Federal Office
for Information Security

Test Criteria Catalogue for AI Systems in Finance

Publications Notes

Given the international relevance of trustworthy AI in the financial sector and the widespread applicability of the EU AI Act across member states and beyond, this publication was prepared in English to ensure broader accessibility and facilitate collaboration with international stakeholders. English serves as the standard language in technical, regulatory, and academic discourse on AI, making it the most appropriate choice for addressing a diverse audience, including researchers, industry professionals, and policymakers across Europe and globally.

Document history

Version	*Date*	*Editor*	*Description*
1.0	27.05.2025	Henry Kleineidam	Provisional version
2.0	18.06.2026	Ulf Menzler	First complete version including lifecycle information and references to financial regulation.

Federal Office for Information Security
P.O. Box 20 03 63, 53133 Bonn, Germany
E-Mail: ki-kontakt@bsi.bund.de
Internet: https://www.bsi.bund.de

Table of Contents

I. Catalogue Structure

1 Introduction

The following criteria catalogue was developed as part of BSI Project 587 “Development and testing of test criteria, requirements and methods for AI systems in the financial sector”, in the following referred to as **“AICRIV Finanz”**.

One of the priorities of this project is the development of evaluation criteria and requirements for AI applications and systems in the financial sector. Building on existing approaches and information sources, e.g., standards, guidelines and regulations, a set of criteria for secure and trustworthy AI systems in finance was created and summarised in the AICRIV Finanz criteria catalogue.

The catalogue provides practical criteria for testing AI systems and suggests suitable test methods and tools for technical and document-based testing. Additionally, a process for applying the catalogue is described. The catalogue’s applicability was tested on real world use cases from the financial sector to ensure the testing criteria are generalisable, understandable and relevant and to demonstrate their practical feasibility. The insights gained from this, along with further project experiences, were incorporated into the final design of the criteria catalogue and the application process.

The focus is on developing specific testing approaches that can be applied to both the selected use cases and to AI models in finance in general. The evaluation criteria and requirements are motivated by specific threat scenarios and attack vectors and cover security-related risks that affect the trustworthy use of AI systems. In addition to traditional IT security, the catalogue also addresses AI-specific aspects such as data quality, functionality, security, transparency, and bias. The aim is to provide a comprehensive basis for evaluating and securing AI systems in the financial sector.

This document presents a detailed overview of the developed criteria catalogue and its application. Part I of this document describes the development of the catalogue, including the resources used and the process for deriving relevant content such as generic criteria, procedures, and testing tools. Moreover, the catalogue structure is explained in detail and the assignment of the selected criteria to the requirements of the EU AI Act (AI Regulation) is shown. Part I contains the AICRIV Finanz catalogue with the evaluation criteria.

2 Criteria Dimensions

The catalogue is thematically divided into 10 dimensions. Although there are thematic overlaps, each developed criterion is assigned to a single dimension:

- **AI Security & Robustness**: Robustness of the AI application against both natural and adversarial perturbations of the input, as well as the analysis and documentation of potential risks of the AI application.
- **IT Security**: Classical IT security of the overall system and its components, including access controls, firewalls, and security updates.
- **Monitoring**: Technical monitoring of the AI application, as well as monitoring the behaviour of the AI application to analyze performance and ensure security.
- **Performance**: Correct functionality and efficiency of the AI application, i.e. quality of output, as well as reproducibility and reliability.
- **Governance**: Structural Organizational framework conditions for the responsible and secure use of AI applications.
- **Human Oversight**: Human monitoring of the AI application and the involvement of individuals in the development and deployment of the AI application.
- **Fairness**: Ensuring that the operation of an AI application complies with specified ethical values and that there is sufficient diversity in the data.
- **Transparency**: Transparency in the decisions of the AI application, especially towards the user, but also other stakeholders. This includes disclosure of how the AI application operates, but also liability issues, instructions and performance indicators.
- **Data Quality & Management**: Ensuring data quality and responsible data management, as well as compliance with current data protection regulations.
- **Development**: Requirements for the development phase and corresponding documentation of the development of the AI application, ranging from architecture and training process to validation, testing, and deployment.

3 Example Entry

This chapter provides an example entry showing the structure of a criteria entry from the catalogue. The individual information fields within the entry contain brief descriptions of the corresponding content.

EXMP-01: Example Criterion

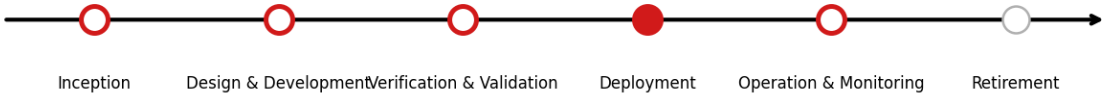


Figure 1: Example Lifecycle phase diagram

Evaluation requirement:

Requirement that needs to be fulfilled for compliance with this criteria to be achieved.

Evaluation method:

Test-based or Document-based: Instructions for the testing and evaluation of the criterion

Supportive guidance:

Supplementary Definitions:

Supplementary definitions: Explanation or definition of terminology

Exemplary [parameters]:

Examples or options for the variables in the []-brackets to be filled with

Additional guidance regarding the requirement:

Guidance and additional information that help with understanding of the criterion

Evaluation tools:

Suggestions for evaluation tools and methods for testing of the evaluation criteria

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

References to corresponding articles of the EU AI Act

References to DORA

References to DORA

Re-evaluation triggers:

Abbreviations for the triggers.

Each evaluation criterion is assigned a unique **Criterion ID** and a **Criterion name**. The ID consists of an abbreviation for the dimension the criterion is assigned to and a sequence number, e.g., TR-07 for the seventh criterion in the transparency dimension. The name of the criterion is another means of identifying the criterion and provides a first indication of its topic.

A **graphical illustration** depicts the relevant life cycle phases for the criterion. During the stated phases, the criterion shall be considered and prepared for the upcoming evaluation during the recommended life cycle stage (single phase).

The information given in the **Evaluation requirement** field specifically describes the test objective. The requirement is formulated generically, i.e., it may have to be adapted for the evaluation of a specific system. It does not contain any specific references to current methodological approaches, which helps to ensure that the requirements remain relevant in the long term. The evaluation requirement may contain parameters in square brackets that must be filled in based on the nature of the target of the evaluation and the evolving state of the art.

The **Evaluation method** gives specific instructions to the tester as to how the fulfillment of the evaluation requirement can be checked and informs the user whether the criteria is test based or document based.

The **Supportive guidance** contains definitions that are helpful for understanding the evaluation requirement and additional information on fulfilling the requirement. This could be, for example, recommendations on what specific information should be included in a particular documentation. These notes support the user of the catalogue in evaluating the criteria. In addition, common examples for the placeholders in the evaluation requirement can be found.

The **Evaluation tool** field supplements the evaluation method with relevant technical tools that can be used to check the evaluation requirement.

The Relevance based on use case **parametrization** states the relevant questions and answers of the use case questionnaire that determine whether the criterion should be applied. The questionnaire, along with the general approach in which it is applied, is presented in Chapter 5.

The **Related Regulatory References** fields contains thematically related requirements from the AI Regulation, DORA and other financial regulation and thus offers users a direct link between regulatory requirements and the AICRIV criteria catalogue. Further information on this mapping is described in Chapter 4.

Finally, the **Re-evaluation triggers** are a list of abbreviations for the trigger scenarios that warrant considerations of re-testing the respective criteria upon their occurrence. For more information, see Chapter 6.

4 Mapping of the Evaluation Criteria to Regulatory Requirements

The financial sector in the EU is subject to extensive regulatory requirements. In the context of AI systems, the EU AI Act[1] and DORA[2] are particularly noteworthy. Furthermore, AI systems used e.g., for creditworthiness assessments and capital requirement calculations are subject to additional regulatory frameworks, such as the Capital Requirements Regulation (CRR). As many of the evaluation criteria in the catalogue are thematically related to regulatory requirements set out in the AI Act, DORA and other financial sector regulations, the catalogue provides a mapping of these criteria to the relevant regulatory provisions. Therefore, meeting the evaluation criteria can help meet the corresponding regulatory provision to a certain extent. In the case of regulatory requirements from legal sources (such as DORA), the mapping is based on individual articles (e.g. DORA Art. 5).

Important: The referenced regulatory provisions are thematically related to the evaluation criterion, according to the author's interpretation. Fulfilling an evaluation criterion can contribute to fulfilling the corresponding regulatory provision to a certain extent. However, this does not imply that a regulatory provision is fully met simply by fulfilling the corresponding evaluation criterion. Whether and to what extent the regulatory provisions are met by fulfilling the evaluation criteria is at the discretion of the auditor, notified body or competent authority examining compliance with the given regulatory requirements. Furthermore, it cannot be anticipated how future guidance and interpretations may change the overlap between evaluation criteria and mapped regulations. Finally, it cannot be guaranteed that the regulatory references are correct, accurate and complete.

The AI Act sets out extensive provisions in connection with the development, placing on the market and deployment of AI systems, many of which are aimed at providers of such systems (see in particular Arts. 4, 6-22, 50, 72 and 73 AI Act). A natural or legal person is deemed to be a provider of an AI system if it develops an AI system and places it on the market or puts it into service for its own purposes. The mere deployment of (non-self-developed) AI systems is also subject to certain provisions (see in particular Arts. 4, 26, 27, 49, 50 and 86 of the AI Act). The majority of the provisions only apply to AI systems that are used in predefined high-risk areas (so-called high-risk AI systems). Outside the high-risk areas, the development and deployment of certain AI systems is subject to transparency provisions (e.g., when chatbots interact with natural persons). In addition, both providers and deployers of AI systems must ensure that their employees have a sufficient level of AI literacy (Art. 4 AI Act). A central and overarching provision of the AI Act for providers of high-risk AI systems is the maintenance of a quality management system (QMS, Art. 17 AI Act) to ensure permanent conformity with the AI Act. It can be assumed that most of the aspects addressed in the criteria catalogue are part of good quality management of AI systems. For this reason, the QMS provision was assigned to the majority of the criteria. It should be noted that for financial institutions, the AI Act assumes that some essential components of a QMS are in place, as they are already subject to extensive internal governance provisions under financial sector-specific regulation (see Art. 17 (4) EU AI Act). In addition to the provisions for AI systems, the AI Act also formulates provisions for the development and placing on the market of ”general purpose AI models” (so-called ”GPAI models”, see Art. 53 and Art. 55 AI Act), which are aimed at providers of such models. A natural or legal person is deemed to be a provider of a GPAI model if it places a GPAI model on the market (see Art. 3(4) AI Act). As providers of GPAI models are not included in the addressees of the criteria catalogue, the mapping is limited to the provisions for AI systems; the provisions for GPAI models were not considered further.

The Digital Operational Resilience Act (DORA) is an EU regulation aimed at enhancing the operational resilience of financial institutions by ensuring their ability to withstand, respond to, and recover from IT-related disruptions. It establishes a comprehensive, principle-based framework with broad requirements

[1] Regulation (EU) 2024/1689 (original version from 12.07.2024)
[2] Regulation (EU) 2022/2554 (original version from 27.12.2022)

around Information and Communication Technology (ICT) risk management, incident reporting, third-party risk and operational resilience testing. As ICT systems, AI systems are also subject to the technology-neutral DORA requirements. Complementing the DORA framework, Germany's financial supervisory authority, BaFin, has issued an orientation document on AI that provides additional supervisory guidance on managing ICT risks associated with AI systems in the financial sector. For the mapping, the following DORA legal sources were considered:

- Regulation (EU) 2022/2554 (original version from 14.12.2022) (Digital Operational Resilience Act, short: DORA)
- Commission Delegated Regulation (EU) 2024/1774 (consolidated version from 25.06.2024) (short: RMF-RTS)

When mapping the DORA requirements, it was assumed that users of the catalogue would be fully subject to the DORA requirements on the risk management framework (Arts. 5-15, DORA). For this reason simplified risk management (Art. 16, DORA) was removed from consideration and the RMF-RTS relating to the simplified risk management framework were not considered further. Therefore, for proportionality reasons, certain referenced provisions may not apply to all users of the catalogue.

If financial institutions use AI systems e.g., to determine capital requirements as part of the Internal Ratings Based (IRB) approach (in accordance with the Capital Requirements Regulation, CRR), these systems/models are subject to additional, extensive regulatory requirements. Particular attention must be paid to the provisions set out in the following regulatory sources.

- Regulation (EU) No 575/2013 (consolidated version 01.01.2025) (Capital Requirements Regulation, short: CRR)
- Commission Delegated Regulation (EU) 2022/439 (original version from 18.03.2022) (short: IRBA-RTS)
- EBA Guidelines on PD/LGD estimation (version from 20.11.2017) (short: EBA/GL/2017/16)
- ECB Guide to internal models (version from July 2025) (short: EGIM)

AI systems that are used (only) in the context of risk management for the credit assessment of debtors (without influencing capital requirements) are subject to fewer regulations but still must meet certain requirements. In this context, provisions from the following regulatory sources were considered.

- EBA Guidelines on loan origination and monitoring (version from 25.08.2022) (in short: EBA/GL/2020/06)
- Mindestanforderungen an das Risikomanagement (BA) (Rundschreiben 06/2024 (BA) vom 29.05.2024), (in short: MaRisk)

The mapped AI Act and DORA provisions have been directly incorporated into the criteria catalogue (see Chapter 7), while the mapping of the remaining provisions can be found in Appendix 18.

In this context, consideration should also be given to the mapping between the AI Act and the regulatory framework for the financial sector provided by the European Banking Authority (EBA)[3]. However, the level of granularity and the scope of the considered regulatory requirements differ slightly from the mapping presented in this catalogue.

[3] Outcome of EBA's AI Act mapping exercise (EBA/2925/D/5384, Link) (accessed on 27.01.2026)

5 A Risk-based, Holistic and Generalizable Audit Approach

The criteria catalogue is also intended to serve as a self-assessment tool for market participants, e.g., as preparation for an upcoming audit. Not all evaluation criteria are equally relevant for all AI systems. To apply the catalogue, a corresponding generalised testing approach was therefore developed, which can be applied to various scenarios and use cases in the financial industry. For this purpose, the evaluation process is adapted based on the risk and characteristics of the system under consideration. The aim is to create a flexible testing approach that can be transferred to different application areas and that meets the specific and diverse provisions of AI applications in the financial sector.

For the evaluation of the system, whether as part of a self-assessment or by external auditors, a certain level of technical understanding is generally required, particularly in the field of artificial intelligence. Basic knowledge of AI and the existing system is sufficient for answering the initial questionnaire to classify the use case and select suitable evaluation criteria. For the actual evaluation, however, more in-depth specialist knowledge is required in order to implement the described test procedures and the supportive guidance, to use recommended test tools if necessary, to interpret test results correctly and to make a well-founded final assessment.

Due to the heterogeneity of use cases, e.g., in terms of data foundation, area of application, architecture, etc., each evaluation of a use case requires an individual selection of relevant criteria to ensure targeted and efficient testing. At the beginning, a questionnaire helps to identify the relevant criteria for the evaluation process. This allows relevant criteria to be filtered out specifically for the respective use case. For example, the questionnaire inquires whether personal data is processed to evaluate the relevance of fairness-related criteria.

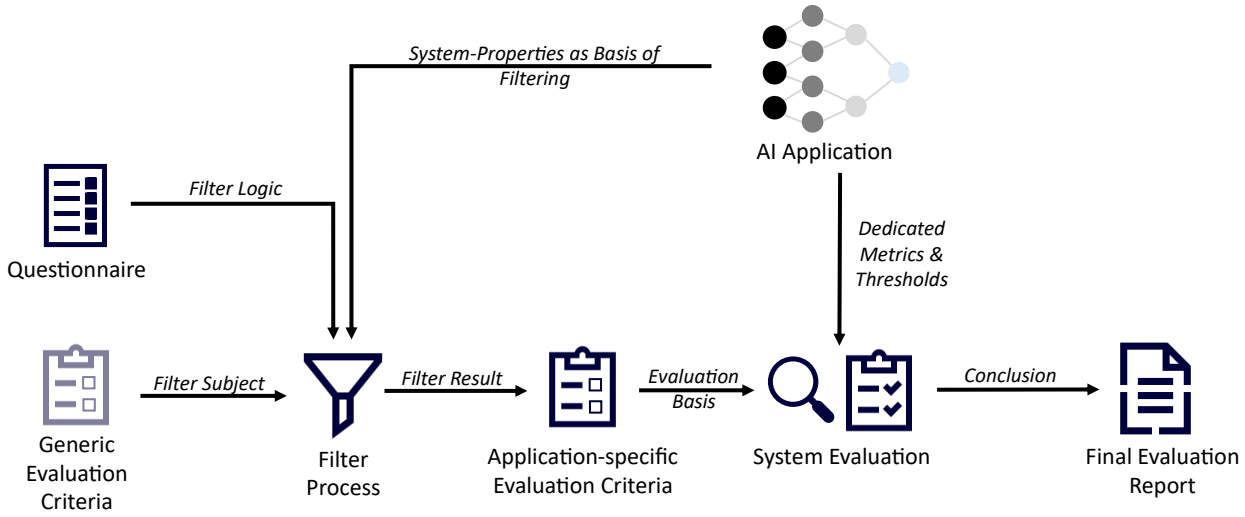


Figure 2: Application of the criteria catalogue as a process-based representation.

The process of an evaluation using the AICRIV criteria catalogue begins, as shown in Figure 2, with the filtering or parametrization of the use case to be tested. The relevant criteria are then output, which must be adapted according to the system under consideration so that they can be applied directly in a (practical) evaluation. A test-based criterion requires a specific method or tool for testing that can be adapted using examples from the supportive guidance. Furthermore, thresholds must be defined according to the selected method or metric and in line with the use case requirements under consideration, which define when a test is considered to be passed. If necessary, the evaluation method is also specified accordingly, and a suitable evaluation tool is selected to check the requirement. In this way, the generic criteria are tailored to the use case under consideration.

In the final step, the criteria catalogue is applied in practice. The evaluation method suggests how the criterion should be checked. The evaluation tool section provides examples for means of evaluation to investigate the fulfilment of the criterion. When testing the AI application using the specific criteria, particular focus should be placed on correct documentation in order to ensure transparency, traceability and consistency in the implementation and testing. A final evaluation report documents the procedure and findings for all evaluated criteria.

5.1 Questionnaire for Filtering Relevant Criteria

[Q1] Does the system integrate Generative AI Foundation Models like Large Language Models?

- Yes: The system integrates at least one Generative AI Foundation Model.
- No: Other approaches, such as Deep Learning or Shallow Learning Methods, were used.

[Q2] Is the architecture of the involved AI models fully known?

- Yes: The architecture is fully known.
- No: The architecture is not or only partially known.

[Q3] Are the parameters (e.g., weights or probability scores) of involved AI models known?

- Yes: There is full knowledge of the parameters of the involved AI models.
- No: There is only partial or zero knowledge of parameters of the involved AI models (e.g. usage of a third-party model with API access providing model results and probability scores or other meta-information / providing model results only).

[Q4] Is the training data fully available for inspection?

- Yes: The training data is fully available and can be inspected.
- No: The training data is only partially known or completely unknown.

[Q5] Has at least one integrated model been trained, even partially?

- Yes: Training of at least one integrated model was performed partially (e.g., with Fine-tuning) or from scratch.
- No: Not at all (e.g. because of using pre-trained / off-the-shelf models).

[Q6] Was personally identifiable or sensitive data used in the training of the involved AI models?

- Yes: Personal data was used to train the involved AI models.
- No: Personal data was not used to train the involved AI models.

[Q7] Does the AI system interact directly with external consumers (e.g., customers)?

- Yes: The model interacts directly with external consumers.
- No: The system does not interact directly with external consumers.

[Q8] Is the model required to explain the reasons for certain conclusions?

- Yes: The model must explain reasoning for specific inference results.
- No: A black box result by the system is sufficient.

[Q9] Does the AI system act completely autonomously (human-out-of-the-loop)?

- Yes: The AI system operates autonomously with no human oversight.
- No: The system integrates a human (either human-on-the-loop, human-in-the-loop, or human control).

[Q10] Does the system have an impact on society?

- Yes: The model has a societal impact.
- No: The model has no societal impact.

[Q11] Is the AI system built/operated on a hybrid infrastructure integrating on-premise as well as cloud components?

- Yes: The system is built/operated on a hybrid infrastructure.
- No: The system is built/operated on either an on-premise or cloud infrastructure.

6 Considerations for the Lifecycle of AI Systems

The increasing utilization of AI systems in safety-critical and business-critical domains necessitates robust and systematic approaches for assessing their safety, quality, and trustworthiness. To address this requirement, the evaluation process should be conceived as an integral component of the AI system life cycle rather than as an isolated activity. Embedding the proposed evaluation scheme within the life cycle enables methodical and efficient testing, reduces redundant verification activities, and ensures that assessments are conducted at appropriate points during the life cycle.

In order to facilitate this, each criterion is explicitly linked to specific phases of the AI life cycle. For every criterion, one phase is designated as the preferred evaluation phase positioned as late as feasible to capture the system's final configuration, while still allowing adequate time for corrective actions prior to deployment. The preceding phases primarily serve preparatory purposes, such as ensuring tool compatibility, collecting relevant evidence, and compiling the documentation required for subsequent verification. In certain cases, additional phases following the main evaluation stage are included to facilitate targeted follow-up checks, thereby ensuring continued compliance without necessitating a full reassessment.

Figure 3: AI life cycle phases as considered for the criteria catalogue.

This systematic alignment of evaluation activities with life cycle phases enhances transparency, efficiency, and reproducibility. The reduction of testing effort is achieved not by diminishing the substantive scope of evaluations, but through structured process design, early planning, and the reuse of prior results across iterative development cycles. As a result, the integration of the evaluation process into the AI life cycle ensures sustained assurance of safety, quality, and regulatory compliance over the entire system life span.

Re-evaluations, as part of the AI system life cycle, ensure that systems continue to meet safety, functionality, and compliance requirements throughout their operation. Unlike the initial assessment, which provides a comprehensive evaluation at the start, re-evaluations focus on verifying specific aspects of the system that may have been affected by changes in its context, components, or regulatory environment. Since continuous assessment is rarely feasible, re-evaluations enable a balanced approach providing timely assurance of system integrity without unnecessary resource consumption.

Re-evaluations can occur periodically, e.g., when mandated by regulatory conditions, or may be caused by specific triggers. These are specific events or conditions that indicate the potential need to reassess parts of the system. Such triggers may include, for example, modifications to the system architecture or model, detected malfunctions, updates to legal or regulatory requirements, or other significant operational changes. However, the presence of a trigger does not automatically necessitate a full reassessment. Instead, an impact analysis determines whether the trigger has a relevant effect on the system and, if so, to what extent a re-evaluation is required. For more information about re-evaluations and their triggers, we refer to the accompanying study of the AICRIV project.

To ensure a structured and transparent process, every criterion within the test criteria catalogue has specific triggers assigned to it. Table 1 shows the list of possible triggers assigned. These indicate when a re-evaluation of the corresponding criterion may be necessary. This systematic mapping allows evaluators to focus their efforts on affected areas, ensuring that re-evaluations are both precise and resource efficient. Again, the required need and scope of the re-assessment must be evaluated individually for every requirement.

