitoring hinder failure analysis and root cause identification.
<input checked="" type="checkbox"/>	Model Instability & Poor Generalization	Model may perform unpredictably or fail to generalize to unseen conditions, undermining functional safety.

Model instability & poor generalization model may pe

Bereits durchgeführte Such- und Assessmenteinstellungen, können durch „Save Settings“ abgespeichert werden, dazu:

1. Select-Kästchen auswählen
- ↓
2. Namen vergeben
- ↓
3. Mit Save Settings speichern

AssessML: Speichern und Laden

Erklärung des Speichervorgangs

Gespeicherte Settings, werden unten aufgelistet. Um gespeicherte Settings abzurufen:

Choose Phase(s):
Select...

Choose Task(s):
Select...

Failure Mode Selection	Task Assessment
Failure Mode and Compliance Area Selection	
Select	Description
<input checked="" type="checkbox"/> AI Regulation	
<input checked="" type="checkbox"/> Lack of AI Transparency	Failure to provide explanations for AI decisions, violating AI Act transparency requirements.
<input checked="" type="checkbox"/> Bias & Fairness Non-Compliance	High-risk AI models reinforcing biases, leading to discrimination and non-compliance with AI Act fairness guidelines.
<input checked="" type="checkbox"/> Lack of Model Risk Assessment	Failure to categorize AI models under AI Act risk levels and document safety concerns.
<input checked="" type="checkbox"/> Security Vulnerabilities in AI	Exposure to adversarial attacks compromising AI safety and trustworthiness.
<input checked="" type="checkbox"/> Non-compliant Human Oversight Mechanisms	Failure to implement necessary human-in-the-loop controls in high-risk AI applications.
<input checked="" type="checkbox"/> Insufficient AI Impact Assessment	Failure to conduct proper risk-benefit analysis for high-risk AI applications.
<input checked="" type="checkbox"/> Lack of Ethical Safeguards in AI	Failure to ensure ethical principles (e.g., non-harm, fairness) in AI decision-making.
<input checked="" type="checkbox"/> Inadequate Monitoring of Deployed AI Systems	Lack of continuous assessment of AI systems in real-world settings.
<input checked="" type="checkbox"/> Lack of Auditability & Documentation	Inability to provide detailed logs of data usage, model training, validation etc. hindering compliance audits.
<input checked="" type="checkbox"/> Data Protection	
<input checked="" type="checkbox"/> Unauthorized Data Access	Lack of strict access control leading to potential data breaches and privacy violations.
<input checked="" type="checkbox"/> Failure to Handle Right to Erasure	Non-compliance with GDPR's Right to be Forgotten, leading to legal risks.
<input checked="" type="checkbox"/> Personal Data Leakage	ML models unintentionally exposing personal data during training or inference.
<input checked="" type="checkbox"/> Lack of Data Minimization	Collecting and storing excessive personal data, violating GDPR principles.
<input checked="" type="checkbox"/> Failure in Data Anonymization & Pseudonymization	Inadequate de-identification techniques lead to re-identifiable personal data.
<input checked="" type="checkbox"/> Non-Transparent User Consent Management	Lack of clear user consent tracking for data collection and processing.
<input checked="" type="checkbox"/> Cross-border Data Transfer Violations	Failure to comply with international data transfer rules under GDPR/Schrems II.
<input checked="" type="checkbox"/> Functional Safety	
<input checked="" type="checkbox"/> Data Integrity Failure	Corrupt or inconsistent data leading to incorrect model predictions, endangering safety-critical systems.
<input checked="" type="checkbox"/> Lack of Real-Time Processing	Failure to meet real-time inference requirements in autonomous or safety-critical applications.
<input checked="" type="checkbox"/> Failure Mode Undetection	Inability to detect and mitigate failure conditions in deployed AI models, leading to high-risk scenarios.
<input checked="" type="checkbox"/> Inconsistent Model Reproducibility	Non-deterministic model outputs lead to unpredictable AI system behavior.
<input checked="" type="checkbox"/> Insufficient Edge Case Handling	AI models fail to correctly handle rare but critical safety scenarios, leading to unreliable operation.
<input checked="" type="checkbox"/> Undetected Model Performance Drift	Model performance degrades over time without triggering corrective actions, leading to unsafe decisions.
<input checked="" type="checkbox"/> Insufficient Redundancy & Failover Mechanisms	Lack of backup AI models or failover strategies in critical systems.
<input checked="" type="checkbox"/> Inadequate Error Handling & Logging	Poor logging and monitoring hinder failure analysis and root cause identification.
<input checked="" type="checkbox"/> Model Instability & Poor Generalization	Model may perform unpredictably or fail to generalize to unseen conditions, undermining functional safety.



1. Setting-Namen eingeben



2. Load Settings klicken



3. Dashboard lädt Konfiguration

Choose Phase(s):
x Training & Validation x

Choose Task(s):
x Deep Learning & ML Frameworks x

Choose Tool(s):
Select...

Failure Mode Selection	Task Assessment	Tool Assessment	Assessment Summary
Failure Mode and Compliance Area Selection			
Select	Description		
<input checked="" type="checkbox"/> AI Regulation			
<input checked="" type="checkbox"/> Lack of AI Transparency	Failure to provide explanations for AI decisions, violating AI Act transparency requirements.		
<input checked="" type="checkbox"/> Bias & Fairness Non-Compliance	High-risk AI models reinforcing biases, leading to discrimination and non-compliance with AI Act fairness guidelines.		
<input type="checkbox"/> Data Protection			
<input checked="" type="checkbox"/> Personal Data Leakage	ML models unintentionally exposing personal data during training or inference.		
<input type="checkbox"/> Functional Safety			
<input checked="" type="checkbox"/> Lack of Real-Time Processing	Failure to meet real-time inference requirements in autonomous or safety-critical applications.		
<input checked="" type="checkbox"/> Inadequate Error Handling & Logging	Poor logging and monitoring hinder failure analysis and root cause identification.		
<input checked="" type="checkbox"/> Model Instability & Poor Generalization	Model may perform unpredictably or fail to generalize to unseen conditions, undermining functional safety.		

Setting_1 Save Settings Load Settings

Setting_1 Save Settings Load Settings

Settings saved to Setting_1

Setting_1 Save Settings Load Settings

Settings saved to Setting_1