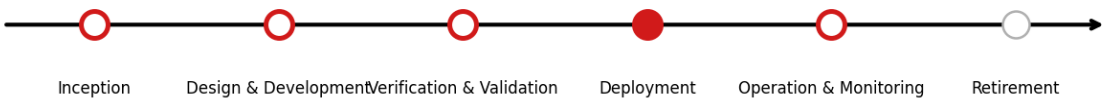
Table 1: Triggers for re-evaluation of criteria and their abbreviations within the catalogue.

Trigger	***Abbreviation***
System expansion	SE
System reduction	SR
Internal guidelines	IG
Legal requirements	LR
Standards, norms, or best practices	SN
Model architecture	MA
Application context	AC
New vulnerability	NV
Technical standards	TS
Advanced tools	AT
Training data	TD
Model parameters	MP
Error messages	EM
Unexpected results	UR
Unstable performance	UP
Ethical guidelines	EG
User group or permissions	UG
Sociatel impact	SI

I. Criteria Catalogue

7 AI Security & Robustness

AISR-01: Handling of AI specific security incidents



Lifecycle overview for criteria Lifecycle overview for criteria AISR-01

Evaluation requirement:

The system shall comply with applicable (cloud) standards and regulations, ensuring necessary security controls, incident management, and auditing. Any AI model security incidents shall be promptly addressed, logged, and documented.

Evaluation method:

Document-based: Verify that the system is compliant to necessary standards and regulations. Check whether an adequate process for handling security incidents is in place and previous identified security incidents related to the AI model were addressed and documented.

Supportive guidance:

Supplementary Definitions:

Security Controls:
Measures and mechanisms implemented to safeguard systems, data, and processes against threats, ensuring confidentiality, integrity, and availability as per applicable regulations and standards.

AI-Specific Security Incidents:
Events where AI systems are compromised, misused, or malfunction due to vulnerabilities, attacks, or breaches, potentially impacting model integrity, data privacy, or system behaviour.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

Regulations & standards: e.g. Cloud Computing Compliance Criteria Catalogue (C5), NIST, NIS-2.
For more information on the handling of security incidents, for example, the Policy for Security Incident Management (C5-SIM-01) from C5 can be consulted.

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 13, 15, 17, 26, 73

DORA (Level 1) References:

DORA Art. 10, 17

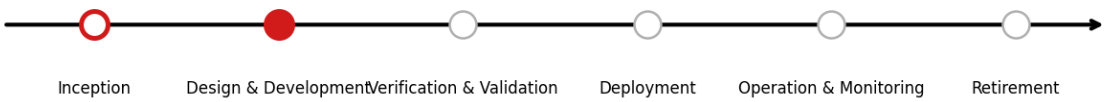
DORA (Level 2) References:

RMF-RTS Art. 12, 22, 23

Re-evaluation triggers:

IG, LR, SE, SN, SR

AISR-02: Assessment of threats in third party models



Lifecycle overview for criteria Lifecycle overview for criteria AISR-02

Evaluation requirement:

If the training data and/or model originate from third party, they should be analyzed for potential attack vectors and the results shall be documented.

Evaluation method:

Document-based: Ensure that third-party components are analysed for potential attack vectors. Ensure adequate documentation of the analysis.

Supportive guidance:

Supplementary Definitions:

Third Party:
An external entity (not directly affiliated with the system's owner), responsible for providing training data, models, or other services.

Attack Vectors:
Paths or methods leveraged by adversaries to exploit vulnerabilities in systems, such as data poisoning, model manipulation, or injection of malicious triggers or backdoors.

Exemplary [parameters]:

N/A

Additional guidance regarding the requirement:

Considerations should be at least given to:
- Triggers;
- Backdoors;
- Anomalies in the training data.

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

- There is only partial or zero knowledge of parameters of the involved AI models OR
- The training data is only partially known or completely unknown OR
- The training of at least one integrated model was performed partially or not at all.

Reference to EU AI Act:

AI Act Arts. 10, 11, 15, 17, 18, 25

DORA (Level 1) References:

DORA Art. 9 and Art. 28

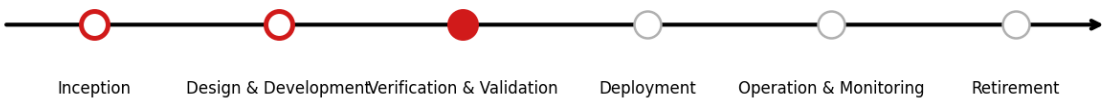
DORA (Level 2) References:

RMF-RTS Art. 11

Re-evaluation triggers:

AC, AT, MA, NV, TS

AISR-03: Risks from non-specified inputs



Lifecycle overview for criteria Lifecycle overview for criteria AISR-03

Evaluation requirement:

The behaviour of the system for, and the risk associated with, inputs that do not conform to the specified requirements shall be analysed with [methods] and documented.

Evaluation method:

Document-based: Check that the behaviour of the system for inputs that do not meet the specified requirements is analysed and documented.

Supportive guidance:

Supplementary Definitions:

Associated Risks:
The likelihood and potential impact of adverse effects, such as security vulnerabilities or functional failures, caused by handling non-conforming inputs.

Non-Conforming Inputs:
Inputs that deviate from defined standards, such as malformed, corrupted, or improperly formatted data, potentially affecting system integrity or performance.

Exemplary [parameters]:

[methods]: Testing frameworks, Robustness assessment tools, American Fuzzy Lop ++

Additional guidance regarding the requirement:

Damaged or manipulated inputs can be corrected if it is safely possible.

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 9, 10, 11, 13, 15, 17, 18

DORA (Level 1) References:

N/A

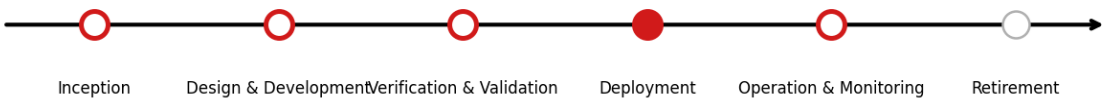
DORA (Level 2) References:

N/A

Re-evaluation triggers:

EM, LR, MA, MP, TD, UP, UR

AISR-04: Assessment of security threats



Lifecycle overview for criteria Lifecycle overview for criteria AISR-04

Evaluation requirement:

Procedures shall be implemented to continuously monitor and assess new threats related to the AI model(s) and AI system.

Evaluation method:

Document-based: Verify that procedures are in place to continuously monitor and assess emerging threats. Ensure that the process covers all potential threats as defined in the supportive guidance.

Supportive guidance:

Supplementary Definitions:

Not applicable.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

All potential threats to the AI system should be carefully considered, including design and technical faults, environmental risks, and cyber threats. A thorough review of data and features is essential to identify vulnerabilities and threats throughout the entire AI lifecycle.
These threats can include specific technical faults (e.g., hardware information leakage) and cyber threats (e.g., DDoS attacks, backdoors).
In relation to privacy threats, the analysis should focus on identifying and documenting vulnerabilities in the ML model that could lead to unintended personal inferences or potential side-channel attacks, which might result in the leakage of personal data.

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 9, 11, 15, 17, 18

DORA (Level 1) References:

DORA Art. 8, 9, 45

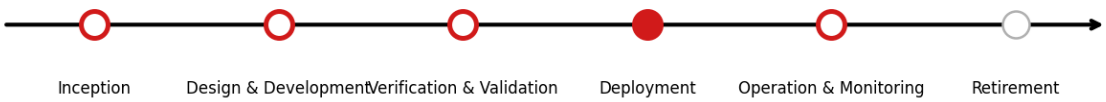
DORA (Level 2) References:

RMF-RTS Art. 3, 18

Re-evaluation triggers:

AT, EG, IG, LR, NV, SN, TS

AISR-05: Documented security review protocol



Lifecycle overview for criteria Lifecycle overview for criteria AISR-05

Evaluation requirement:

The frequency, scope, and method(s) for security and robustness reviews of productive models shall be documented.

Evaluation method:

Document-based: Ensure adequate documentation of the frequency, scope and methodologies for independent safety reviews of product models.

Supportive guidance:

Supplementary Definitions:

Security Reviews:
Systematic assessments to detect and address vulnerabilities in productive models, ensuring they are safeguarded against potential threats, breaches, and unauthorized access.

Robustness Reviews:
Complete evaluations of a model's ability to maintain functionality and performance under adversarial attacks, unexpected inputs, or challenging conditions.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

Not applicable.

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 9, 11, 13, 15, 17, 18

DORA (Level 1) References:

N/A

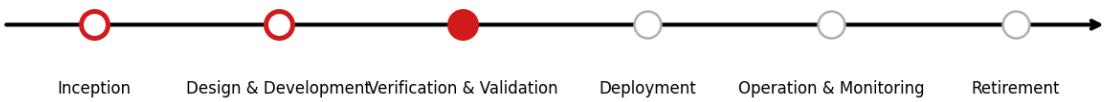
DORA (Level 2) References:

RMF-RTS Art. 3

Re-evaluation triggers:

AC, AT, EM, IG, LR, MA, MP, NV, SN, TD, TS, UP

AISR-06: Formal verification



Lifecycle overview for criteria Lifecycle overview for criteria AISR-06

Evaluation requirement:

The possibility of verifying the correctness of the output by formal verification, at least of modules of the full system shall be considered.

Evaluation method:

Test-based: Verify that formal verification has been considered to verify the correctness of the output. Evaluate the formal guarantees globally (for the whole system) or module-wise using the suggested evaluation tools from the supportive guidance.

Supportive guidance:

Supplementary Definitions:

Formal Verification:
The use of mathematical techniques to ensure that the system or its components operate as intended by checking against predefined rules and specifications.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

Formal verification can be conducted by mapping functions to smaller AI models that can be individually checked and verified.

Evaluation tools:

- ReLUPlex - can be used to formally verify neural network properties by checking activation functions and constraints in individual AI model components.
- Abstract Domain Transformers - can be used to perform static analysis and verify AI model behavior against predefined specifications.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 9, 15, 17

DORA (Level 1) References:

N/A

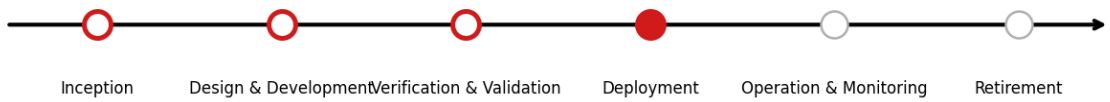
DORA (Level 2) References:

N/A

Re-evaluation triggers:

LR, MA, MP, SN, TD, TS

AISR-07: White box robustness



Lifecycle overview for criteria Lifecycle overview for criteria AISR-07

Evaluation requirement:

The attack surface for white-box evasion attacks shall be assessed with appropriate [attack vectors].

Evaluation method:

Test-based: Verify that the attack surface has been assessed for white box attacks. Verify that the system was tested against relevant white-box attacks and demonstrates adequate resistance.

Supportive guidance:

Supplementary Definitions:

White-Box:
Scenarios where attackers have full access to the internal structure, parameters, and algorithms of a system or model, allowing detailed analysis and exploitation.

Evasion Attacks:
Methods where attackers modify inputs to trick a system or model into making wrong or unexpected decisions without changing the system itself.

Exemplary [parameters]:

[attack vectors]: Standard Adversarial Attacks - FGSM (Fast gradients sign method), PGD (Project gradient descent), BIM (Basic Iterative Method), CW (Carlini-Wagner).

Additional guidance regarding the requirement:

Not applicable.

Evaluation tools:

- CleverHans - can be used to generate adversarial examples and evaluate the AI model's robustness against common white-box evasion attacks.
- Adversarial Robustness Toolbox - can be used to systematically test AI models with a range of adversarial attacks and measure their resistance to white-box evasion techniques.

Relevance based on use case parametrization:

- The architecture is fully known AND
- There is full knowledge of parameters of the involved AI models.

Reference to EU AI Act:

AI Act Arts. 9, 15, 17

DORA (Level 1) References:

DORA Art. 25

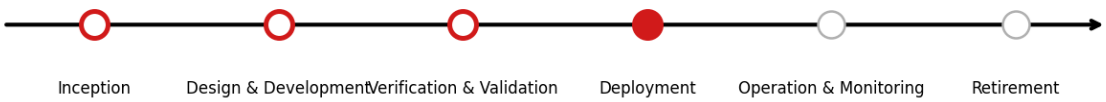
DORA (Level 2) References:

N/A

Re-evaluation triggers:

AT, IG, LR, MA, MP, NV, SN, TD, TS, UG

AISR-08: Gray and black box robustness via surrogate model



Lifecycle overview for criteria Lifecycle overview for criteria AISR-08

Evaluation requirement:

The attack surface for gray and black box evasion attacks shall be assessed with appropriate [attack vectors].

Evaluation method:

Test-based: Verify that the attack surface has been assessed for gray and black box attacks. Verify that the system was tested against relevant gray-/black-box attacks and demonstrates adequate resistance.

Supportive guidance:

Supplementary Definitions:

Gray Box:
A scenario where attackers have partial knowledge of the system or model, such as access to some parameters, architecture details, or limited input-output behaviour.

Black Box:
A scenario where attackers have no internal knowledge of the system or model and rely on observable input-output behaviour.

Evasion Attacks:
Methods where attackers modify inputs to trick a system or model into making wrong or unexpected decisions without changing the system itself.

Exemplary [parameters]:

[attack vectors]: e.g., Standard gray and black box attacks - Transfer attack with similar (pretrained) surrogate model, Training of a (fine-tuned) surrogate model if high bandwidth with original model is available, Z00 (Zero order attack), Boundary Attack, Single Pixel Attack.

Additional guidance regarding the requirement:

The known information should be assumed realistically (obtainable by an attacker).

Evaluation tools:

- CleverHans - can be used to simulate gray-box and black-box attacks using techniques such as transfer attacks, surrogate models, and boundary attacks.
- Adversarial Robustness Toolbox - can be used to conduct advanced black-box and gray-box evasion tests.

Relevance based on use case parametrization:

- The architecture is not or only partially known OR
- There is only partial or zero knowledge of parameters of the involved AI models.

Reference to EU AI Act:

AI Act Arts. 9, 15, 17

DORA (Level 1) References:

DORA Art. 25

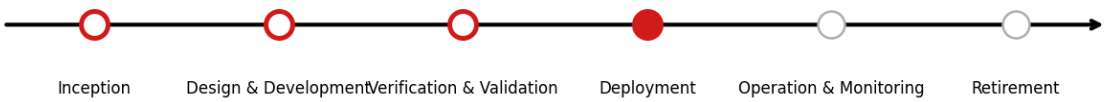
DORA (Level 2) References:

N/A

Re-evaluation triggers:

AT, IG, LR, MA, MP, NV, SN, TD, TS, UG

AISR-09: White box attacks



Lifecycle overview for criteria Lifecycle overview for criteria AISR-09

Evaluation requirement:

The attack surface for white-box information extraction attacks shall be assessed with appropriate [attack vectors].

Evaluation method:

Test-based: Verify that the attack surface has been assessed for white-box information extraction attacks. Verify that relevant attack vectors have been evaluated and that the sensitive information, such as training data, model parameters, or underlying logic, can not be extracted.

Supportive guidance:

Supplementary Definitions:

White-Box:
Scenarios where attackers have full access to the internal structure, parameters, and algorithms of a system or model, allowing detailed analysis and exploitation.

Extraction Attacks:
Methods where attackers try to obtain sensitive information from a model, such as its training data, parameters, or underlying logic, by analyzing its responses or internal structure.

Exemplary [parameters]:

[attack vectors]: e.g., Membership inference attacks, Property/Attributed Inference attacks, Reconstruction Attacks (of training samples).

Additional guidance regarding the requirement:

Not applicable.

Evaluation tools:

- Adversarial Robustness Toolbox - can be used to conduct membership inference, property inference, and reconstruction attacks to test the model's vulnerability to information extraction.
- TensorFlow Privacy - can be used to assess differential privacy protections and analyze the model's susceptibility to white-box extraction techniques.
- Privacy Meter - can be used to evaluate membership inference risks and determine if attackers can infer whether specific data was used in model training.

Relevance based on use case parametrization:

- The architecture is fully known AND
- There is full knowledge of parameters of the involved AI models.

Reference to EU AI Act:

AI Act Arts. 9, 15, 17

DORA (Level 1) References:

DORA Art. 25

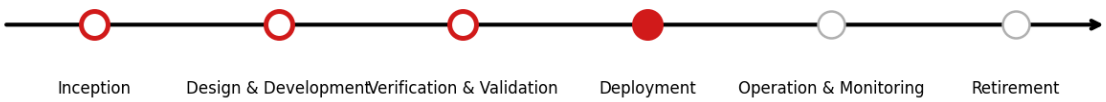
DORA (Level 2) References:

N/A

Re-evaluation triggers:

AC, AT, LR, MA, MP, NV, SN, TD, TS, UG

AISR-10: Gray and black box attacks



Lifecycle overview for criteria Lifecycle overview for criteria AISR-10

Evaluation requirement:

The attack surface for gray and black box information extraction attacks shall be assessed with appropriate [attack vectors].

Evaluation method:

Test-based: Verify that the attack surface has been assessed for grey-/and black-box information extraction attacks. Verify that relevant attack vectors have been evaluated and that the sensitive information, such as training data, model parameters, or underlying logic, can not be extracted.

Supportive guidance:

Supplementary Definitions:

Gray Box:
A scenario where attackers have partial knowledge of the system or model, such as access to some parameters, architecture details, or limited input-output behaviour.

Black Box:
A scenario where attackers have no internal knowledge of the system or model and rely on observable input-output behaviour.

Extraction Attacks:
Methods where attackers try to obtain sensitive information from a model, such as its training data, parameters, or underlying logic, by analyzing its responses or internal structure.

Exemplary [parameters]:

[attack vectors]: e.g., Label-Only Membership Inference Attacks, Model Extraction, Prompt Injection, Divergence Attack.

Additional guidance regarding the requirement:

The known information should be assumed realistically (obtainable by an attacker). The information extracted does not need to be limited to training data but can be all data that is connected to the AI model in any way.

Evaluation tools:

- Adversarial Robustness Toolbox - can be used to conduct label-only membership inference, model extraction attacks, and divergence attacks to evaluate how much information can be extracted with limited access.
- TensorFlow Privacy - can be used to assess privacy-preserving techniques and analyze how effective they are against gray-box and black-box extraction attacks.

- Privacy Meter - can be used to evaluate the risk of membership inference and determine whether attackers can infer training data presence in black-box scenarios.

Relevance based on use case parametrization:

- The architecture is not or only partially known OR
- There is only partial or zero knowledge of parameters of the involved AI models.

Reference to EU AI Act:

AI Act Arts. 9, 15, 17

DORA (Level 1) References:

DORA Art. 25

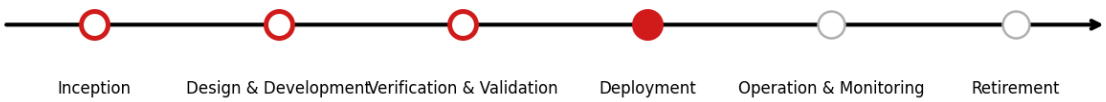
DORA (Level 2) References:

N/A

Re-evaluation triggers:

AC, AT, LR, MA, MP, NV, SN, TD, TS, UG

AISR-11: Backdoor attacks



Lifecycle overview for criteria Lifecycle overview for criteria AISR-11

Evaluation requirement:

The attack surface for backdoor attacks shall be assessed with appropriate [attack vectors].

Evaluation method:

Test-based: Verify that the attack surface for backdoor attacks, e.g., data poisoning or tampering, has been assessed. Verify that relevant attack vectors have been evaluated exploiting the inserted vulnerabilities and the system demonstrates adequate resistance.

Supportive guidance:

Supplementary Definitions:

Backdoor attack: A backdoor attack is a type of data poisoning attack where an adversary implants a hidden pattern (trigger) in the training data so that the model behaves normally on clean inputs but produces incorrect or malicious outputs when the trigger is present in the input. The goal is to create a model that is covertly controlled under specific conditions while remaining undetected during regular use.

Exemplary [parameters]:

[attack vectors]: e.g., Data poisoning or data tampering through backdoors.

Additional guidance regarding the requirement:

Not applicable.

Evaluation tools:

- Adversarial Robustness Toolbox (ART) - can be used to conduct data poisoning and backdoor detection, simulating adversarial manipulations of training datasets.
- LabelFix - can be used to detect and analyze mislabeled or tampered data.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 9, 15, 17

DORA (Level 1) References:

DORA Art. 25

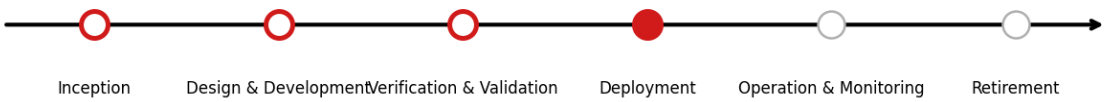
DORA (Level 2) References:

N/A

Re-evaluation triggers:

AT, IG, LR, MA, MP, NV, SN, TD, TS, UG

AISR-12: Robustness on corner cases



Lifecycle overview for criteria Lifecycle overview for criteria AISR-12

Evaluation requirement:

The robustness of the system on corner cases and/or boundary values shall be assessed with suitable [measures] and documented.

Evaluation method:

Test-based: Verify if the AI system maintains reliable and stable when exposed to extreme or unexpected inputs and assess that the robustness is compliant with the minimum acceptable robustness of the system.

Supportive guidance:

Supplementary Definitions:

Robustness:
Robustness in AI systems refers to the ability to function reliably and consistently even under uncertain conditions, such as unexpected inputs or disturbances. A robust AI system delivers consistent and correct results, even when faced with noisy or manipulated data such as adversarial attacks. The goal is to ensure that the system is not easily disrupted by external influences.

Boundary Values:
Inputs at the edge of an allowed range or threshold, used to evaluate a system's performance and stability under extreme conditions.

Exemplary [parameters]:

[measures]: Empirical statistical robustness tests for identified corner cases, testing on the boundary values of the system (if quantifiable).

Additional guidance regarding the requirement:

Not applicable.

Evaluation tools:

- Adversarial Robustness Toolbox - can be used to generate adversarial perturbations and edge-case inputs to test model robustness.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 9, 11, 15, 17, 18

DORA (Level 1) References:

DORA Art. 25

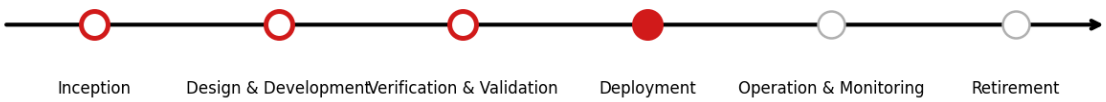
DORA (Level 2) References:

N/A

Re-evaluation triggers:

AT, IG, LR, MA, MP, NV, SN, TD, TS, UG

AISR-13: I/O security



Lifecycle overview for criteria Lifecycle overview for criteria AISR-13

Evaluation requirement:

Input and output shall be protected from tampering by classical [measures].

Evaluation method:

Test-based: Verify if classical security measures are in place to prevent unauthorized alteration or manipulation of data.

Supportive guidance:

Supplementary Definitions:

Tampering:
The unauthorized alteration, manipulation, or interference with data, inputs, or outputs to compromise the integrity, functionality, or security of the system.

Exemplary [parameters]:

[measures]: e.g., Data encryption, input validation and sanitization, authentication and authorization (see ITSEC-36), integrity checks, rate limiting and throttling, secure APIs and endpoints, audit logging and monitoring, human in the loop oversight.

Additional guidance regarding the requirement:

It is worthwhile to assess the ability of a standard user or even a developer to tamper with data.

Evaluation tools:

- Taint Analysis - can be used to track untrusted data sources and ensure that there is input validation, sanitization, and protection against tampering.
- Static Code Analysis - can be used to detect security vulnerabilities related to data integrity, authentication, authorization, and secure API handling.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 9, 15, 17

DORA (Level 1) References:

DORA Art. 9, 10

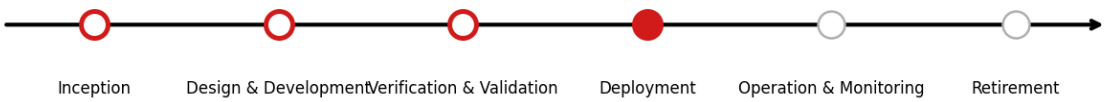
DORA (Level 2) References:

RMF-RTS Art. 3, 6, 12, 13, 14

Re-evaluation triggers:

AT, LR, SE, SN, SR, TS

AISR-14: Countermeasures against evasion attacks



Lifecycle overview for criteria Lifecycle overview for criteria AISR-14

Evaluation requirement:

[Countermeasures] against evasion attacks shall be implemented and documented.

Evaluation method:

Test-based: Verify that countermeasures against adversarial attacks are in place and work properly.

Supportive guidance:

Supplementary Definitions:

Evasion Attacks:
Methods where attackers modify inputs to trick a system or model into making wrong or unexpected decisions without changing the system itself.

Exemplary [parameters]:

[countermeasures]: e.g., Adversarial training, anomaly detection, employing multiple AI systems with different architectures or training data redundantly against adversarial attacks.

Additional guidance regarding the requirement:

Not applicable.

Evaluation tools:

- CleverHans - can be used to generate adversarial examples and test the effectiveness of implemented countermeasures.
- Adversarial Robustness Toolbox - can be used to conduct structured adversarial attacks and evaluate the model's resistance to evasion attempts.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 9, 11, 13, 15, 17, 18

DORA (Level 1) References:

N/A

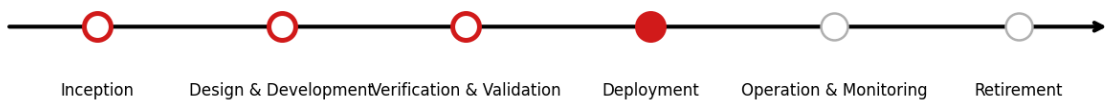
DORA (Level 2) References:

N/A

Re-evaluation triggers:

AT, LR, MA, MP, NV, SN, TD, TS

AISR-15: Countermeasures against information extraction attacks



Lifecycle overview for criteria Lifecycle overview for criteria AISR-15

Evaluation requirement:

[Countermeasures] against information extraction attacks shall be implemented and documented.

Evaluation method:

Test-based: Verify that effective measures are in place to prevent unauthorized extraction of training data, data used by the model, model parameters, or internal logic.

Supportive guidance:

Supplementary Definitions:

Extraction Attacks:
Methods where attackers try to obtain sensitive information from a model, such as its training data, parameters, or underlying logic, by analyzing its responses or internal structure.

Exemplary [parameters]:

[countermeasures]: e.g.,
- Differential Privacy (epsilon-DP training (e.g., DP-SGD), Rényi Differential Privacy)
- Regularization Techniques (L2/L1 norm regularization, Dropout / DropConnect, Label smoothing)
- Data Augmentation (Synthetic data generation (GAN-based), Mixup and CutMix strategies)
- Knowledge Distillation (Use of student-teacher model architectures where only the student is exposed to queries)
- Model Compression/Pruning (Reduces memorization of individual training samples)
- Ensemble Models with Randomization (Stochastic selection among several models at inference time)
- Response Filtering (Limit to top-k or top-p outputs, Output obfuscation or rounding of probabilities)
- Access Restriction on Confidence Scores (Return only class labels instead of full confidence scores, Add noise to model outputs (especially probabilistic outputs))
- Rate Limiting / Throttling (Restrict frequency of API access or number of queries)
- Query Auditing and Anomaly Detection (Track query patterns to detect suspicious sequences indicative of extraction attempts)
- Dynamic Watermarking (Insert unique, hard-to-replicate patterns into model behavior to identify extraction or unauthorized use)
- Dataset Condensation or Sanitization (Remove outliers or memorization-prone samples from training datasets)
-Guardrails against LLM prompt injection

Additional guidance regarding the requirement:

Not applicable.

Evaluation tools:

- Privacy Meter - can be used to evaluate the effectiveness of membership inference attack countermeasures.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 9, 11, 13, 15, 17, 18

DORA (Level 1) References:

DORA Art. 9, 10

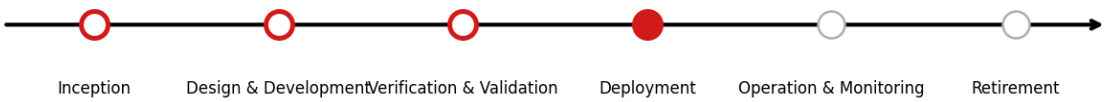
DORA (Level 2) References:

RMF-RTS Art. 3, 11, 21

Re-evaluation triggers:

AT, LR, MA, MP, NV, SN, TD, TS

AISR-16: Countermeasures against data poisoning attacks



Lifecycle overview for criteria Lifecycle overview for criteria AISR-16

Evaluation requirement:

[Countermeasures] against data poisoning attacks shall be implemented and documented.

Evaluation method:

Test-based: Verify that suitable measures to detect and mitigate data poisoning are in place and work properly.

Supportive guidance:

Supplementary Definitions:

Poisoning Attacks:
A type of adversarial attack where malicious data is injected into the training dataset to corrupt the model's learning process, leading to degraded performance, biased outputs, or hidden malicious behaviours.

Exemplary [parameters]:

[countermeasures]: e.g.,
- Data Collection & Ingestion (Access control to datasets (e.g., strict role-based access to data and version histories)
- Data provenance tracking using cryptographic methods (e.g., blockchain or signed datasets)
Source validation and authentication, especially for third-party or crowd-sourced data, Differential access policies for sensitive training data
- Data sanitization and filtering (e.g., removing outliers, duplicates, or inconsistent entries)
- Use of trusted data sources and certified data providers)
- Poisoning Detection & Analysis (Anomaly detection models for spotting suspicious data points (e.g., clustering, distance-based, or graph-based methods)
- Influence function analysis to measure training point impact, Backdoor detection tools (e.g., activation clustering, spectral signatures)
- Statistical validation (e.g., comparing class distributions)
- Data slicing and canary examples to detect poisoning, Model behavior auditing after fine-tuning)
- Training-Time Defenses (Robust training algorithms (e.g., trimmed loss minimization, differential privacy, Deep k-NN), Noise injection in training,)
- Post-Training Monitoring & Evaluation (Model fingerprinting and output distribution tracking, Shadow model comparisons, Periodic re-evaluation against clean benchmarks, Interpretability tools for suspect behavior (e.g., SHAP, LIME))

Additional guidance regarding the requirement:

Possible attacks: e.g., availability poisoning, targeted poisoning, backdoor poisoning, model poisoning, model skewing.

Evaluation tools:

- LabelFix - can be used to detect mislabeled or manipulated training samples that may indicate poisoning attempts.
- Dioptra - can be used to simulate and evaluate poisoning attacks, enabling assessment of existing countermeasures and model robustness.

Relevance based on use case parametrization:

At least partial training was performed.

Reference to EU AI Act:

AI Act Arts. 9, 11, 13, 15, 17, 18

DORA (Level 1) References:

DORA Art. 9, 10

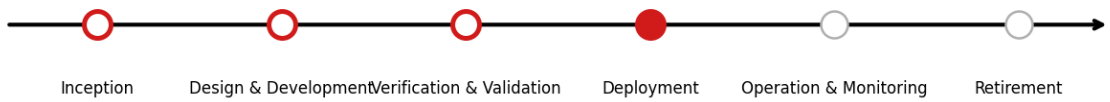
DORA (Level 2) References:

RMF-RTS Art. 3, 11, 21

Re-evaluation triggers:

AT, LR, MA, MP, NV, SN, TD, TS

AISR-17: Countermeasures against model theft attacks



Lifecycle overview for criteria Lifecycle overview for criteria AISR-17

Evaluation requirement:

[Countermeasures] against model theft attacks shall be implemented and documented, where the deployer is responsible for the intellectual property of the model or where the license terms of a third-party or open-source model require such protection.

Evaluation method:

Test-based: Verify that suitable countermeasures against model theft attacks are in place and work properly.

Supportive guidance:

Supplementary Definitions:

Model Theft Attacks:
Attempts by attackers to replicate or steal a machine learning model by exploiting access to its outputs, structure, or functionality.

Exemplary [parameters]:

[countermeasures]: e.g., access control, model obfuscation, watermarking, homomorphic encryption, limited queries/output.

Additional guidance regarding the requirement:

Not applicable.

Evaluation tools:

- Model Inversion Attack - those techniques can be used to assess whether sensitive model information can be extracted, evaluating the effectiveness of defenses.
- Adversarial Robustness Toolbox - can be used to simulate model theft attempts.
- Knockoff Nets - can be used to evaluate model extraction risks by replicating the behavior of a black-box model using query-based attacks.

Relevance based on use case parametrization:

Yes, training of at least one integrated model was performed partially (e.g., with finetuning) or from scratch.

Reference to EU AI Act:

AI Act Arts. 9, 11, 13, 15, 17, 18

DORA (Level 1) References:

DORA Art. 9, 10

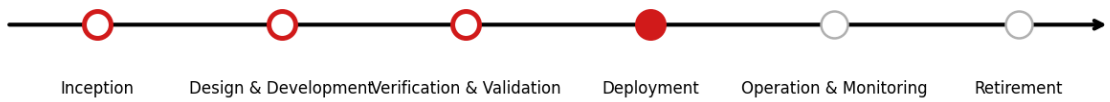
DORA (Level 2) References:

RMF-RTS Art. 3, 11, 21

Re-evaluation triggers:

AT, LR, MA, MP, NV, SN, TD, TS

AISR-18: Countermeasures against output integrity attacks



Lifecycle overview for criteria Lifecycle overview for criteria AISR-18

Evaluation requirement:

[Countermeasures] against output integrity attacks shall be implemented and documented.

Evaluation method:

Test-based: Verify that suitable countermeasures against output integrity attacks, i.e., attacks that manipulate the output of the AI system are in place and work properly.

Supportive guidance:

Supplementary Definitions:

Output Integrity Attack:
An attack that occurs when an adversary manipulates or alters the final results produced by the AI model without interfering with the input data or performing adversarial attacks on the model itself. In this type of attack, the input data remains untouched, but the generated output is modified or tampered with before it reaches the end user or decision-making system.

Exemplary [parameters]:

[countermeasures]: e.g., Encryption, secure communication channels, output monitoring.

Additional guidance regarding the requirement:

Not applicable.

Evaluation tools:

- Penetration Testing Tools (e.g., OWASP ZAP) - penetration testing can be used to identify vulnerabilities in web applications and APIs where output data may be intercepted or manipulated.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 9, 11, 13, 15, 17, 18

DORA (Level 1) References:

DORA Art. 9, 10

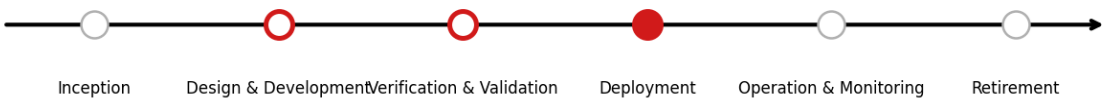
DORA (Level 2) References:

RMF-RTS Art. 3, 11, 21

Re-evaluation triggers:

AT, LR, MA, MP, NV, SN, TD, TS

AISR-19: Residual risk mitigation



Lifecycle overview for criteria Lifecycle overview for criteria AISR-19

Evaluation requirement:

If countermeasures tested and derived against attacks do not lower the risk to an acceptable level, or if there are no specific countermeasures available, alternative measures not linked to specific threat scenarios shall be developed, implemented, and tested.

Evaluation method:

Document-based: Analyse whether the countermeasures tested and derived against attacks reduce the risk to an acceptable level. Verify that alternative measures are implemented if not.

Supportive guidance:

Supplementary Definitions:

Not applicable.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

Suitable alternative measures eventually need to be investigated for the individual combination of use case and residual risk. An initial literature research can be performed and based on the results a possible solution can be developed inhouse or by an external service provider.

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 9, 15, 17

DORA (Level 1) References:

DORA Art. 6

DORA (Level 2) References:

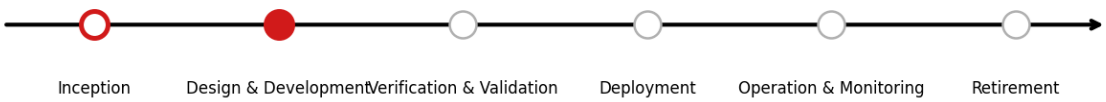
RMF-RTS Art. 3, 27

Re-evaluation triggers:

AT, LR, MA, MP, NV, SN, TD, TS

8 Data Quality & Management

DQM-01: Data planning



Lifecycle overview for criteria Lifecycle overview for criteria DQM-01

Evaluation requirement:

The data planning process shall be documented

Evaluation method:

Document-based: Verify that the data planning was performed and documented accordingly. Considerations should be given to the listed elements in the supportive guidance.

Supportive guidance:

Supplementary Definitions:

Data planning: The process of strategically defining how data will be collected, processed, managed, stored, analyzed, and maintained to meet organizational objectives and compliance requirements.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

The data planning process should consider at least the following elements:
- The data model or data architecture necessary to achieve the data requirements;
- Plan for acquiring the necessary data as identified by the data requirements;
- Roles, skills and people necessary to execute the data quality process;
- IT and other resources necessary to execute the data quality process;
- Time and budget necessary to execute the data quality process;
- Acquiring the data according to the data requirements;
- Executing data quality measures according to the data quality model;
- Meeting legal requirements;
- Meeting other data requirements.

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

At least partial training was performed.

Reference to EU AI Act:

AI Act Arts. 10, 11, 17, 18

DORA (Level 1) References:

N/A

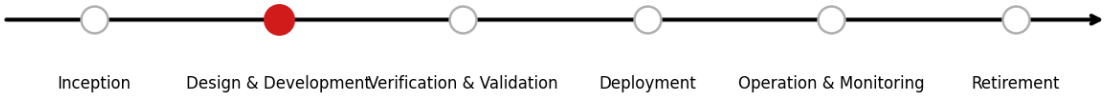
DORA (Level 2) References:

N/A

Re-evaluation triggers:

IG, LR, TD

DQM-02: Data acquisition



Lifecycle overview for criteria Lifecycle overview for criteria DQM-02

Evaluation requirement:

The data acquisition shall be documented.

Evaluation method:

Document-based: Verify that the data acquisition is adequately documented, see the supporting guidance for an indication of what should be included in the documentation.

Supportive guidance:

Supplementary Definitions:

Data acquisition: The process of collecting, measuring or obtaining raw data from various sources to be used for training and validating machine learning models.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

Considerations regarding data provenance should be given to:
- Compliance with the elements defined in the data planning process;
- Adherence to the principles of the data quality process;
- Data Ownership and Responsibility;
- Data Licensing and Usage Rights;
- Key data attributes identified during the data requirements process, such as provenance, bias, consistency, reliability, validity, data types, schema, format, and low data noise;
- The source of the data, whether it is internal or external;
- Verification of the trustworthiness of each data source used;
- The data collection process, specifying if it was conducted by humans, automated sensors, or both;
- The method of data collection, such as surveys or streaming;
- In the case of personal data, the original purpose of data collection.

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

At least partial training was performed.

Reference to EU AI Act:

AI Act Arts. 10, 11, 17, 18

DORA (Level 1) References:

N/A

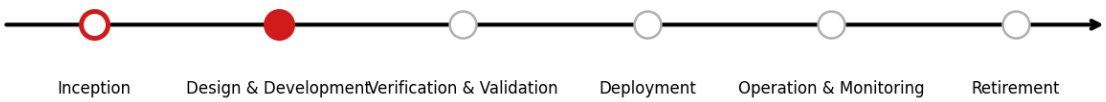
DORA (Level 2) References:

N/A

Re-evaluation triggers:

IG, LR, TD

DQM-03: Data quality requirements



Lifecycle overview for criteria Lifecycle overview for criteria DQM-03

Evaluation requirement:

Data Requirements and data quality management shall be documented

Evaluation method:

Document-based: Verify that the data requirements were developed and documented accordingly. Considerations should be given to the listed elements in the supportive guidance.

Supportive guidance:

Supplementary Definitions:

Data quality: The degree to which data meets the standards for (among others) consistency, informativeness, representativeness, timeliness, validity, reliability, completeness.
Data Management: The practice of organizing, storing, protecting and maintaining data to ensure its quality, accessibility, and usability throughout its lifecycle.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

Data requirements should consider at least the following elements:
- Definition of quality standards, including accuracy, consistency, completeness, and timeliness of the data;
- Specification of necessary attributes or features that the data must contain to support model training;
- Determination of the volume of data needed to achieve reliable results and generalization;
- Assessment of statistical measures such as mean, variance, and distribution characteristics to understand dataset behavior;
- Assurance that the dataset accurately represents the population in terms of demographics, behaviors, and geographies relevant to the problem;
- Compliance with regulations (e.g., GDPR) and ethical guidelines regarding data privacy and usage;
- Clear guidelines on which data to include or exclude based on specific attributes or quality metrics;
- Uncertainty and variability assessment;
- Annotation Requirements for how data should be labelled or categorized, including quality assurance measures;
- Inclusion and exclusion criteria for data based on relevant attributes;
- Technical inclusion and exclusion criteria;
- Guidelines on whether and how synthetic data can be used to augment the dataset;
- Processes for evaluating, resolving, and closing data quality issues;
- Regular reviews and validations of the quality and relevance of the data throughout the model lifecycle.

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

At least partial training was performed.

Reference to EU AI Act:

AI Act Arts. 10, 11, 13, 17, 18

DORA (Level 1) References:

N/A

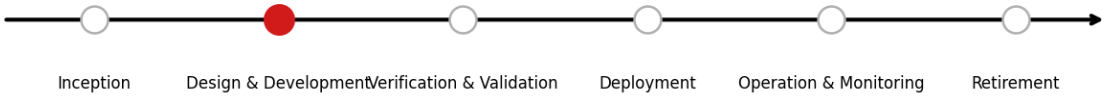
DORA (Level 2) References:

N/A

Re-evaluation triggers:

IG, LR, SN, TD, TS

DQM-04: Data quality assessment



Lifecycle overview for criteria Lifecycle overview for criteria DQM-04

Evaluation requirement:

The quality of gathered data shall be assessed and documented with [measures] and [corrective measures] shall be in place to ensure stability.

Evaluation method:

Test-based: Verify that a process for continuous assessment of data quality is in place and documented. Also verify that data subject to this process meets consistency, reliability, and bias-free standards, with corrective actions in place. The required data properties and key figures about the data that shall be validated are based on the data requirements (see DQM-03) directly or indirectly induced by performance/ functional requirements of the system. Examples for testing tools and corresponding data properties to test can be found in the column "Evaluation tool".

Supportive guidance:

Supplementary Definitions:

Stability: In the context of data, it refers to the degree to which data remains consistent, reliable and unchanged over time or under varying conditions. It is critical for maintaining trust in data-driven processes, ensuring consistency, and supporting robust decision-making.

Exemplary [parameters]:

[measures]: Data Validation, Consistency Check, Bias Detection Tools, Statistical (Power) Analysis, cross-validation of labels
[corrective measures]: LabelFix, Bias Correction Tools, Trigger model retraining when data shift is detected.

Additional guidance regarding the requirement:

Requirements according to DQM-03.

Evaluation tools:

- TensorFlow Data Validation - can be used to analyze dataset distribution, detect anomalies..
- LabelFix - can be used to identify inconsistencies, and ensure labeling quality.

Statistical Power Analysis - can be used to assess data sufficiency and reliability.
- Power Analysis Tools - can be used to verify if the dataset is large enough to detect meaningful statistical effects and support robust decision-making.

Dimensionality Reduction Analysis (verify stability under varying data conditions):
- PCA (Principal Component Analysis) - can be used to analyze variance in datasets and detect instability due to redundant or highly correlated features.

- t-SNE - can be applied for high-dimensional data visualization to detect clustering inconsistencies that may indicate bias or data drift.

Fairness and Bias Detection Tools (verify data fairness and detect bias-related instability):
- Fairlearn - can be used to valuate the fairness of datasets by measuring disparate impact and mitigating algorithmic bias.
- IBM AI Fairness 360 - can be used for advanced bias detection, fairness metrics analysis.

Data Quality Assessment Tools (evaluate correctness, consistency, and reliability of labeled data):
- Talend Data Quality - can be used to detect anomalies, check data consistency, and validate dataset structure.

Relevance based on use case parametrization:

At least partial training was performed.

Reference to EU AI Act:

AI Act Arts. 10, 15, 17

DORA (Level 1) References:

N/A

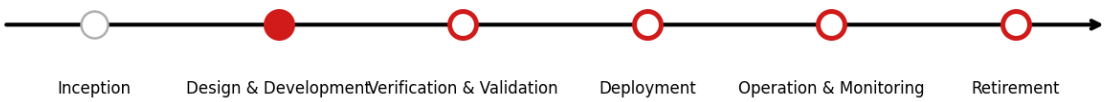
DORA (Level 2) References:

N/A

Re-evaluation triggers:

IG, LR, SN, TD, TS

DQM-05: Data tracking procedures



Lifecycle overview for criteria Lifecycle overview for criteria DQM-05

Evaluation requirement:

The traceability of data shall be ensured and documented.

Evaluation method:

Document-based: Verify that traceability of data is ensured regarding the factors stated in the supportive guidance.

Supportive guidance:

Supplementary Definitions:

Traceability: Refers to the ability to track, document, and verify the origin, history and transformation of data throughout its lifecycle.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

Some factors to consider regarding traceability:
- Clear identification of all data records through indexing;
- Recording and authorization of all data changes;
- Placing the entire data processing software and libraries under version control;
- Consistency of metadata;
- Safe delivery or transfer of data.

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

At least partial training was performed.

Reference to EU AI Act:

AI Act Arts. 10, 12, 17, 19

DORA (Level 1) References:

N/A

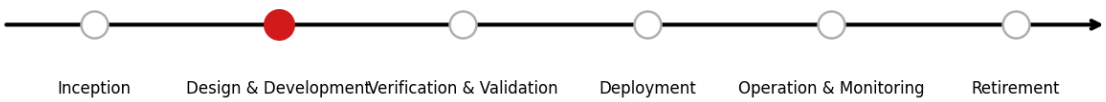
DORA (Level 2) References:

N/A

Re-evaluation triggers:

IG, LR, TD

DQM-06: Data integrity



Lifecycle overview for criteria Lifecycle overview for criteria DQM-06

Evaluation requirement:

The integrity of the data shall be ensured through suitable [measures].

Evaluation method:

Document-based: Verify that methods preserving the integrity of the used data are in place and are being used throughout the whole lifecycle.

Supportive guidance:

Supplementary Definitions:

Integrity: In terms of data, integrity ensures that data remains unchanged, uncorrupted, and true to its original form through collection, storage, transmission and processing.

Exemplary [parameters]:

[measures]: e.g., Hashing, digital signatures, version control, backups.

Additional guidance regarding the requirement:

Data integrity should be ensured through measures such as hashing, digital signatures, version control, and backups.

The following tools can be used for the implementation:
- MLOps can be used to evaluate whether data integrity checks, such as versioning, hashing, and digital signatures, are integrated into the machine learning lifecycle.
- DevOps can be used to assess if CI/CD pipelines incorporate data integrity measures, such as automated checks for corruption and rollback mechanisms.

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

At least partial training was performed.

Reference to EU AI Act:

AI Act Arts. 10, 17, 19, 26

DORA (Level 1) References:

N/A

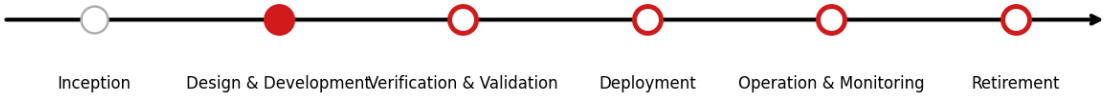
DORA (Level 2) References:

RMF-RTS Art. 1, 2

Re-evaluation triggers:

EG, IG, LR, SN, TD, TS

DQM-07: Data decommissioning



Lifecycle overview for criteria Lifecycle overview for criteria DQM-07

Evaluation requirement:

The data decommissioning process shall be documented.

Evaluation method:

Document-based: Ensure that data can be decommissioned. Verify that the decommission plan is created, reviewed and met.

Supportive guidance:

Supplementary Definitions:

Data decommissioning: The process of securely and systematically retiring, archiving or disposing of data that is no longer needed.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

Depending on the needs and (regulatory) requirements a plan for decommission of datasets shall be in place.
Every decommission plan shall include
-Clear descriptions and means for identification of the datasets that are covered by the decommision plan
-Requirements for the decommission of data, e.g., end of use, privacy/copyright infringements, etc.
-Involved personnel/roles and corresponding responsibilities
-Description of the decommissioning process with the main steps of identification of datasets, archiving of datasets (scope/extent of archiving, storage medium, encryption, etc.) and data deletion (e.g., means of erasure such as digital override or physical destruction of media)

Every conducted data decommission shall be documented with a clear assignment to the datasets that were subject to the decommissioning.
Furthermore, a recovery plan for decommissioned (and archived) data should be implemented if necessary.

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

At least partial training was performed.

Reference to EU AI Act:

AI Act Arts. 10, 11, 17, 18

DORA (Level 1) References:

N/A

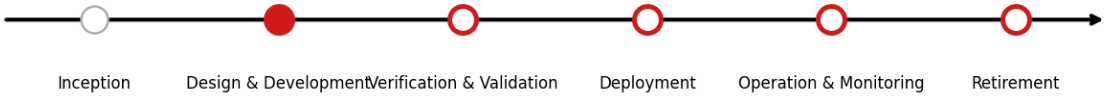
DORA (Level 2) References:

RMF-RTS Art. 11

Re-evaluation triggers:

EG, IG, LR, SN, TD, TS

DQM-08: Users' consent to the use of their data



Lifecycle overview for criteria Lifecycle overview for criteria DQM-08

Evaluation requirement:

It shall be ensured that individuals provide active informed consent for the AI system to process and store their data. It shall be ensured that individuals can perform their right to revoke this consent by allowing users to request the deletion of their data and to stop its processing.

Evaluation method:

Document-based: Verify that individuals actively consent to processing the data and that it is logged appropriately. Ensure that data can be deleted if it is requested by a user.

Supportive guidance:

Supplementary Definitions:

Active informed consent: A type of consent where individuals actively and explicitly agree to the processing and storage of their data after being fully informed about the nature, purpose, and implications of that processing.
Revocation of consent: The ability of individuals to withdraw their previously granted consent at any time, making further processing or storage of their data unlawful unless another legal basis exists.
Right to deletion (Right to be forgotten): The right of individuals to have their personal data erased or deleted when it is no longer necessary for the purposes for which it was collected, or when consent has been withdrawn.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

Individuals should be given clear, understandable information about:
- What data is being processed
- Why it is being processed
- How long it will be stored
- Who will have access to the data
- Any potential risk or consequences

The applicable law for this can be taken from the GDPR.

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

- Personal data was used to train the involved AI models.

Reference to EU AI Act:

AI Act Art. 10

DORA (Level 1) References:

N/A

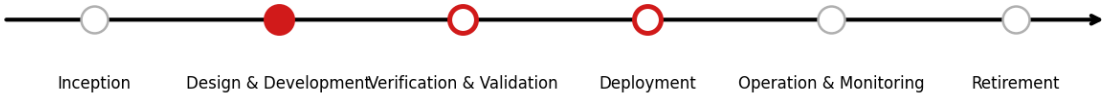
DORA (Level 2) References:

N/A

Re-evaluation triggers:

EG, IG, LR, UG

DQM-09: Users' right to be forgotten



Lifecycle overview for criteria Lifecycle overview for criteria DQM-09

Evaluation requirement:

Measures shall be in place to realize a user's right to be forgotten.

Evaluation method:

Document-based: Verify that a process for users to revoke their consent to usage of their personal data is in place, and verify that there is a process and adequate methods to delete the user's data from datasets and the model.
Additionally, it shall be verified that no information can be extracted by employing privacy attacks (e.g., membership inference, prompt injection) customized to a specific individual.

Supportive guidance:

Supplementary Definitions:

Right to be forgotten:
The right of individuals to have their personal data erased or deleted when it is no longer necessary for the purposes for which it was collected, or when consent has been withdrawn (GDPR Art. 17).

Exemplary [parameters]:

[measures]: e.g., distillation, retraining, federated learning, machine unlearning methods

Additional guidance regarding the requirement:

Not applicable.

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

- Personal data was used to train the involved AI models.

Reference to EU AI Act:

AI Act Art. 10

DORA (Level 1) References:

N/A

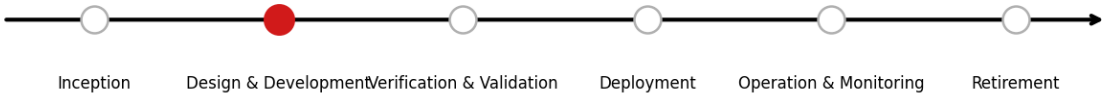
DORA (Level 2) References:

N/A

Re-evaluation triggers:

EG, IG, LR, SN, TS

DQM-10: Data sensitivity and necessity filtering



Lifecycle overview for criteria Lifecycle overview for criteria DQM-10

Evaluation requirement:

Prior to model training, sensitive data that are not essential for the intended purpose of the system shall be removed or (pseudo-)anonymized to ensure compliance with data privacy regulations.

Evaluation method:

Document-based: Verify that a functioning process is in place to remove all unneeded sensitive information from the data.

Supportive guidance:

Supplementary Definitions:

Sensitive data: Data that, if disclosed, could harm or infringe on privacy rights.

Anonymization: The process of irreversibly transforming data so that an individual can no loner be identified, directly or indirectly.

Pseudonymization: The process of replacing identifiable data with pseudonyms or identifiers so that individuals cannot be identified without additional information.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

Examples for sensitive data are among others:
- Personal data such as names, addresses
- Health information
- Financial information (e.g., credit card number)
- Biometric data (e.g., fingerprints, facial recognition)
- Racial or ethnic origin, political opinions, religious beliefs

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

- Personal data was used to train the involved AI models. OR
- At least partial training was performed.

Reference to EU AI Act:

AI Act Art. 10

DORA (Level 1) References:

N/A

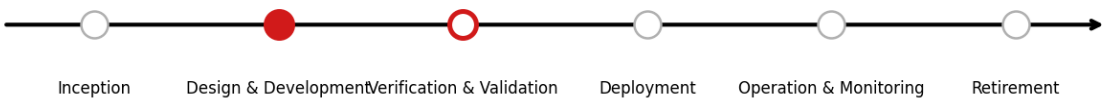
DORA (Level 2) References:

N/A

Re-evaluation triggers:

EG, IG, LR, TD

DQM-11: Documentation of development data splits



Lifecycle overview for criteria Lifecycle overview for criteria DQM-11

Evaluation requirement:

A method for data splitting of the development data into training, validation and test data shall be defined and documented.

Evaluation method:

Document-based: Review the documentation of the chosen method for data splitting, review the given reasoning and validate that the defined method was used to create the respective datasets. Validate that training and test dataset are distinct from each other.

Supportive guidance:

Supplementary Definitions:

Data splitting: Process of partitioning data in portions for model training, validation and testing.
Training set: Used for the model training.
Validation set: Used for hyperparameter tuning.
Test set: Used for (final) evaluation of the model.
Data split ratios: Proportions of data allocated to training, validation, and test sets.
Random splitting: Random division into tain, validation and test set.
Stratification splitting: Splitting data in such a way that data property distributions are equal accross train, validation and test set.
Cross-validation: Splitting the data into multiple subsets and training and testing on diverse combinations of these subsets.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

A suitable data splitting method, e.g., random splitting, stratification splitting or cross-validation, shall be chosen and documented. A reasoning for the chosen method and the data split (ratios) shall be given. Defined data requirements that are relevant to the development data (see DQM-03 & DQM-4) shall also hold for the subsets for training, validation and testing resulting from the splitting process. Under all circumstances, for an unbiased evaluation with the ability to detect potential overfitting and validate the system's ability to generalize, the set of test data shall be distinct from the used training data.

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

At least partial training was performed.

Reference to EU AI Act:

AI Act Arts. 10, 11, 17, 18

DORA (Level 1) References:

N/A

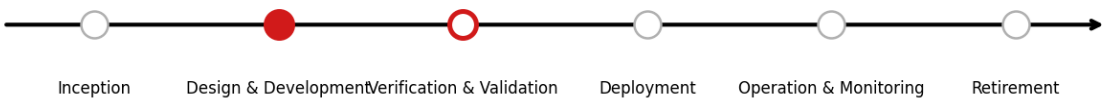
DORA (Level 2) References:

RMF-RTS Art. 8, 16

Re-evaluation triggers:

IG, LR, SN, TD, TS

DQM-12: Documentation of data preprocessing steps



Lifecycle overview for criteria Lifecycle overview for criteria DQM-12

Evaluation requirement:

The data preparation process shall be documented.

Evaluation method:

Document-based: Verify that there is documentation describing the process of the (pre)processing of data. Ensure that it is adequately documented, see supporting guidance.

Supportive guidance:

Supplementary Definitions:

Data Preparation: The process of cleaning, transforming, and organizing raw data into a format suitable for analysis, modeling, or machine learning. This step ensures that data is of high quality and ready for use.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

Data preparation process should consider:
- Dataset composition;
- Data labelling;
- Data annotation;
- Data quality assessment relative to the data quality measure targets established in the data requirements process;
- Data quality improvement such as data cleaning; data standardization, data normalization, data imputation
- Data de-identification to protect privacy;
- Data encoding;
- Individual processing steps such as conversions, transformations, aggregations, normalization, format conversions, feature calculation, conversion of numerical data into categories;
- Filtering, e.g.: detection and processing values with different measurement scales;
- Documentation of assumptions made, in particular with regard to the information that the data should measure and represent;

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

At least partial training was performed.

Reference to EU AI Act:

AI Act Arts. 10, 11, 17, 18

DORA (Level 1) References:

N/A

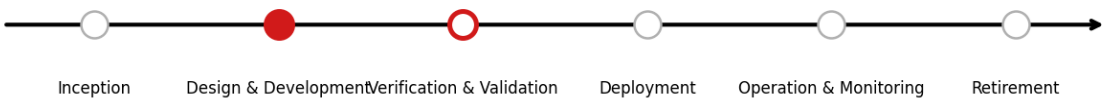
DORA (Level 2) References:

N/A

Re-evaluation triggers:

IG, LR, SN, TD, TS

DQM-13: Data provisioning



Lifecycle overview for criteria Lifecycle overview for criteria DQM-13

Evaluation requirement:

The data provisioning process shall be documented.

Evaluation method:

Document-based: Ensure that the data requirements were met. Verify that the data is available for the model and that the data quality is consistent when transferring data.

Supportive guidance:

Supplementary Definitions:

Data provisioning: The process of making data available to users, applications, or systems in a secure, efficient, and controlled manner. This can involve extracting, preparing, and delivering data from various sources to its intended destination.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

Considerations should be given to:
Data requirements should be met before provisioning.
Data provisioning includes:
- Ensuring data quality when transferring or moving the data from one system to another;
- Making data available for the model.

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

At least partial training was performed.

Reference to EU AI Act:

AI Act Arts. 10, 11, 17, 18

DORA (Level 1) References:

N/A

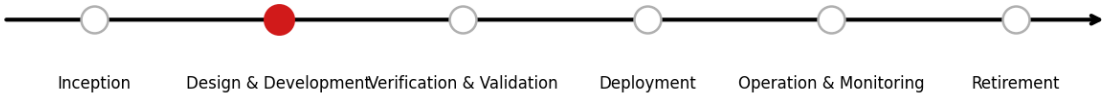
DORA (Level 2) References:

N/A

Re-evaluation triggers:

IG, LR, SN, TD, TS

DQM-14: Data characteristics



Lifecycle overview for criteria Lifecycle overview for criteria DQM-14

Evaluation requirement:

Relevant data characteristics shall be documented.

Evaluation method:

Document-based: Verify that data characteristics are documented in accordance to the supportive guidance.

Supportive guidance:

Supplementary Definitions:

Data characteristics: Attributes or properties that describe the nature, structure, quality, and behavior of data. Documenting these characteristics helps ensure that data is well-understood, appropriately used, and suitable for its intended purpose.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

Some factors to consider in the documentation:
- Its dynamics: e.g., static, dynamic with occasional updates, dynamic real-time updates;
- Its scale: e.g., very large (Exabyte), Large (tens of petabyte), Medium (hundreds of GB), Small (tens of GB or smaller);
- Its structure: e.g., unstructured, semi-structured, structured, complex structured data;
- Its format: e.g. CSV, JSON, XML etc.;
- Its privacy risks: e.g., identified data, pseudonymised data, unlinked pseudonymised data, anonymised data, aggregated data;
- Its sensitivity, e.g. protected, confidential, public;
- Its degree of standardization: standardized data format, non-standardized data format, standardized dataset metadate, non-standardized dataset;
- The feature's data types: e.g. Tabular static numerical data, Temporal (longitudinal) data, image data, text data etc.;
- Dependencies between features, especially in tabular data;
- Its metadata.

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 10, 11, 13, 17, 18, 26

DORA (Level 1) References:

N/A

DORA (Level 2) References:

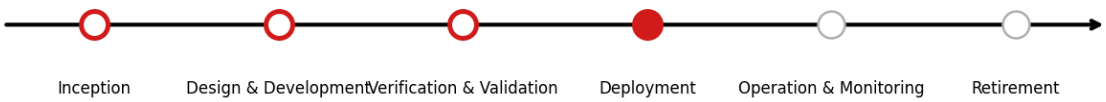
N/A

Re-evaluation triggers:

IG, LR, SN, TD, TS

9 Development

DEV-01: Technical model documentation



Lifecycle overview for criteria Lifecycle overview for criteria DEV-01

Evaluation requirement:

A description of the AI system shall be provided including relevant technical conditions.

Evaluation method:

Document-based: A description of the AI system is provided, including the aspects stated in the supportive guidance.

Supportive guidance:

Supplementary Definitions:

Not applicable.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

Aspects to consider in the documentation:
- Initial description of the AI system including relevant technical conditions and circumstances;
- Objective and scope definition;
- Intended usage environment for the system;
- Expected output ranges of the system;
- Training approach (central or federated);
- Type of decision making (probabilistic, deterministic);
- System architecture;
- System requirements;
- Relevant developer choices;
- Input data requirements;
- Expected ranges of the output;
- Development and lifecycle process;
- Scale of current deployment: pilot, narrow, broad, widespread;
- Relevant dependencies of the system on other models not directly developed or data not directly used;
- The versions of relevant software or firmware, and any requirements related to version updates;
- Information about the model(s) used in the system;
- The description of the hardware on which the AI system is intended to run.

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 10, 11, 13, 17, 18

DORA (Level 1) References:

N/A

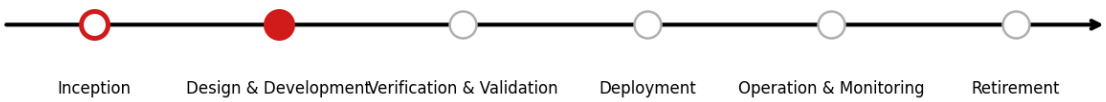
DORA (Level 2) References:

RMF-RTS Art. 4, 5, 16

Re-evaluation triggers:

AC, MA, MP, SE, SR, TD, UG

DEV-02: Documentation of pretrained model usage



Lifecycle overview for criteria Lifecycle overview for criteria DEV-02

Evaluation requirement:

If a pretrained model is used, its usage shall be justified and documented.

Evaluation method:

Document-based: The documentation must justify, or reference, why in general a pretrained model is chosen for the use case and a justification for the selected pretrained model shall be given. This reasoning shall include the evaluation of the Service Level Agreement as stated in the supportive guidance.

Supportive guidance:

Supplementary Definitions:

Not applicable.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

The documentation should at least cover:
- Why pretraining is appropriate for the task,
- An evaluation of the Service Level Agreement (if applicable) focusing on Availability (Ensuring the systems are reliably accessible as per agreed-upon terms), Scalability (Confirming the infrastructure can handle increased loads and scale as needed), Performance (Assessing the efficiency and responsiveness of the systems under various conditions).
- An exit and continuity strategy outlining how critical functionality will be maintained if the pretrained-model provider experiences a prolonged outage or ends the service.

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

- The architecture is not or only partially known.
- There is only partial or zero knowledge of parameters of the involved AI models.
- The training of at least one integrated model was performed partially or not at all.

Reference to EU AI Act:

AI Act Arts. 10, 11, 13, 17, 18

DORA (Level 1) References:

N/A

DORA (Level 2) References:

N/A

Re-evaluation triggers:

AC, IG, LR, MA, SE

DEV-03: Model development policy

Lifecycle overview for criteria Lifecycle overview for criteria DEV-03

Evaluation requirement:

Model development policies and instructions shall be documented.

Evaluation method:

Document-based: Verify completeness of the documentation for model development policies and instructions.

Supportive guidance:

Supplementary Definitions:

Not applicable.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

The development policies and instructions should at least cover:
- Requirements for training material (datasets, guides needed for system training);
- System robustness;
- Design;
- Development;
- Deployment;
- Verification (application tests, code review etc.);
- Validation;
- Relevant subsystems;
- Review and approval procedures by management.

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 11, 13, 17, 18

DORA (Level 1) References:

N/A

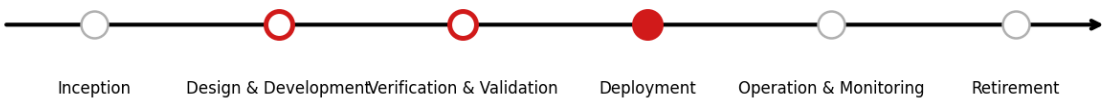
DORA (Level 2) References:

RMF-RTS Art. 15, 16

Re-evaluation triggers:

IG, LR, MA, SE

DEV-04: Runtime environment specification



Lifecycle overview for criteria Lifecycle overview for criteria DEV-04

Evaluation requirement:

The necessary runtime environment shall be determined and documented regarding hardware and software, to ensure compatibility and performance.

Evaluation method:

Document-based: Verify that the hardware and software runtime environment has been identified, as described in the supportive guidance and documented accordingly.

Supportive guidance:

Supplementary Definitions:

Runtime Environment:
The combination of hardware and software configurations required to execute a system or application, including operating systems, frameworks, libraries, and physical devices essential for proper functionality and performance.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

Aspects to consider regarding:
- Hardware: Screen size, screen resolution, storage, network connection;
- Software: Operating system, browser, run-time environments such as Java Run-time Environment or .NET, SOUP (Software of Unknown Provenance, including libraries and frameworks)
- Package dependencies: requirements.txt, dockerfiles

Tool which could be used during implementation:
Hardware Detection & enumeration:
- HWInfo
- OS-tools (msinfo32, lspci)

SBOM generation tools:
- Syft
- SPDX
- CycloneDX

Code dependency Tracker:
- dependency-track

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 11, 13, 17, 18

DORA (Level 1) References:

N/A

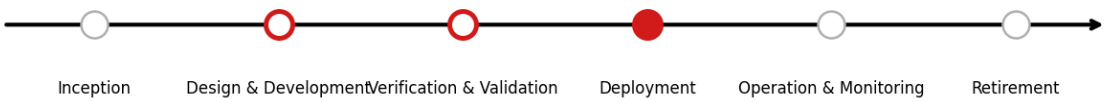
DORA (Level 2) References:

N/A

Re-evaluation triggers:

EM, SE, SR, UP, UR

DEV-05: Model logging and versioning



Lifecycle overview for criteria Lifecycle overview for criteria DEV-05

Evaluation requirement:

All model executions, whether for development, testing, validation, or using third-party models, shall be versioned and traceable, with sufficient metadata, configurations, datasets, and artifacts recorded to enable reproducible results.

Evaluation method:

Document-based: Verify that all model executions for development, testing, validation, corresponding metadata and configurations files, the used data, and all artifacts (performance logs and reports, docker container images) are stored and that there is a consistent versioning scheme in place.

Supportive guidance:

Supplementary Definitions:

Not applicable.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

- Model identification: Record provider, exact model name/version (for LLMs e.g., gpt-4o-2024-12, gpt-5-preview123, claude-3.5-sonnet, llama-3-70b), SDK/library version, and runtime parameters (e.g., "temperature"). For internals model introduce consistent identification scheme (e.g. based on architecture "CNN-51Layer-ReLU")
- Data & configuration tracking: Version datasets, code, configuration files, preprocessing steps, and environment details (OS, dependencies, hardware).
- Run logging: For all training, test, and validation runs, log unique run IDs, dataset versions, model checkpoints, hyperparameters, seeds, metrics, and associated artifacts.
- Reproducibility: Define acceptable tolerance levels for reruns and maintain the ability to replay a run from stored versions.
- Versioning policy: Assign a new version to every trained or updated model; clearly distinguish pre-production from production; avoid floating “latest” tags.
- Production traceability: For deployed models, including third-party endpoints, log model version and configuration to enable post-hoc analysis.

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 11, 13, 17, 18

DORA (Level 1) References:

N/A

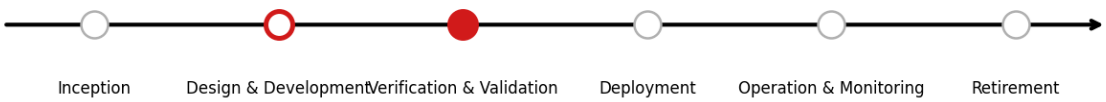
DORA (Level 2) References:

RMF-RTS Art. 15, 16, 17

Re-evaluation triggers:

EM, IG, LR, MA, MP, TD, UP, UR

DEV-06: Model selection process



Lifecycle overview for criteria Lifecycle overview for criteria DEV-06

Evaluation requirement:

Where appropriate, multiple model types shall be trained and compared and the final model choice shall be justified.

Evaluation method:

Document-based: Verify that a set of possible model candidates was defined. The different models should be trained. Check that the final choice of model is justified by comparing, among other things, correctness, robustness and explainability.

Supportive guidance:

Supplementary Definitions:

Multiple Models:
The practice of training and evaluating different machine learning algorithms or architectures on the same task to compare their performance, robustness, and suitability before selecting the final model.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

Define metrics and compare e.g., regarding:
- Correctness;
- Robustness;
- Explainability.

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 12, 19

DORA (Level 1) References:

N/A

DORA (Level 2) References:

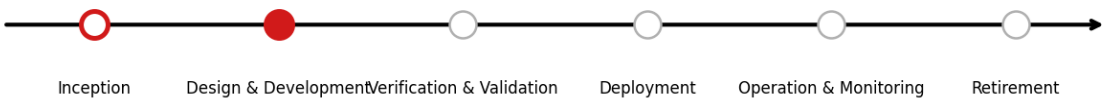
N/A

Re-evaluation triggers:

MA, SN, TS, UP, UR

10 Fairness

FAIR-01: Potential bias and impact assessment



Lifecycle overview for criteria Lifecycle overview for criteria FAIR-01

Evaluation requirement:

Sensitive features shall be identified and potential biases against individuals or groups of individuals shall be assessed and documented.

Evaluation method:

Document-based: Review the documentation whether the user demographic is correctly identified. Verify that all of the potential biases were considered during the bias analysis and that sensitive features were identified.

Supportive guidance:

Supplementary Definitions:

A selection of potential biases:
- Individual bias = The tendency of an AI system to make skewed decisions based on patterns learned from biased data related to individual preferences or characteristics, e.g. when a person with a university degree from a less known university applies to a job that is not recognized by an AI system.
- Group bias = The tendency of an AI system to favor or disadvantage certain groups based on patterns learned from biased data associated with group characteristics such as gender, ethnicity, or age.
- Automation bias = Occurs when a human decision-maker prefers recommendations made by an automated decision-making system, even when the system makes errors, over non-automated information.
- Group attribution bias = Occurs when a human assumes that what is true for one individual or object is also true for all individuals or objects in the group.
- Sampling bias = Occurs when data records are not collected randomly from the intended population.
- Coverage bias = Happens when the population represented in a dataset does not match the population that the machine learning model is making predictions about.
- Non-normality bias = Arises when statistical methods are applied to non-normally distributed data.
- Simpson's paradox = Manifests when a trend indicated in individual groups of data reverses when the groups of data are combined.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

In order to identify potential biases, the context, purpose and relevant protected attributes for the use case should be considered.

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

- Personal data was used to train the involved AI models OR
- The model has a societal impact.

Reference to EU AI Act:

AI Act Arts. 9, 10, 11, 15, 17, 18

DORA (Level 1) References:

N/A

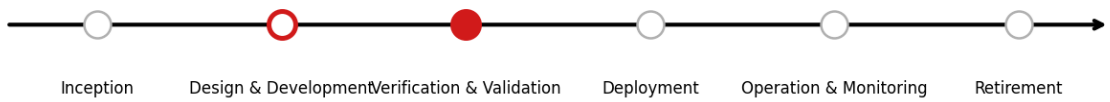
DORA (Level 2) References:

N/A

Re-evaluation triggers:

EG, IG, LR, SI, TD, UG

FAIR-02: Effective bias assessment



Lifecycle overview for criteria Lifecycle overview for criteria FAIR-02

Evaluation requirement:

The level of effective bias with respect to identified sensitive features in the data shall be assessed with [measures].
The selection of [measures] and reason for their selection shall be included in the system description.

Evaluation method:

Test-based: Assess whether sensitive features have been identified and are complete.
Check the system description and/or the source code whether appropriate measures have been identified and correctly used to measure and/or mitigate the potential bias or fairness problems.

Supportive guidance:

Supplementary Definitions:

Sensitive features: features that were identified during the assessment in FAIR-01, e.g., gender, ethnicity, religion etc.

Exemplary [parameters]:

[measures]: e.g., Absolute Standardized Mean Difference (ASMD), Adverse Impact Ratio

Additional guidance regarding the requirement:

Not applicable.

Evaluation tools:

- Aequitas - can be used as an audit tool to check if fairness reports match expected demographic group fairness assessments.
- IBM AI Fairness 360 can be used for independent bias analysis to cross-check whether the AI system's bias detection aligns with industry standards.

Relevance based on use case parametrization:

- Personal data was used to train the involved AI models OR
- The model has a societal impact.

Reference to EU AI Act:

AI Act Arts. 9, 10, 11, 13, 15, 17, 18

DORA (Level 1) References:

N/A

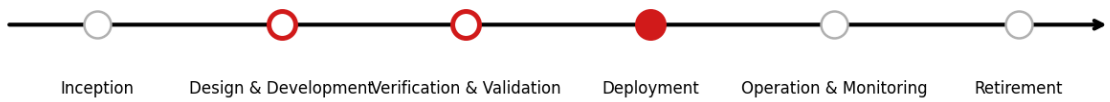
DORA (Level 2) References:

N/A

Re-evaluation triggers:

EG, IG, LR, SI, SN, TD, UG

FAIR-03: Bias mitigation



Lifecycle overview for criteria Lifecycle overview for criteria FAIR-03

Evaluation requirement:

If a critical level of bias is detected, the system shall test and implement a [mitigation measure]. Results should be compared using both [bias measures] and [performance measures]. Bias that is considered critical for functionality shall be documented in the system description.

Evaluation method:

Test-based: The detection mechanism must be in place and it must trigger the mitigation strategy, once the threshold is reached. The mitigation must now adjust aspects of the system to balance the output.

Supportive guidance:

Supplementary Definitions:

Critical level of bias:
A threshold or degree of bias that, if exceeded, poses a significant risk to fairness, functionality, or compliance with ethical and legal standards.

Exemplary [parameters]:

[mitigation measure]: e.g., Post-Stratification, Covariate Balancing Propensity Score, Inverse Propensity Score Weighting.
[bias measure]: see FAIR-02.
[performance measures]: see MON-01.

Additional guidance regarding the requirement:

Not applicable.

Evaluation tools:

- Fair Training methods, e.g. "Learning Certified Individually Fair Representations" - can be used to verify whether fairness training methods work by evaluating formal fairness guarantees on specific test samples.

- balance python package - the implemented diagnostics and evaluation methods such as statistical summaries of weight and covariates distributions can be used to indentify potential biases.

- AWS Sagemaker Clarify - the implemented methods and reporting can be used for pre- and post-training data and model training Bias detection.

Relevance based on use case parametrization:

- Personal data was used to train the involved AI models OR
- The model has a societal impact.

Reference to EU AI Act:

AI Act Arts. 9, 10, 11, 13, 15, 17, 18

DORA (Level 1) References:

N/A

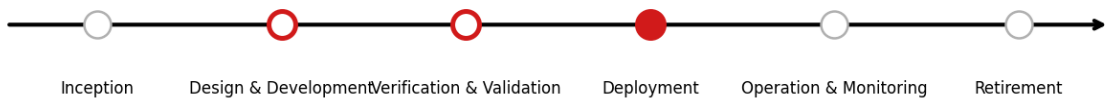
DORA (Level 2) References:

N/A

Re-evaluation triggers:

EG, IG, LR, SI, SN, UG, UP, UR

FAIR-04: Accessibility



Lifecycle overview for criteria Lifecycle overview for criteria FAIR-04

Evaluation requirement:

The AI system shall ensure equal accessibility and inclusivity for relevant user groups.

Evaluation method:

Test-based: Assess whether the potential user groups have been identified and whether any potential accessibility or inclusivity challenges have been recognized.
Evaluate whether the system's design ensures accessibility for all identified user groups, including those with disabilities and diverse language needs.

Supportive guidance:

Supplementary Definitions:

All user groups:
All people including marginalized groups, people with disabilities (e.g., accessible to screen readers, including alt text for images, colour-blind friendly palettes, etc.).

Accessibility:
The system should be operable by people with a wide range of impairments, including visual impairments, hearing impairments and physical impairments; it should support multiple languages.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

Analysis shall be conducted for all user groups. Most relevant to customer facing products and systems.

Evaluation tools:

- WAVE - can be used to check if the system is compliant by evaluating screen reader compatibility, contrast ratios, and missing alt text.
- Google Lighthouse - can be used to assess accessibility performance using automated audits, focusing on usability, contrast, ARIA attributes, and keyboard navigation.

Relevance based on use case parametrization:

- The system interacts directly with external consumers OR
- The system integrates a human (either human-on-the-loop, human-in-the-loop or human control).

Reference to EU AI Act:

AI Act Arts. 9, 11, 13, 18

DORA (Level 1) References:

N/A

DORA (Level 2) References:

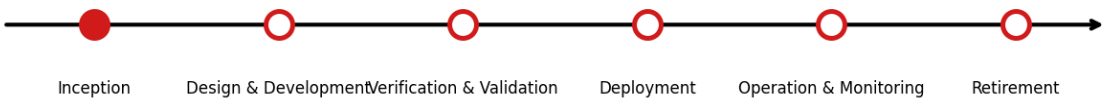
N/A

Re-evaluation triggers:

AC, IG, LR, UG

11 Governance

GOV-01: Periodic review of the (internal) AI policy



Lifecycle overview for criteria Lifecycle overview for criteria GOV-01

Evaluation requirement:

An AI policy shall be established and reviewed periodically to ensure its continuing suitability, adequacy and effectiveness.

Evaluation method:

Document-based: Verify that AI policies are established and communicated effectively.
Evaluate the process for periodic reviews and assess its effectiveness

Supportive guidance:

Supplementary Definitions:

AI policy:
A set of guidelines and principles designed to govern the development, deployment, and use of artificial intelligence systems.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

The goal of the AI policiy is to ensure ethical, legal, and operational compliance, aligning AI practices with organizational goals and societal standards. Regular reviews help maintain its relevance and effectiveness in addressing emerging risks, technologies, and regulatory changes.

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 9, 17

DORA (Level 1) References:

N/A

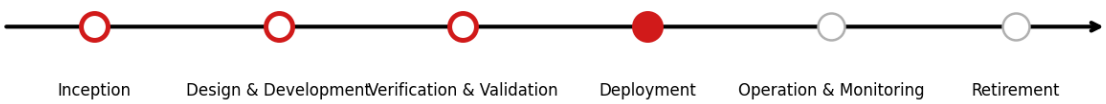
DORA (Level 2) References:

N/A

Re-evaluation triggers:

IG, SE, SR

GOV-02: Establishment of AI management system and regular audits



Lifecycle overview for criteria Lifecycle overview for criteria GOV-02

Evaluation requirement:

A Quality Management System shall be established and audited (internal or external) on a regular basis.

Evaluation method:

Document-based: Verify whether an AI Management System or a QMS that includes the AI system is established.
Check that the review process of this Q/MS is effective and appropriate.

Supportive guidance:

Supplementary Definitions:

Quality Management System (QMS):
A structured framework that is concerned with the implementation, monitoring, and governance of AI systems.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

The QMS shall include at least:
- The name and address of the provider and, if the application is lodged by an authorised representative, also their name and address;
- The list of AI systems covered under the same quality management system;
- The technical documentation for each AI system covered under the same quality management system
- A description of the procedures in place to ensure that the quality management system remains adequate and effective

The review of the QMS shall include at least:
- The changes, if any, since the last review;
- Changes in the AI system that are relevant to the AI management system;
- Changes in the needs and expectations of the AI system that are relevant to the AI management system;
- Information on the performance of the AI system.

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Art. 17

DORA (Level 1) References:

N/A

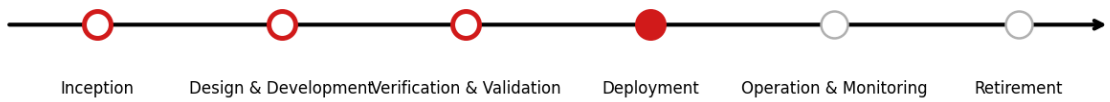
DORA (Level 2) References:

N/A

Re-evaluation triggers:

EM, IG, LR, SE, SR, UP, UR

GOV-03: Attacks with GenAI



Lifecycle overview for criteria Lifecycle overview for criteria GOV-03

Evaluation requirement:

The potential risks associated with the use of General Purpose AI (GPAI) / Generative AI (GenAI) shall be assessed and incorporated into the design of incident response and defense strategies.

Evaluation method:

Document-based: Verify that the threat of GPAI/GenAI has been considered in the design of incidents and defence strategies.
Ensure that the considerations have been properly documented.

Supportive guidance:

Supplementary Definitions:

Generative AI (GenAI):
A type of artificial intelligence that creates new content, such as text, images, or audio, by learning patterns from existing data.

General Purpose AI (GPAI):
Artificial intelligence systems designed to perform a wide range of tasks across various domains. These systems are adaptable and can solve diverse problems using the same core model, making them versatile for multiple applications.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

It could be considered, e.g., how GenAI could be used for attacks on the business's customers or clients through spoofing or GenAI generated content etc.
The OWASP LLM AI Cybersecurity & Governance Checklist can be consulted for more information.

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

- The system integrates at least one Generative AI Foundation Model.

Reference to EU AI Act:

AI Act Arts. 9, 17

DORA (Level 1) References:

DORA Art. 8

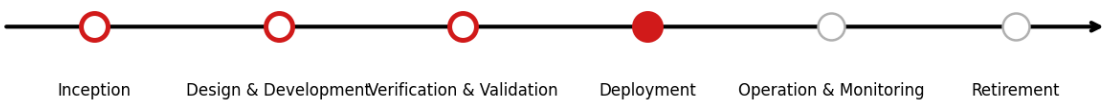
DORA (Level 2) References:

N/A

Re-evaluation triggers:

AT, LR, MA, NV, SN, TS

GOV-04: Compliance of the AI service with general cloud computing compliance criteria



Lifecycle overview for criteria Lifecycle overview for criteria GOV-04

Evaluation requirement:

The system should be compliant with general cloud computing compliance criteria, as set out in the Cloud Computing Compliance Criteria Catalogue (C5)

Evaluation method:

Document-based: Verify that the system is compliant to necessary standards and regulations.

Supportive guidance:

Supplementary Definitions:

The Cloud Computing Compliance Criteria Catalogue (C5) by BSI outlines minimum security requirements for cloud applications.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

Complying with the C5 Catalogue enhances trust and credibility in your system by demonstrating adherence to high-security standards. It strengthens market competitiveness. Additionally, compliance facilitates alignment with international frameworks, such as ISO 27001 and SOC 2, simplifying global standardization efforts.
By adhering to the C5 requirements, organizations can effectively mitigate risks associated with cyberattacks, operational disruptions, and reputational damage. Conversely, non-compliance may indirectly lead to violations of GDPR or industry-specific regulations, resulting in significant financial penalties and legal consequences.

A Tool which could be used during implementation is clouditor, an assurance tool that checks whether cloud-based services and applications are securely configured

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

- The system is built/operated on a hybrid infrastructure.

Reference to EU AI Act:

AI Act Art. 17

DORA (Level 1) References:

N/A

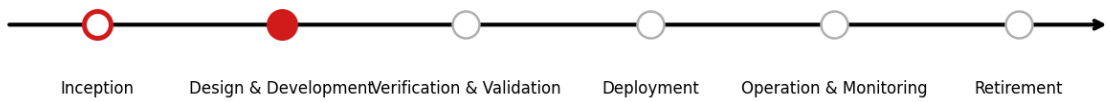
DORA (Level 2) References:

N/A

Re-evaluation triggers:

IG, LR, SE, SN, SR

GOV-05: Impact on IP rights



Lifecycle overview for criteria Lifecycle overview for criteria GOV-05

Evaluation requirement:

All possibilities for potential impact on IP rights shall be considered and documented.

Evaluation method:

Document-based: Verify that an analysis for possible IP right infringements has been performed.
Evaluate if appropriate measures against the possible infringements have been taken.

Supportive guidance:

Supplementary Definitions:

IP (Intellectual property):
Concerns content that is subject to copyright, trademark, or patent protection.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

Considerations for potential impact on IP rights are ownership, licensing, copyright, patents, trademarks, and trade secrets throughout the lifecycle of the system.

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 9, 17

DORA (Level 1) References:

N/A

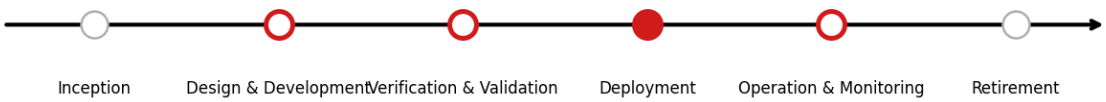
DORA (Level 2) References:

N/A

Re-evaluation triggers:

IG, LR

GOV-06: System description/User guidance



Lifecycle overview for criteria Lifecycle overview for criteria GOV-06

Evaluation requirement:

A system description shall be created and made available to the user.

Evaluation method:

Document-based: Verify that a system description including relevant aspects regarding the system is in place and made available to the user.

Supportive guidance:

Supplementary Definitions:

System description:
A detailed document ensuring transparency and ease of understanding.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

The system description should include e.g.,:
- System overview: brief introduction to the system, purpose and intended use cases, key features and functionalities
- Technical Architecture: Description of the used infrastructure, components of the system, technologies and frameworks used
- User interface description
- User instructions: instructions on how to use the system, examples of input data and expected outputs
- Troubleshooting and FAQs
- Data Management: information on data input formats and sources, data storage, security measures, and privacy considerations
- System limitations
- System evaluations: Levels of performance, Performance metrics used, Bias detection metrics used, detected Biases that are considered critical for functionality
- Compliance and ethical considerations

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 11, 13, 17, 18

DORA (Level 1) References:

N/A

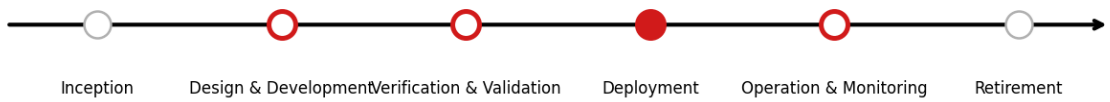
DORA (Level 2) References:

RMF-RTS Art. 16

Re-evaluation triggers:

AC, MA, MP, SE, SR, TD, UG

GOV-07: Qualitative description of system



Lifecycle overview for criteria Lifecycle overview for criteria GOV-07

Evaluation requirement:

The AI system and it's intended purpose shall be described including relevant qualitative aspects.

Evaluation method:

Document-based: Verify that a complete and comprehensive description of the AI system is provided. Including its intended function/purpose, the user group, user's competences, etc.. See supportive guidance for more information.

Supportive guidance:

Supplementary Definitions:

Intended purpose:
The specific goal or function that the system is designed to achieve. It defines why something was created or implemented and what outcomes are expected from its use.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

The AI system and its intended purpose are clearly documented to ensure a comprehensive understanding.

Relevant qualitative aspects to consider in the documentation:
- Intended function/purpose;
- User group;
- The user's existing skills;
- The user's needs;
- Needed user's competency;
- The user's potential interaction with the system;
- Business function;
- The system's impacts on critical functions and activities of the organization;
- Specific risks of harm that can have an impact on natural persons;
- Implemented human oversight measures;
- Arrangements for internal governance;
- Established complaint mechanisms.

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 11, 13, 17, 18

DORA (Level 1) References:

N/A

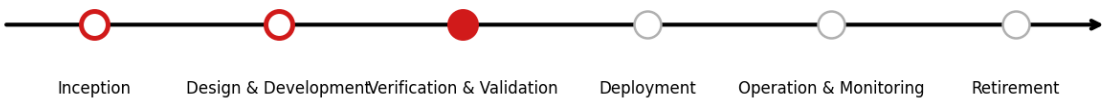
DORA (Level 2) References:

N/A

Re-evaluation triggers:

AC, MA, MP, SE, SR, TD, UG

GOV-08: Alternatives to ML approach



Lifecycle overview for criteria Lifecycle overview for criteria GOV-08

Evaluation requirement:

The alternatives to AI shall be listed and evaluated regarding utility, security, and performance. The superiority of ML methods shall be justified while taking its associated risks into account.

Evaluation method:

Document-based: Verify that an analysis was performed to identify and compare possible alternatives to AI.

Supportive guidance:

Supplementary Definitions:

Machine Learning (ML):
A type of AI where systems learn from data to make predictions or decisions without explicit programming. It includes methods like supervised, unsupervised, and reinforcement learning.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

Alternatives to ML can, e.g., be rule-based systems, statistical methods, optimization algorithms, hardcoded algorithms, heuristic methods, simulation methods, human decision-making.

Considerations should be given at least to:
- Utility;
- Security;
- Performance.

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 9, 17

DORA (Level 1) References:

N/A

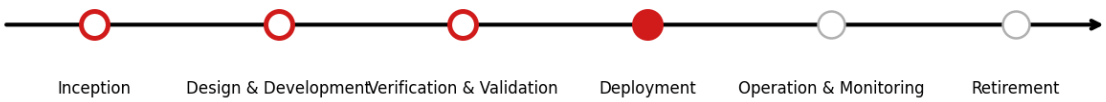
DORA (Level 2) References:

N/A

Re-evaluation triggers:

IG, LR

GOV-09: Catalogue of AI services and tools



Lifecycle overview for criteria Lifecycle overview for criteria GOV-09

Evaluation requirement:

An asset inventory including all internally developed, external or third-party solutions shall be conducted for the overall system.

Evaluation method:

Document-based: Verify that the asset inventory contains all relevant information, see supportive guidance.

Supportive guidance:

Supplementary Definitions:

Asset inventory:
A detailed list of all resources, components, or items owned or used within a system. It includes physical assets, software, data, and third-party solutions.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

The information provided for each module should include at least:
- Its core application area;
- Its use;
- Its appearance
in the overall system.

Tools which could be used for implementation:
SBOM generation tools:
- Syft
- SPDX
- CycloneDX

Code dependency Tracker:
- dependency-track

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 11, 13, 17, 18

DORA (Level 1) References:

N/A

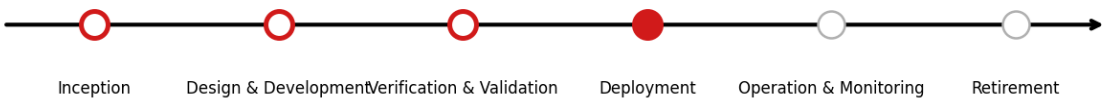
DORA (Level 2) References:

RMF-RTS Art 4

Re-evaluation triggers:

AT, EG, IG, LR, MA, NV, SN, TS

GOV-10: Impact Analysis for ML model inventory



Lifecycle overview for criteria Lifecycle overview for criteria GOV-10

Evaluation requirement:

For each asset in the ML model inventory, an impact analysis shall be performed to determine the impact if the model is compromised.

Evaluation method:

Document-based: Verify that each asset in the ML model inventory has been assessed with an impact analysis for the event of a compromised model.

Supportive guidance:

Supplementary Definitions:

Impact analysis:
An assessment of how changes or risks affect a system's functionality, security, and performance, helping identify risks and plan mitigations.
Compromised:
An Attacker has found a weakness in the AI model or system that can be abused.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

Possible impacts include:
- Data breaches;
- Model manipulation;
- Operational disruption;
- Financial loss

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 9, 17

DORA (Level 1) References:

DORA Art. 11

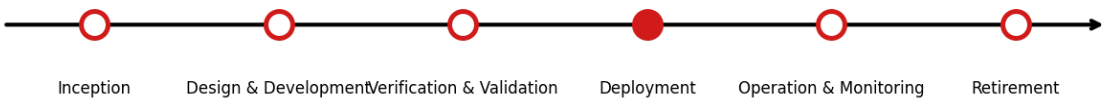
DORA (Level 2) References:

RMF-RTS Art. 25, 26

Re-evaluation triggers:

AT, EG, IG, LR, NV, SN, TS

GOV-11: Human resource competencies



Lifecycle overview for criteria Lifecycle overview for criteria GOV-11

Evaluation requirement:

Information on all persons involved in the life cycles of the AI system shall be documented. Their competencies shall be ensured and documented through the necessary qualifications and trainings.

Evaluation method:

Document-based: The necessary (personnel) roles for the AI lifecycle are defined including their necessary qualification.
Assess whether the personnel has the necessary qualifications corresponding to their role, e.g. by reviewing certificates, project experiences or evidence for related trainings.

Supportive guidance:

Supplementary Definitions:

Life cycle:
The process from design and development through deployment, operation, and eventual decommissioning.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

Involved personnel includes, among others, the people who
- have participated in the various stages of development, deployment, operation, change management, maintenance, transfer and decommissioning, continuous improvement of the AI system, as well as verification and integration of the AI system;
- are involved in the decision-making process (e.g. in the form of human-in-the-loop, human-on-the-loop, human-in-command).

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 4, 17, 26

DORA (Level 1) References:

N/A

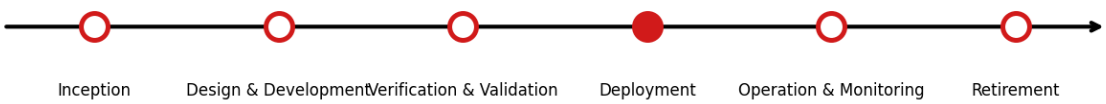
DORA (Level 2) References:

N/A

Re-evaluation triggers:

EM, IG, LR, MA, MP, SE, SR, TD, UG, UP, UR

GOV-12: Allocation of responsibilities among all involved parties



Lifecycle overview for criteria Lifecycle overview for criteria GOV-12

Evaluation requirement:

It shall be ensured and documented how responsibilities within the life cycle of the AI system are allocated between the organization, its partners, suppliers, customers and third parties.

Evaluation method:

Document-based: Verify that all involved stakeholders in the full AI lifecycle are identified.
Assess whether their responsibilities are clearly described and communicated.

Supportive guidance:

Supplementary Definitions:

Life cycle:
The process from design and development through deployment, operation, and eventual decommissioning.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

Defining roles and responsibilities is critical for ensuring accountability throughout the organization for its role with respect to the AI system throughout its life cycle.
Examples of areas that can require defined roles and responsibilities can include:
- Risk management;
- AI system impact assessments;
- Asset and resource management;
- Security;
- Safety;
- Privacy;
- Development;
- Performance;
- Human oversight;
- Supplier relationships;
- Demonstrate its ability to consistently fulfil legal requirements;
- Data quality management (during the whole life cycle).

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 17, 25

DORA (Level 1) References:

DORA Art. 2

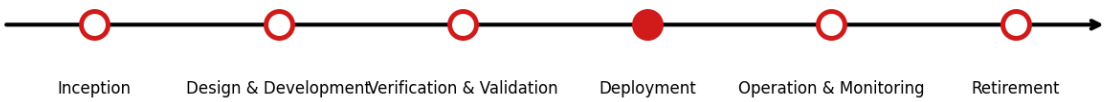
DORA (Level 2) References:

RMF-RTS Art. 15

Re-evaluation triggers:

IG, LR, SE, SR

GOV-13: AI security awareness training



Lifecycle overview for criteria Lifecycle overview for criteria GOV-13

Evaluation requirement:

Educational initiatives and practical exercises to enhance cybersecurity skills and raise awareness about risks along the entire lifecycle and supply chain of the AI systems for all AI stakeholders shall be conducted.

Evaluation method:

Document-based: Verify that AI stakeholders are educated with initiatives and practical exercises to enhance cybersecurity skills throughout the AI lifecycle.

Supportive guidance:

Supplementary Definitions:

Supply chain:
The network of entities and processes involved in creating, delivering, and maintaining a system.

Life cycle:
The process from design and development through deployment, operation, and eventual decommissioning.

Exemplary [parameters]:

Not applicable-

Additional guidance regarding the requirement:

Relevant parties should have been identified in GOV-12.

Educationak initiatives include, e.g., Workshops and training sessions (internal or external), certification programs and cybersecurity awareness campaigns.
In addition, practical exercises can be, e.g., phishing simulations, incident response drills, penetratin testing labs, data recovery scenarios and secure coding challanges.

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 4, 26

DORA (Level 1) References:

DORA Art. 13

DORA (Level 2) References:

N/A

Re-evaluation triggers:

EM, IG, LR, SE, SR

12 Human Oversight

HO-01: Human oversight

Lifecycle overview for criteria Lifecycle overview for criteria HO-01

Evaluation requirement:

The system shall be monitored through human oversight.

Evaluation method:

Document-based: Verify that a process is in place for a human to monitor the system.

Supportive guidance:

Supplementary Definitions:

Human oversight:
The process in which humans closely monitor AI systems and intervene when necessary to ensure safe, ethical, and compliant operations.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

Especially meaningful decisions that have a direct impact on people should include effective human oversight.
Human oversight can help with detecting anomalies beyond automation, responding in unforeseen scenarios, contextual decision-making, improving system performance and providing accountability.

Evaluation tools:

Not applicable

Relevance based on use case parametrization:

- The AI system operates autonomously with no human oversight OR
- The model has a societal impact.

Reference to EU AI Act:

AI Act Arts. 13, 14, 26

DORA (Level 1) References:

N/A

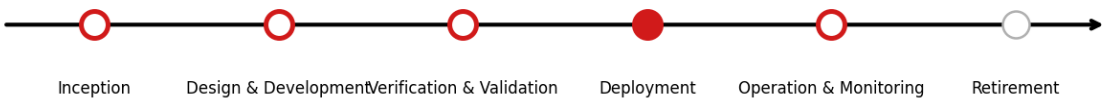
DORA (Level 2) References:

N/A

Re-evaluation triggers:

AC, EM, IG, LR, SN, UR

HO-02: Modification of automated decisions



Lifecycle overview for criteria Lifecycle overview for criteria HO-02

Evaluation requirement:

In the case of automated decision-making, [procedures] shall be put in place to allow authorized personnel and/or users of the AI service to override the system.

Evaluation method:

Test-based: Verify that a procedure is in place to override the system decision if necessary. Ensure that only authorized persons are able to perform the procedure.

Supportive guidance:

Supplementary Definitions:

Override the system:
Update or modify decisions or completely stop the operation of the system.

Exemplary [parameters]:

[Procedures]: e.g.,
- Clear Escalation Paths;
- Incident Reporting Systems;
- Manual Override Buttons;
- Fail-Safe Triggers

Additional guidance regarding the requirement:

Ensuring authorized personnel can override AI systems is critical to mitigate risks and prevent harm caused by incorrect or unforeseen decisions. Overrides enable human judgment in complex or high-stakes scenarios, ensuring fairness, safety, and compliance with regulations like GDPR. They also build trust and transparency by providing a safeguard against biases, malfunctions, or failures in automated systems.

Evaluation tools:

- OWASP ZAP - can be used to test access control vulnerabilities and ensure that override mechanisms cannot be misused or bypassed by unauthorized users.
- Metasploit - can be used to perform penetration testing on manual override systems, escalation paths, and fail-safe triggers to identify security gaps.)
- Burp Suite - can be used to analyze and validate API security, ensuring that override commands function correctly and securely.
- Gremlin - chaos testing can be used to simulate system failures and validate whether override mechanisms function effectively under real-world stress conditions.

Relevance based on use case parametrization:

- The AI system operates autonomously with no human oversight.

Reference to EU AI Act:

AI Act Arts. 13, 14, 26

DORA (Level 1) References:

N/A

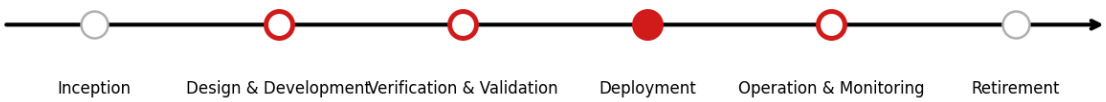
DORA (Level 2) References:

N/A

Re-evaluation triggers:

AC, AT, EM, IG, LR, TS, UG, UR

HO-03: Monitoring of oversight and overriding processes



Lifecycle overview for criteria Lifecycle overview for criteria HO-03

Evaluation requirement:

There should be clear specification and documentation of individuals that are authorized to override decisions made by the system.

Evaluation method:

Document-based: Verify that the documentation contains the specification about people who are authorized to override the decision made by the system.

Supportive guidance:

Supplementary Definitions:

Override the system:
Update or modify decisions or completely stop the operation of the system.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

Considerations should be given to:
- Potential biases that may result from this arrangement;
- Only authorized persons should have the ability to correct outputs based on their roles and responsibilities;
- Observation mechanisms to observe project team members who can oversee and override AI system decisions;
- Access removal controls should be in place to quickly revoke individual access.

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

- The AI system operates autonomously with no human oversight.

Reference to EU AI Act:

AI Act Arts. 13, 14, 26

DORA (Level 1) References:

N/A

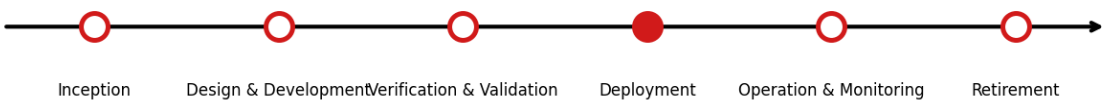
DORA (Level 2) References:

N/A

Re-evaluation triggers:

IG, LR, UG

HO-04: Attribution of ethical and legal responsibility in AI system lifecycle



Lifecycle overview for criteria Lifecycle overview for criteria HO-04

Evaluation requirement:

At all stages of the AI system's lifecycle, the physical persons or existing legal entities holding responsibility shall be clearly identified and documented.

Evaluation method:

Document-based: Verify that the physical persons or legal entities are determined and documented throughout all stages of the AI cycle.

Supportive guidance:

Supplementary Definitions:

Life cycle:
The process from design and development through deployment, operation, and eventual decommissioning.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

Specific individuals or entities bearing such responsibility within the project team shall be clearly identified.
A suitable way to document could be a RACI chart (who is responsible, who is accountable, who should be consulted, and who should be informed)

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Art. 17

DORA (Level 1) References:

N/A

DORA (Level 2) References:

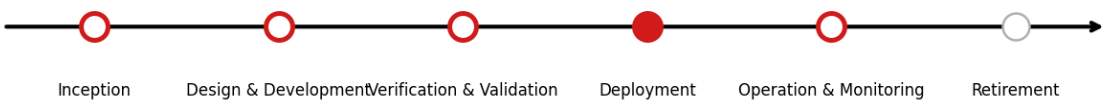
N/A

Re-evaluation triggers:

IG, LR, MA, MP, SE, SR, TD

13 IT-Security

ITSEC-01: Common standards and frameworks for privacy impact



Lifecycle overview for criteria Lifecycle overview for criteria ITSEC-01

Evaluation requirement:

Common standards, frameworks and best practices for PIA (privacy impact analysis) should be considered.

Evaluation method:

Document-based: Verify that a PIA is carried out taking into account common standards, frameworks and best practices, see supportive guidance for relevant standards.

Supportive guidance:

Supplementary Definitions:

Privacy Impact Analysis (PIA):
A systematic process to identify, assess, and mitigate risks to individuals' privacy posed by a system. It evaluates how personal data is collected, stored, processed, and shared, ensuring compliance with data protection laws and safeguarding individuals' rights

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

Relevant standards (among others) are:
- ISO/IEC 29134
- ISO/IEC 27001
- NIST Privacy Framework
- General Data Protection Regulation (GDPR)

Relevant frameworks and practices (among others) could be:
- GDPR DPIA Guidance
- OECD Guidelines on Privacy

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 10, 17, 26

DORA (Level 1) References:

N/A

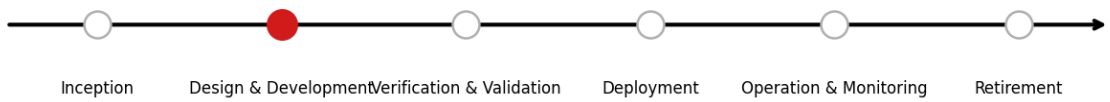
DORA (Level 2) References:

N/A

Re-evaluation triggers:

EG, NV, SN, TS

ITSEC-02: Supply chain security



Lifecycle overview for criteria Lifecycle overview for criteria ITSEC-02

Evaluation requirement:

The supply chain involved in the development shall be analyzed for potential security weaknesses, ensuring all components and dependencies are secure.

Evaluation method:

Document-based: Verify that the supply chain is analyzed for potential security weaknesses. Ensure that the analysis includes all SOUP or OTS components.

Supportive guidance:

Supplementary Definitions:

Supply chain:
The network of entities and processes involved in creating, delivering, and maintaining a system.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

This should also include a verification process for all SOUP (Software of Unknown Provenance) or OTS (Off-The-Shelf) components.

Tools which could be used during implementation:
- Snyk
- OWASP Dependency-Check
- Black Duck
- WhiteSource

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 9, 15, 17

DORA (Level 1) References:

N/A

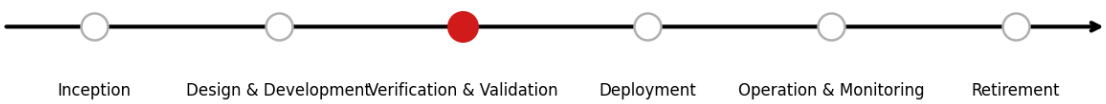
DORA (Level 2) References:

N/A

Re-evaluation triggers:

SE, SR

ITSEC-03: Risks from unspecified usage environments and users



Lifecycle overview for criteria Lifecycle overview for criteria ITSEC-03

Evaluation requirement:

The risks arising from unspecified usage environments and non-specified users shall be analyzed.

Evaluation method:

Document-based: Verify that the different types of risk that could arise from unspecified usage environments are analysed.

Supportive guidance:

Supplementary Definitions:

Unspecified Usage Environments:
Physical or operational settings that are not explicitly defined.

Non-Specified Users:
Individuals who are either untrained or lack proper authorization to access the system.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

Potential risks are:
- Misuse of the system;
- Security risks;
- Safety concerns;
- Regulatory non-compliance;
- Bias or discrimination

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 9, 15, 17

DORA (Level 1) References:

DORA Art. 9, 10

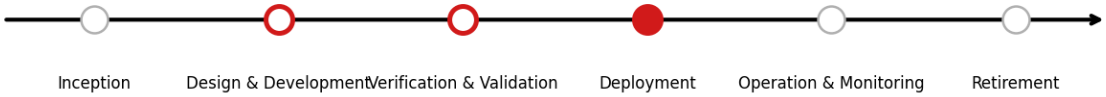
DORA (Level 2) References:

RMF-RTS Art. 21

Re-evaluation triggers:

AC, UG

ITSEC-04: Secure GenAI-App integration



Lifecycle overview for criteria Lifecycle overview for criteria ITSEC-04

Evaluation requirement:

Connections with existing systems and databases shall be safeguarded with secure integrations at all AI system interfaces.

Evaluation method:

Document-based: Verify that secure integrations are in place to protect connections between the AI system and other existing systems or databases.

Supportive guidance:

Supplementary Definitions:

System interfaces:
Points of interaction where a system communicates with other systems, applications, or users.

Secure Integrations:
The process of safely connecting systems or components to ensure data integrity, confidentiality, and functionality.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

System interfaces can be:
- LLM trust boundaries
- User Interface
- Hardware Interfaces
- Webhooks
- Input/Output Interfaces

Secure integration could be:
- Encryptied data transfer
- API security
- Firewalls and network segmentation
- Endpoint security

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 9, 15, 17

DORA (Level 1) References:

N/A

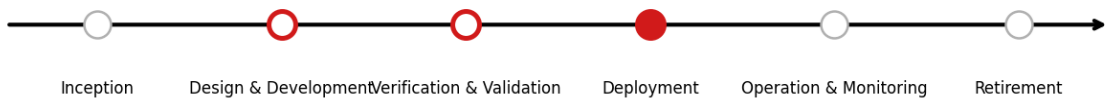
DORA (Level 2) References:

RMF-RTS Art. 13, 14

Re-evaluation triggers:

LR, SE, SR

ITSEC-05: Conventional network security protection



Lifecycle overview for criteria Lifecycle overview for criteria ITSEC-05

Evaluation requirement:

The network shall be protected by conventional IT security [measures].

Evaluation method:

Test-based: Verify that the network is protected by classical IT security measures, e.g., via access control, maleware detection mechanisms, security incident management, etc.

Supportive guidance:

Supplementary Definitions:

Conventional IT security:
Traditional methods and practices for protecting systems, data, and infrastructure from unauthorized access, misuse, or harm.

Exemplary [parameters]:

[measures]: e.g., Firewalls, virtual private networks, network segmentation, access control.

Additional guidance regarding the requirement:

A protected network is important to maintain data confidentioality and garnatee operational continuity along with cost reduction in the sense of mitigation financial loss due to e.g., data breachers, ransomware or downtime.
Possible attacks and threats include e.g. Malware, Phishing attacks, ransomware, (distributed) denial of service, unauthorized acess.

Tools that can be used for the implementation:
- Intrusion Detection System (IDS)
- Firewall
- Network Access Control (NAC)
- Security Information and Event Management (SIEM)

Evaluation tools:

- nmap - can be used to scan for open ports, detect misconfigurations, and identify potential attack vectors in the network.
- Nessus - can be used for vulnerability scanning to detect security gaps, outdated software, and misconfigured security controls.
- Wireshark - can be used to analyze network traffic and detect anomalies that could indicate security threats such as unauthorized access or malware communication.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 9, 15, 17

DORA (Level 1) References:

N/A

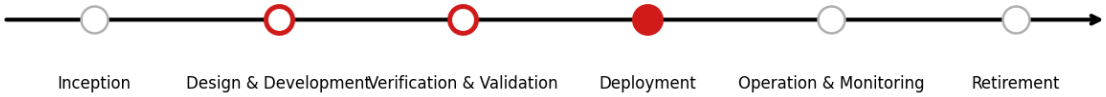
DORA (Level 2) References:

RMF-RTS Art. 13, 14

Re-evaluation triggers:

EM, LR

ITSEC-06: Gateway controls



Lifecycle overview for criteria Lifecycle overview for criteria ITSEC-06

Evaluation requirement:

[Mechanisms] shall be sufficiently implemented and documented to enable a graduated set of controls.

Evaluation method:

Document-based: Verify that the security mechanisms specified in the supportive guidance are adequately implemented and documented.

Supportive guidance:

Supplementary Definitions:

Graduated set of controls:
Security measures that were implemented in layers or tiers.

Exemplary [parameters]:

[mechanisms]: e.g. Secure gateway, VPN, and routing for machine learning systems.

Additional guidance regarding the requirement:

Possible levels are:
- risk levels
- sensitivity of the system or data
- criticality of the system or data

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 11, 13, 15, 17, 18, 26

DORA (Level 1) References:

N/A

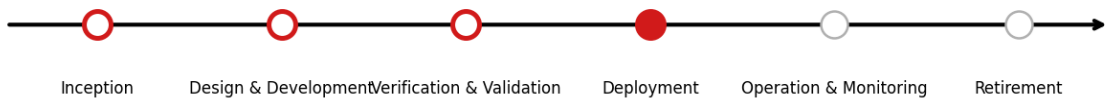
DORA (Level 2) References:

RMF-RTS Art. 13, 14

Re-evaluation triggers:

LR, SE, SR

ITSEC-07: Availability and disaster recovery



Lifecycle overview for criteria Lifecycle overview for criteria ITSEC-07

Evaluation requirement:

The AI system shall comply with the defined availability standards for its IT infrastructure, applications and data. It should also have suitable [emergency plans] in place to restore the AI system and the model in the event of a failure.

Evaluation method:

Document-based: Verify that the AI system meets the defined availability standards for its IT infrastructure. Ensure that a process is in place to restore the AI system and model in the event of a failure, such as a backup plan.

Supportive guidance:

Supplementary Definitions:

Event of a failure:
An incident where a system, component, or process does not perform as intended, leading to disrupted functionality, reduced performance, or complete inoperability. Common types of failures include hardware failures, software failures, security failures, network failures or data failures. Emergency plans shall explicitly cover scenarios in which externally hosted or third-party AI model providers become unavailable or terminate their service. Contingency measures may include hot-swap to an alternative provider, activation of a locally hosted fallback model.

Exemplary [parameters]:

[emergency plans]: e.g. Backup plans.

Additional guidance regarding the requirement:

Relevant standards (among others) could be:
- ISO/IEC 27001
- NIST Cybersecurity Framework
- CIS Controls
- Risk Management Framework

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 9, 15, 17

DORA (Level 1) References:

DORA Art. 9, 12

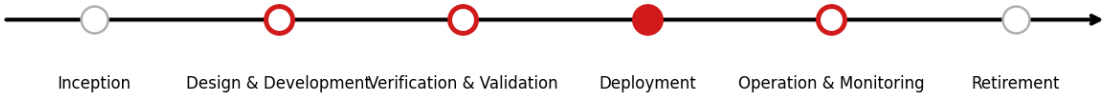
DORA (Level 2) References:

RMF-RTS Art. 9, 24, 26

Re-evaluation triggers:

AT, EM, IG, LR, SN, TS

ITSEC-08: System availability assessment



Lifecycle overview for criteria Lifecycle overview for criteria ITSEC-08

Evaluation requirement:

The availability of the AI system shall be constantly evaluated with [suitable methods] and documented.

Evaluation method:

Document-based: Verify that a procedure is in place to evaluate the availability of the system. This can include monitoring tools that monitor system properties related to the relevant business assets related to the system. In order to test the procedure, a stress test, causing the system to be unavailable shall be conducted and the reaction shall be evaluated.

Supportive guidance:

Supplementary Definitions:

Availability:
Ability of a system, to remain accessible and operational for authorized users whenever needed.

Exemplary [parameters]:

[suitable methods]: Uptime monitoring, latency monitoring, throughput monitoring, error rate monitoring, stress testing, scalability checks, maintenance scheduling, system health checks.

Additional guidance regarding the requirement:

Applicability depends on availability requirement of use case or system.
A high availability rate will improve business reputation, increase user satisfaction and build competitive advantage. Downtime of the system can lead to disrupt workflows, delay processes, and halt critical services, which can directly or indirectly lead to financial losses. A high availability rate is also required to be compliant with many legal and regulatory standards.

Tools which could be used for the implementation:
Uptime Monitoring Tools to check logs, historical uptime data, and alerts to ensure that monitoring is in place and functioning correctly: Nagios
Chaos Engineering Tools to inject faults into the system to test the suitable methods: Gremlin, Metasploit

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 9, 11, 15, 17, 18

DORA (Level 1) References:

N/A

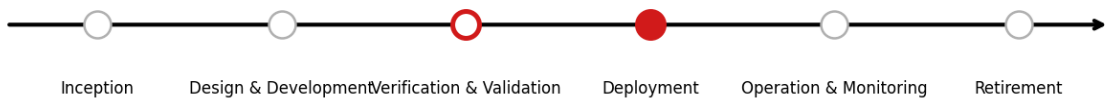
DORA (Level 2) References:

RMF-RTS Art. 9, 25

Re-evaluation triggers:

EM, SN, TS

ITSEC-09: Infrastructure security evaluation



Lifecycle overview for criteria Lifecycle overview for criteria ITSEC-09

Evaluation requirement:

Resilience testing shall be performed with [suitable techniques] to ensure that systems can withstand and recover from disruptions.

Evaluation method:

Test-based: Verify that resilience testing is carried out. Check that appropriate techniques, see supportive guidance, are implemented correctly by ensuring if the system can withstand and recover from disruptions.

Supportive guidance:

Supplementary Definitions:

Disruptions:
For example server crashes, network outages, high traffic or load, hardware failures, power failures etc.

Exemplary [parameters]:

[suitable techniques]: e.g., Fault injection, failover testing, load and stress testing.

Additional guidance regarding the requirement:

A measurement for resilience can be Performance. Applicability depends on availability requirement of use case or system.

Evaluation tools:

- Chaos Monkey - can be used to randomly terminate system instances and evaluate the system's ability to handle unexpected failures.
- Gremlin - can be used to conduct controlled fault injection and assess system resilience under various failure scenarios.
- Failure Injection Testing - can be used to simulate infrastructure-level failures and analyze system response and recovery strategies.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 9, 10, 15, 17

DORA (Level 1) References:

DORA Art. 24, 25, 26

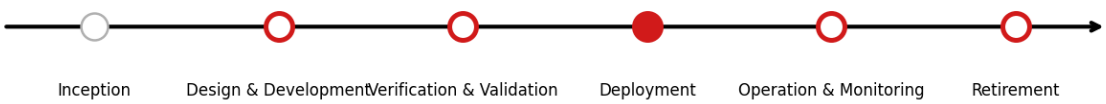
DORA (Level 2) References:

N/A

Re-evaluation triggers:

AC, EM, SN, TS, UG, UP, UR

ITSEC-10: Data classification and protection based on sensitivity



Lifecycle overview for criteria Lifecycle overview for criteria ITSEC-10

Evaluation requirement:

Data must be classified and protected based on its sensitivity. User permissions must be managed effectively, and appropriate safeguards must be in place to ensure that access controls and defense-in-depth measures are robust.

Evaluation method:

Document-based: Verify if appropriate measures for the classification and protection of data based on its sensitivity are in place. Furthermore, verify that protective measure to safeguard the sensitive data are in place and test their functionality. And User roles ensure only appropriate data access is possible.

Supportive guidance:

Supplementary Definitions:

Sensitive data:
Information that requires extra protection due to its potential to cause harm if accessed or disclosed without authorization, e.g. presonal or biometric data.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

Appropriate safeguards can be:
- Role-based access crontrol (RBAC)
- Session management
- Encryption
- Physical Security

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 10, 15

DORA (Level 1) References:

N/A

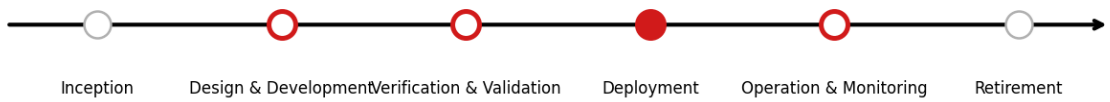
DORA (Level 2) References:

RMF-RTS Art. 4, 11

Re-evaluation triggers:

EG, IG, LR, SN, TD, UG

ITSEC-11: Ensuring data authenticity



Lifecycle overview for criteria Lifecycle overview for criteria ITSEC-11

Evaluation requirement:

The system shall ensure the authenticity of data.

Evaluation method:

Test-based: Verify that every method to add or manipulate data can only be used with sufficient authentication of the user.
Also ensure that data is genuine, unaltered or traceably altered, and originates from a trusted source

Supportive guidance:

Supplementary Definitions:

Authenticity of data:
The quality of being genuine, unaltered/altered with traces, and verifiably from a trusted source.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

Data authenticity can be secured through the use of public key infrastructure (PKI) and digital certificates.

Evaluation tools:

- Wireshark - can be used to analyze network traffic for verifying data integrity, checking if data exchanges include valid digital signatures and certificates.
- PKI Compliance Audit Tools - compliance tools like X.509 certificate validators can be used to check if the system correctly implements PKI for authenticity verification.
- Scapy - can be used to craft and send inauthentic or tampered data packets, testing whether the system correctly detects and rejects altered data.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 10, 15, 17

DORA (Level 1) References:

DORA Art. 9

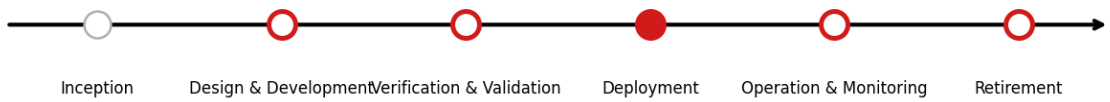
DORA (Level 2) References:

RMF-RTS Art. 20, 21

Re-evaluation triggers:

EG, LR, TD

ITSEC-12: Authorization and access control



Lifecycle overview for criteria Lifecycle overview for criteria ITSEC-12

Evaluation requirement:

An access control concept for the AI System should be created, defining roles, authorizations and responsibilities along the MLOps process and its feedback loops by logging and auditing all activities.

Evaluation method:

Test-based: Verify that an access control concept is in place. The AI system must withstand unauthorised access attempts using simulated attack scenarios. Ensure that least-privilege access controls and defence-in-depth measures effectively restrict unauthorised access by attempting to access resources and data with different privilege levels.

Supportive guidance:

Supplementary Definitions:

Access control concept:
A framework that defines how access to resources, data, or systems is managed and restricted based on user roles, permissions, and security policies.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

Access control mechanisms can include least privilege access controls or defence-in-depth measures.

Evaluation tools:

- OpenSCAP - can be used to audit and assess compliance with access control policies, ensuring that security configurations follow defined standards.
- Metasploit Framework - can be used to simulate unauthorized access attempts and assess how well the system enforces access restrictions.
- BloodHound - can be used to analyze permission structures and identify privilege escalation paths in the AI system's access control.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 12, 15, 17, 19

DORA (Level 1) References:

DORA Art. 9

DORA (Level 2) References:

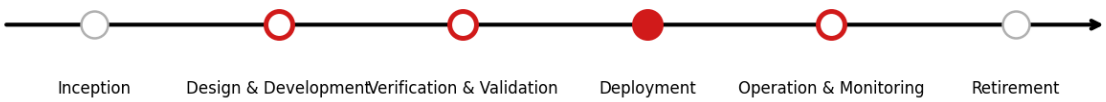
RMF-RTS Art. 20, 21, 25

Re-evaluation triggers:

AC, IG, LR, UG

14 Monitoring

MON-01: Performance development & operation



Lifecycle overview for criteria Lifecycle overview for criteria MON-01

Evaluation requirement:

The performance of the system shall be continuously monitored and documented with [Key Performance Indicators]. Deviations from defined performance requirements shall be made transparent.

Evaluation method:

Test-based: Verify that appropriate measure to continuously monitor and document the system performance are in place (exeplary tools are given in the supportive guidance). Verify the monitored variables and the defined KPIs and test the mechanims to detect (and document) performance deviations from the defined performance requirements.

Supportive guidance:

Supplementary Definitions:

Performance:
Indicator that shows how effectively and efficiently an AI system accomplishes its tasks.

KPIs (Key Performance Indicators):
Metrics to evaluate the performance.

Exemplary [parameters]:

[key performance indicators]:, e.g. F1 Score, Mean Squared Error (MSE), response times, resource consumption.

Additional guidance regarding the requirement:

The system shall demonstrate the same KPIs in development respectively testing and validation and after the deployment in its operating state.

Evaluation tools:

- Google Lighthouse Performance Auditor - can be used to verify if system response times meet defined KPI thresholds and whether Performance degrades under load.
- Apache JMeter - can be used to validate if system Performance remains consistent across different loads and if deviations from expected KPIs are detected.
- ML Benchmark Tools - benchmarking Tools can be used to compare AI model Performance against predefined KPIs and industry standards.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 11, 15, 17, 18, 72

DORA (Level 1) References:

N/A

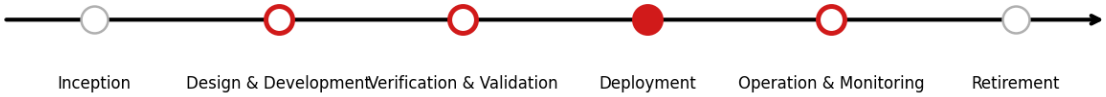
DORA (Level 2) References:

RMF-RTS Art. 9, 34

Re-evaluation triggers:

LR, SE, SN, SR, UP, UR

MON-02: Error detection



Lifecycle overview for criteria Lifecycle overview for criteria MON-02

Evaluation requirement:

The system should have [mechanisms] to allow detection of errors and fail-safe methods to mitigate entire system failures.

Evaluation method:

Test-based: Verify if the AI system has mechanisms for detection of errors and fail-safe methods, e.g. by actively triggering fault conditions, analyzing the system's response, testing for internal consistency or anomalies.

Supportive guidance:

Supplementary Definitions:

Error detection:
Criteria and mechanisms for (automatic) error detection.

Exemplary [parameters]:

[mechanisms]: e.g., Internal consistency checks, anomaly detection algorithms.

Additional guidance regarding the requirement:

Entire system failures are where the AI system becomes completely non-functional or unable to perform its intended tasks due to Hardware Failures, Software Failures, Security Failures, Network Failures or Data Failures.
A fully automated system will need to self-detect errors while a system that is actively used by a human can flag the errors to its user.

Evaluation tools:

- Chaos Monkey - controlled failures can be injected into the system, such as network disruptions or service crashes, to verify whether self-error detection mechanisms identify and mitigate the issue.
- Apache JMeter - can be used to validate if system performance remains consistent across different loads.
- Failure Injection Testing - can be used to simulate infrastructure-level failures and analyze system response and recovery strategies.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 15, 17

DORA (Level 1) References:

N/A

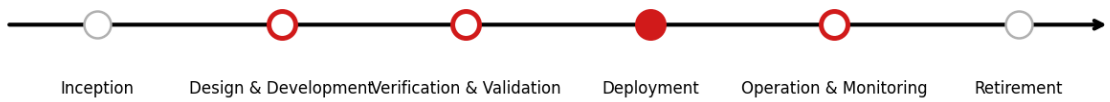
DORA (Level 2) References:

N/A

Re-evaluation triggers:

EM, LR, SE, SN, SR

MON-03: Logging



Lifecycle overview for criteria Lifecycle overview for criteria MON-03

Evaluation requirement:

The AI system shall have consistent logging across all components. Policies and instructions with technical and organizational safeguards for the logging process shall be documented.

Evaluation method:

Document-based: Verify that appropriate monitoring and logging capabilities (tools, storage, etc.) are implemented on a technical and organizational level. Evaluate their effectiveness by targeted test, e.g., by inducing certain events that shall then be verified via the resulting system logs.

Supportive guidance:

Supplementary Definitions:

Consistent logging:
Systematic recording of events, actions, and decision made by the system.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

The log files should contain at least:
- Type of request ;
- Processing times including time stamps and metadata on the user requesting the AI service.
The log files should be kept for an appropriate amount of time (suggestion: at least 3 months) (depending on the system and the amount of data to be logged).
Contents to log shall be specified: e.g., Input, model (state), output, timestamps, system access.

Tools which could be used for implementation:
- Log Analysis and Monitoring Tools to check if the logs include required information and verify that logging is consistent: Loggly
- Log File Integrity Tools to verify that logs are secure, consistent, and meet the technical and organizational safeguard: Tripwire

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 11, 12, 17, 18, 19

DORA (Level 1) References:

N/A

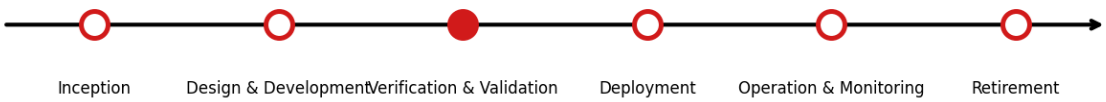
DORA (Level 2) References:

RMF-RTS Art. 12

Re-evaluation triggers:

IG, LR, SE, SN, SR

MON-04: Incident response procedures



Lifecycle overview for criteria Lifecycle overview for criteria MON-04

Evaluation requirement:

Formal incident response procedures shall be implemented to report AI systems incidents throughout the whole lifecycle. The procedures should be tested on a periodic basis, tracking with [Key Performance Indicators].

Evaluation method:

Document-based: Verify that system states that qualify as incidents are defined. Verify that appropriate formal incident response procedures are implemented and documented.
Assess whether the response procedures are tested before deployment and an appropriate periodic plan of repeated testing is established.Verify the incident response procedures by simulating incidents, analyzing response effectiveness, and measuring key performance indicators.

Supportive guidance:

Supplementary Definitions:

Formal incident response procedures:
Structured guidelines and processes for identifying, managing, and mitigating security incidents or breaches.

KPIs (Key Performance Indicators):
Metrics to evaluate the performance.

Exemplary [parameters]:

[KPIs]: e.g. F1 Score, Mean Squared Error (MSE)

Additional guidance regarding the requirement:

AI system incidents could be for example loss of service, loss of equipment, loss of facilities, system malfunctions, system overloads, human errors, and non-compliances with policies or guidelines, breaches of physical security, uncontrolled system changes, software malfunctions, hardware malfunctions, and access violations.

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 9, 12, 17, 19, 26, 72, 73

DORA (Level 1) References:

DORA Art. 17

DORA (Level 2) References:

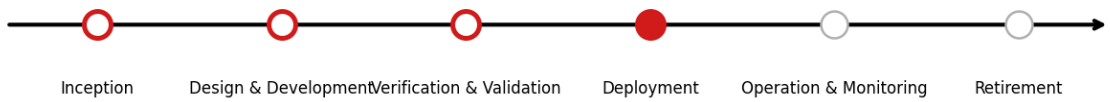
RMF-RTS Art. 22, 23

Re-evaluation triggers:

EM, LR, SE, SN, SR

15 Performance

PERF-01: Definition of performance requirements



Lifecycle overview for criteria Lifecycle overview for criteria PERF-01

Evaluation requirement:

The performance requirements shall be defined and included in the system description with at least the [Key Performance Indicators] used.

Evaluation method:

Document-based: Verify that the key performance indicators, see supportive guidance, are defined and included in the system description.

Supportive guidance:

Supplementary Definitions:

Performance Requirements:
Criteria that define the expected level of functionality and efficiency of the system, often measured through specific metrics like accuracy, speed, reliability, or resource usage, to ensure it meets its intended goals.

KPIs (Key Performance Indicators):
Metrics to evaluate the performance.

Exemplary [parameters]:

[key performance indicators] (KPIs): type of metric to evaluate the performance, e.g. F1 Score, Mean Squared Error (MSE).

Additional guidance regarding the requirement:

Not applicable.

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 11, 15, 17, 18, 72

DORA (Level 1) References:

Verify that the system can successfully detect failures. Ensure that a fallback function, see supportive guidance, is in place in the Event of a failure.

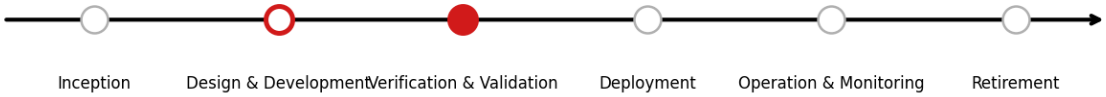
DORA (Level 2) References:

N/A

Re-evaluation triggers:

IG, LR, UP

PERF-02: Generalization performance verification via XAI



Lifecycle overview for criteria Lifecycle overview for criteria PERF-02

Evaluation requirement:

The ability of the AI system to generalize shall be verified by appropriate [XAI methods].

Evaluation method:

Test-based: Verify that an appropriate XAI method was implemented to ensure that the AI system effectively generalizes. This can be done by testing its ability to handle unseen data distributions through adversarial and explainability-driven evaluations.

Supportive guidance:

Supplementary Definitions:

XAI = Explainable AI

Generalize:
The ability of an AI system to perform well on new, unseen data. This demonstrates adaptability beyond examples used in training datasets.

Exemplary [parameters]:

[XAI methods]: e.g., Post-hoc explanations, model explanations or data explanations can help to understand the AI system and its performance regarding generalization. For bigger systems, a partitioning into subsystems that are explained individually is possible.

Additional guidance regarding the requirement:

Insufficient generalizability can, e.g., lead to model collapse in LLMs.

Evaluation tools:

- Foolbox - can be used to perform adversarial robustness testing by applying small, realistic perturbations to input data and verifying if the AI model maintains performance or collapses due to lack of generalization.
- Robustness Gym - can be used to test the AI system under different real-world data perturbations, including paraphrased text, altered image resolutions, and noise injections. Evaluate whether accuracy remains stable across these transformations.
Model Testing Frameworks with XAI Capabilities - can be used to verify that post-hoc explanations are correctly implemented and provide meaningful insights.
- Captum - can be used to perform Integrated Gradients or Layer Conductance analysis on different test cases. If feature attributions shift erratically across near-identical inputs, this suggests poor generalization capability.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 11, 13, 18

DORA (Level 1) References:

N/A

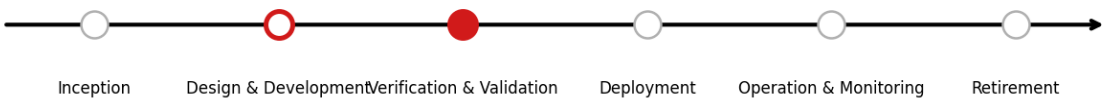
DORA (Level 2) References:

N/A

Re-evaluation triggers:

AT, MA, MP, TD, TS

PERF-03: Measures against overfitting



Lifecycle overview for criteria Lifecycle overview for criteria PERF-03

Evaluation requirement:

[Measures] must be implemented to minimize overfitting and underfitting, ensuring that the model is adequately trained and performs effectively for its intended purpose.

Evaluation method:

Test-based: Verify, through code review and/or performance testing, that measures to avoid or reduce overfitting, underfitting and to ensure performance are implemented. Also check model's generalization ability.

Supportive guidance:

Supplementary Definitions:

Overfitting:
A situation where a model learns the training data too well, including noise and specific details, resulting in poor performance on new, unseen data due to lack of generalization.

Underfitting:
A situation where a model fails to capture the underlying patterns in the training data, leading to poor performance both on the training data and unseen data.

Exemplary [parameters]:

[Measures]: e.g., Cross-validation, regularization.

Additional guidance regarding the requirement:

Not applicable.

Evaluation tools:

- SHAP (SHapley Additive exPlanations) - can be used to analyze feature importance and detect cases where the model overly depends on specific training set patterns, indicating overfitting.
- Adversarial Robustness Toolbox (ART) - can be used to introduce slight perturbations in input data and observe the model's response, checking if small changes cause drastic misclassifications, which would indicate poor generalization.
- DeepCheck - can be used to evaluate model stability by applying distribution shift tests, ensuring that it maintains expected performance on slightly altered data distributions.

Relevance based on use case parametrization:

At least partial training was performed.

Reference to EU AI Act:

AI Act Arts. 10, 15, 17

DORA (Level 1) References:

N/A

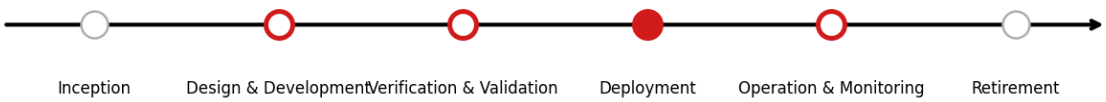
DORA (Level 2) References:

N/A

Re-evaluation triggers:

AT, MA, MP, TD, TS, UP

PERF-04: Periodic retraining and model changes



Lifecycle overview for criteria Lifecycle overview for criteria PERF-04

Evaluation requirement:

It shall be ensured that continuous improvement of the model's performance is achieved through [retraining methods].

Evaluation method:

Document-based: Verify that a retraining method, see supportive guidance, can be performed. Ensure that triggering events, such as model drifts, are detected to appropriately trigger the retraining method and logged accordingly.
Identify if the retraining method mitigated the risk respectively improved performance, e.g., by using the tools stated in the supportive guidance and comparing the retrained version with the former system. Otherwise a procedure should be in place for a domain expert to reevaluate the risk exposure of the model and the model concept. Confirm that all adjustments and model changes are documented.

Supportive guidance:

Supplementary Definitions:

Not applicable.

Exemplary [parameters]:

[retraining methods]: e.g., Periodic retraining, adjustment of the conceptual model or on-demand retraining.

Additional guidance regarding the requirement:

Retraining should in all cases incorporate new ('unseen') data. Retraining is relevant especially when there are model or concept drifts. If issues persist after retraining, experts shall reassess the model, evaluate risks, and document all changes.

Tools which could be used for implementation:
- DeepChecks to perform side-by-side comparisons of pre- and post-retraining model performance, detecting anomalies, biases, and statistical shifts in feature behavior
Data Drift and Model Drift Monitoring Tools to check if triggered when necessary:
- AI Explainability 360 (AIX360) to assess explainability differences before and after retraining, verifying that retraining improves model decisions rather than introducing hidden biases or reduced interpretability.
- Review of Azure Event Grid Triggers
- Review of CI/CD Flows

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

At least partial training was performed.

Reference to EU AI Act:

AI Act Arts. 15, 17

DORA (Level 1) References:

N/A

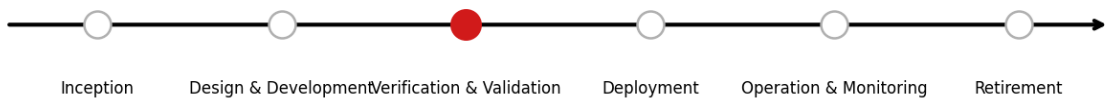
DORA (Level 2) References:

N/A

Re-evaluation triggers:

AT, MA, MP, TD, TS, UP

PERF-05: Testing



Lifecycle overview for criteria Lifecycle overview for criteria PERF-05

Evaluation requirement:

Performance testing of the AI System shall be performed and documented.

Evaluation method:

Document-based: Verify that suitable performance tests were performed and the execution and the corresponding results are documented in a dedictated documentation. The documentation shall comprehensive, so that reproducibility of the performed tests is given.

Supportive guidance:

Supplementary Definitions:

Not applicable.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

Testing should include, for example:
- Stress testing;
- Performance testing on performance KPIs with inputs from the operational design domain of the system;
- Performance testing against expected input perturbations/deviations;
- Performance testing on Out-of-Distribution data;
- Test scenarios/test cases for, e.g. interpretable model decisions.

Tools which could be used for implementation:
- tensorboard

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 9, 11, 15, 17, 18

DORA (Level 1) References:

N/A

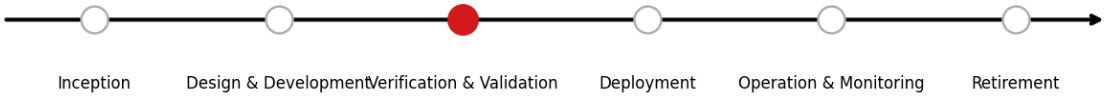
DORA (Level 2) References:

RMF-RTS Art. 15, 16

Re-evaluation triggers:

AT, TS, UP

PERF-06: Assessment of uncertainty



Lifecycle overview for criteria Lifecycle overview for criteria PERF-06

Evaluation requirement:

Uncertainty assessment of the AI system shall be performed with [measures] and documented.

Evaluation method:

Test-based: Verify that an uncertainty assessment was conducted and documented and that it was performed correctly, e.g., by testing whether variations in input features lead to unexpected changes (in the sense of not considered during uncertainty assessment) in the system's predictions.

Supportive guidance:

Supplementary Definitions:

Uncertainty Assessment:
The process of evaluating and quantifying the confidence or reliability of an AI system's predictions or outputs, often by analyzing how variations in inputs or conditions impact its performance and decision-making.

Exemplary [parameters]:

[measures]: e.g., Sensitivity analysis.

Additional guidance regarding the requirement:

E.g., how do changes in individual features affect predictions, particularly for tabular data?
What impact does user-provided data have on future changes to the AI service?
If there are contractually defined performance requirements, deviations from these shall be made transparent to the user.

Evaluation tools:

- InterpretML - can be used to conduct sensitivity analysis and test how variations in individual input features affect the model's predictions. Verify if the uncertainty quantification aligns with the documented assessment.
- Deep Ensemble Tester - can be used to verify if ensemble-based uncertainty quantification produces stable confidence intervals and meaningful variance across multiple perturbed runs.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 11, 15, 17, 18

DORA (Level 1) References:

N/A

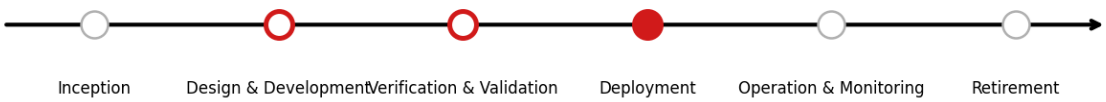
DORA (Level 2) References:

N/A

Re-evaluation triggers:

AT, MA, MP, TD, TS, UR

PERF-07: Error handling



Lifecycle overview for criteria Lifecycle overview for criteria PERF-07

Evaluation requirement:

A procedure shall be in place to detect failures of the system. In case of a detected failure, the system shall answer with [error fallback functions].

Evaluation method:

Test-based: Verify that the system can successfully detect failures. Ensure that a fallback function, see supportive guidance, is in place in the Event of a failure.

Supportive guidance:

Supplementary Definitions:

Procedures (exemplary):
Monitoring and Logging, Self-Checking Mechanisms, Error Detection Algorithms, Threshold-Based Alerts, Redundancy Checks, User Feedback Loops

Exemplary [parameters]:

[error fallback functions]: e.g., Error correction, determination of the erroneous state of the system or deactivation of the system and notification of the bodies and parties concerned.

Additional guidance regarding the requirement:

Not applicable.

Evaluation tools:

- Chaos Monkey - can be used to randomly terminate system components or induce controlled failures to evaluate whether the system detects faults and responds correctly with error fallback mechanisms.
- Metasploit - perform controlled penetration testing to determine if failures lead to unintended system behavior, ensuring fallback mechanisms handle faults securely.
- Selenium - can be used to automate test scripts that simulate erroneous inputs, UI malfunctions, or system crashes, verifying whether failures are detected and appropriate fallback actions are triggered.
- Gremlin - can be use to inject controlled faults into the system, such as CPU spikes, memory leaks, or network failures, to test whether the system identifies these disruptions and executes error fallback functions accordingly.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 9, 11, 15, 17, 18, 72

DORA (Level 1) References:

DORA Art. 10

DORA (Level 2) References:

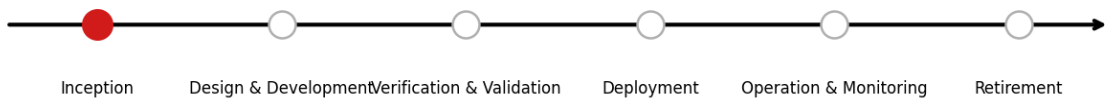
RMF-RTS Art. 12

Re-evaluation triggers:

EM, IG, LR

16 Transparency

TR-01: Assessment of the required degree of explainability



Lifecycle overview for criteria Lifecycle overview for criteria TR-01

Evaluation requirement:

The need and required degree for explanation of the system shall be assessed and documented.

Evaluation method:

Document-based: Verify that the degree of explanation for the system has been assessed. Ensure that the documentation contains the assessed degree.

Supportive guidance:

Supplementary Definitions:

Explanation of the system: An explanation of a system refers to the information provided to make the system's decisions, outputs, or behavior understandable to different stakeholders. This can include describing how, why, and under what conditions a decision or outcome was made.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

During the assessment, considerations should be given to:
- Purpose of the system
- Affected stakeholders
- Potential damages resulting from the system
- Needs and prerequisites for human decision making
- Adequate handling of outliers
- Organization's obligations to reporting information to interested parties and regulators
- The necessary information to users
- Trade-off between model interpretability and performance
- Applicable legal and regulatory requirements and international standards that require the explainability of actions of the AI service
- Justified interest by users, which requires the implementation of methods to improve explainability

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 11, 13, 15, 17, 18, 86

DORA (Level 1) References:

N/A

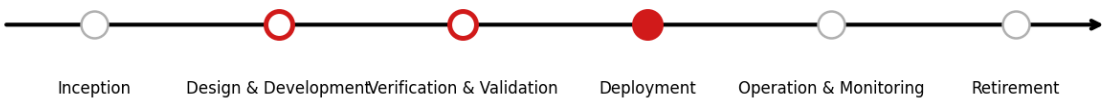
DORA (Level 2) References:

N/A

Re-evaluation triggers:

AC, LR, MA, MP, TD

TR-02: Unique identification of AI actions



Lifecycle overview for criteria Lifecycle overview for criteria TR-02

Evaluation requirement:

Actions carried out by the AI system shall be uniquely identifiable to ensure authenticity.

Evaluation method:

Document-based: Verify that each AI Action is assigned a unique identifier.

Supportive guidance:

Supplementary Definitions:

Action:
Any action or event initiated or completed by the AI system that results in a change of state, data exchange, or processing outcome.

Uniquely Identifiable:
Each transaction has a distinct identifier (e.g., a transaction ID or hash) that makes it distinguishable from all other transactions.

Authenticity:
The assurance that the transaction is genuine, unaltered, and can be verified as having been carried out by the AI system as intended, without unauthorized modifications or tampering.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

Unique identifiers can for example be established with transactions ids, timestamps, digital signatures, user or system metadata involved in the transaction.

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 12, 17, 19

DORA (Level 1) References:

DORA Art. 17

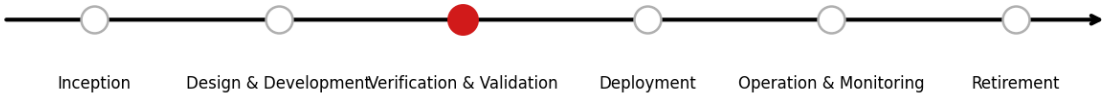
DORA (Level 2) References:

RMF-RTS Art. 23, 25

Re-evaluation triggers:

IG, LR, MA, MP, TD

TR-03: Plausibility check



Lifecycle overview for criteria Lifecycle overview for criteria TR-03

Evaluation requirement:

The output of the AI system shall be plausible for [inputs].

Evaluation method:

Document-based: Verify, using appropriate explanation methods, that the output is reasonable across a variety of input cases. Refer to the provided supportive guidance for further clarification.

Supportive guidance:

Supplementary Definitions:

Plausible: The AI system's output should be credible, reasonable, and appropriate given the context of the inputs.

Exemplary [parameters]:

[inputs]: e.g., corner cases, boundary values, OOD input, expected Inputs, regular input.

Additional guidance regarding the requirement:

The model's decision on failed tests should be explained.

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 13, 15, 17

DORA (Level 1) References:

N/A

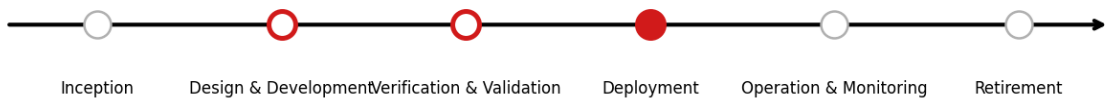
DORA (Level 2) References:

N/A

Re-evaluation triggers:

AT, MA, MP, TD, TS

TR-04: Protect user from overreliance



Lifecycle overview for criteria Lifecycle overview for criteria TR-04

Evaluation requirement:

Measures shall be in place to protect the users from overreliance of the AI system.

Evaluation method:

Document-based: Verify that measures are in place to protect the users from overly reliying on the AI models output. The needed measures depend on the degree of autononmy of the system, human oversight measure put in place and the overall risk of the system.

Supportive guidance:

Supplementary Definitions:

Overreliance:
The tendency of users to depend excessively on the AI system's outputs or recommendations, potentially leading to reduced human judgment, critical thinking, or oversight, even in cases where the AI might produce incorrect or incomplete results.

Exemplary [parameters]:

[measures]: e.g., informing the user about the performance and robustness of the system, integration of checks and questions to ensure that the user questions the AI systems output before proceeding.

Additional guidance regarding the requirement:

Not applicable.

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

- The system interacts directly with external consumers OR
- The system operates autonomously with no human oversight.

Reference to EU AI Act:

AI Act Arts. 14, 15, 17

DORA (Level 1) References:

N/A

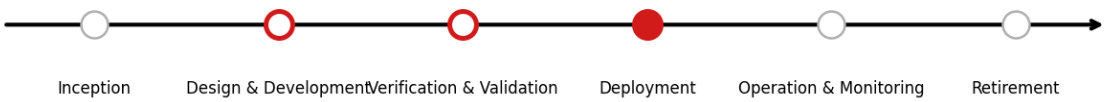
DORA (Level 2) References:

N/A

Re-evaluation triggers:

AC, EG, LR, SN, UG

TR-05: User awareness of AI interaction



Lifecycle overview for criteria Lifecycle overview for criteria TR-05

Evaluation requirement:

Users shall be made fully aware when they are interacting with the AI system and not a human.

Evaluation method:

Document-based: Confirm that users are adequately informed that they are interacting with an AI system.

Supportive guidance:

Supplementary Definitions:

User Awareness:
Ensuring users clearly understand that they are interacting with an AI system rather than a human.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

To ensure user awareness, it would be suitable:
- to explicitly inform the user at the beginning of an interaction,
- use explicit labels for the system,
- maintain visible indicators throughout the interaction.

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

- The system interacts directly with external consumers OR
- The system operates autonomously with no human oversight.

Reference to EU AI Act:

AI Act Art. 50

DORA (Level 1) References:

N/A

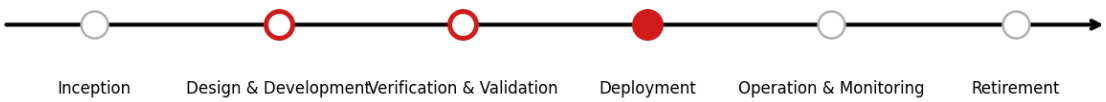
DORA (Level 2) References:

N/A

Re-evaluation triggers:

AC, EG, LR, SN, UG

TR-06: Marking of AI content



Lifecycle overview for criteria Lifecycle overview for criteria TR-06

Evaluation requirement:

AI-generated or AI-modified content shall be clearly identifiable as such. Where applicable it needs to be appropriately marked as AI content using [techniques].

Evaluation method:

Test-based: Review the implementation of techniques as referenced in the supportive guidance, e.g. using huggingface, and assess whether they are appropriate. For example, verify if AI-generated or AI-modified content is appropriately marked as AI content by attempting to remove, alter, or evade to test the robustness of the watermarking method with the proposed evaluation tool.

Supportive guidance:

Supplementary Definitions:

AI-Generated Content:
Content entirely created by an AI system without direct human creation or authorship.

AI-Modified Content:
Content that originates from human input but has been altered, enhanced, or adapted by an AI system.

Exemplary [parameters]:

[techniques]: e.g., DNN (Deep Neural Network) Watermarking, Spread-Spectrum Watermarking, Perturbation-Based Watermarking, Adversarial Watermarking.

Additional guidance regarding the requirement:

Have a look at https://huggingface.co/blog/watermarking

Evaluation tools:

- Watermark-Robustness-Toolbox - can be used to evaluate the resilience of watermarking techniques by simulating various attacks to determine if AI-generated labels remain intact.

Relevance based on use case parametrization:

- The system integrates at least one Generative AI Foundation Model OR
- The system interacts directly with external consumers OR
- The AI system operates autonomously with no human oversight.

Reference to EU AI Act:

AI Act Art. 50

DORA (Level 1) References:

N/A

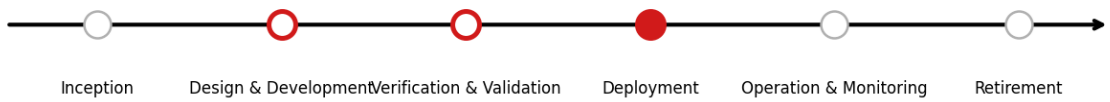
DORA (Level 2) References:

N/A

Re-evaluation triggers:

AC, AT, EG, LR, SN, TS

TR-07: Provision of tailored explanations and transparency



Lifecycle overview for criteria Lifecycle overview for criteria TR-07

Evaluation requirement:

A tailored explanation of why a particular output has been produced by the AI system shall be provided to the users with [explanation method].

Evaluation method:

Test-based: Verify that the explanation methods, see supportive guidance, have been properly implemented, e.g. using Shap, and that these explanations are readily accessible to users.
This can be done by systematically stress-testing the systmens response consistency while checking if explanations remain logically consistent when inputs are slightly altered

Supportive guidance:

Supplementary Definitions:

Tailored Explanation:
A specific, context-aware description provided to users, detailing the reasoning, factors, or data that led the AI system to produce a particular output or decision. The explanation is adapted to the user's needs, expertise, and the complexity of the output

Exemplary [parameters]:

[explanation method]: e.g., Transparency of the data, the AI algorithm, the inference (if applicable); explainability of the AI system; the model's decision at the boundary values; the model's decision on corner cases; feature importance; consider local and global model behaviour if necessary.

Additional guidance regarding the requirement:

Not applicable.

Evaluation tools:

- Counterfactual Testing - can be used to assess if explanations logically shift when similar but distinct cases are presented, ensuring the system maintains a valid rationale for different scenarios.

Relevance based on use case parametrization:

- The model must explain reasoning for specific inference results OR
- The system interacts directly with external consumers OR
- The AI system operates autonomously with no human oversight.

Reference to EU AI Act:

AI Act Arts. 13, 86

DORA (Level 1) References:

N/A

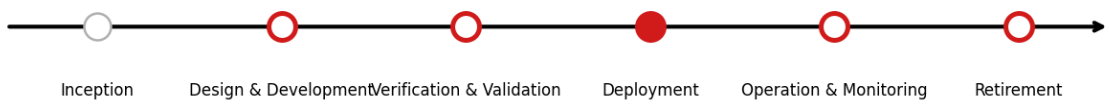
DORA (Level 2) References:

N/A

Re-evaluation triggers:

AC, AT, EG, LR, SN, TS, UG

TR-08: Provision of suitable AI system documentation for relevant parties



Lifecycle overview for criteria Lifecycle overview for criteria TR-08

Evaluation requirement:

The technical documentation of the AI sytsem shall contain all relevant information for each interested party and stakeholder.

Evaluation method:

Document-based: Verify that the documentation includes relevant and necessary information in accordance with the supportive guidance, and ensure that it is accessible for the user.

Supportive guidance:

Supplementary Definitions:

Interested Parties:
Individuals or entities who have a vested interest in the AI system and require documentation to fulfill their roles or responsibilities. Examples include:

- Developers and engineers
- End users
- Compliance officers and regulators
- Business stakeholders
- Supervisory authorities
- Partners

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

Aspects to consider in the documentation/manual:
- Product version = Specific version number or identifier
- Quality measures = Metrics such as accuracy, precision, recall
- User understanding = Comprehension level, ease of use

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 11, 13, 18

DORA (Level 1) References:

N/A

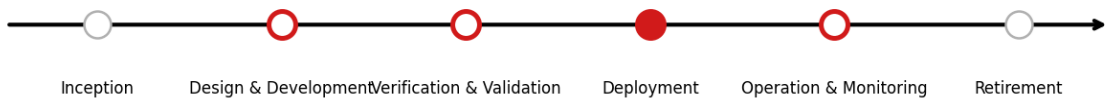
DORA (Level 2) References:

RMF-RTS Art. 16

Re-evaluation triggers:

AC, MA, SE, SR, UG

TR-09: Claims submission and appeal process



Lifecycle overview for criteria Lifecycle overview for criteria TR-09

Evaluation requirement:

The system shall allow individuals impacted by it to submit claims and complaints. The user feedback should be regularly reviewed.

Evaluation method:

Document-based: Verify that there are instructions on how to submit claims and complaints. Verify that claims and complaints can be submitted, e.g. check corresponding tools (see supportive guidance) are implemented, e.g., to evaluate user experience.

Supportive guidance:

Supplementary Definitions:

User Feedback:
Any input, complaint, suggestion, or observation provided by users or stakeholders regarding their experience with the system.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

Details on how to appeal should also be provided to the users.

Tools which could be used during implementation:
- Form and Workflow Testing: Selenium
- Usability Testing Tools: Hotjar
- Accessibility Testing: Wave

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

- The system interacts directly with external consumers OR
- The system operates autonomously with no human oversight OR
- The model has a societal impact.

Reference to EU AI Act:

AI Act Arts. 17, 26

DORA (Level 1) References:

N/A

DORA (Level 2) References:

N/A

Re-evaluation triggers:

AC, EG, LR, UG

TR-10: Notification of system use in terms and conditions and EULA

Lifecycle overview for criteria Lifecycle overview for criteria TR-10

Evaluation requirement:

Existing terms and conditions and EULAs shall be reviewed and updated for any AI considerations for relevant AI systems.

Evaluation method:

Document-based: Verify that AI considerations are addressed and represented in terms, conditions and EULA.

Supportive guidance:

Supplementary Definitions:

Terms and Conditions (T&Cs):
The legal agreement that defines the rules, guidelines, and expectations for users interacting with a service or product. It specifies user rights, responsibilities, and liabilities.

End-User License Agreement (EULA):
A legal contract between the provider of software or services and the user, granting the user rights to use the software while outlining restrictions and obligations.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

Considerations should be given to how the AI handles
- User prompts;
- Output rights and ownership;
- Data privacy;
- Compliance;
- Liability;
- Privacy;
- Limits on how output can be used;
- Privacy and protection of sensitive data.

Tools which could be used during implementation:
- Appropriateness of Language: Grammarly
- Legal and Compliance Check: Compliance.ai

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

N/A

DORA (Level 1) References:

N/A

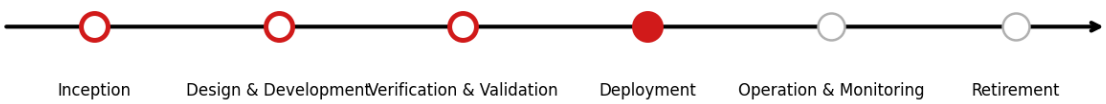
DORA (Level 2) References:

N/A

Re-evaluation triggers:

IG, LR

TR-11: Establishing regular error review and incident reporting



Lifecycle overview for criteria Lifecycle overview for criteria TR-11

Evaluation requirement:

An error report evaluation shall be established and executed.
Discovered incidents should be communicated to users of the AI system.

Evaluation method:

Document-based: Verify that a process is established for receiving and reviewing claims and complaints reports, and verify that a mechanism for reporting incidents back to users is effectively implemented.

Supportive guidance:

Supplementary Definitions:

Error Report Evaluation:
A systematic process for collecting, analyzing, categorizing, and addressing errors or malfunctions within the AI system to ensure continuous improvement, reliability, and safety.

Incident:
Any unexpected or incorrect behavior, malfunction, bias, or failure in the AI system that impacts performance, accuracy, safety, or user experience.

Exemplary [parameters]:

Not applicable.

Additional guidance regarding the requirement:

Errors should be categorized according to their severity, type and impact.

Evaluation tools:

Not applicable.

Relevance based on use case parametrization:

Relevant for all use cases.

Reference to EU AI Act:

AI Act Arts. 9, 12, 17, 19, 26, 72, 73

DORA (Level 1) References:

DORA Art. 17

DORA (Level 2) References:

RMF-RTS Art. 22

Re-evaluation triggers:

EM, LR, SN, TS

17 Glossary

Term	Definition
Access control concept	A framework that defines how access to resources, data, or systems is managed and restricted based on user roles, permissions, and security policies.
Accessibility	The system should be operable by people with a wide range of impairments, including visual impairments, hearing impairments and physical impairments; it should support multiple languages.
Active informed consent	A type of consent where individuals actively and explicitly agree to the processing and storage of their data after being fully informed about the nature, purpose, and implications of that processing.
AI policy	A set of guidelines and principles designed to govern the development, deployment, and use of artificial intelligence systems.
AI-Generated Content	Content entirely created by an AI system without direct human creation or authorship.
AI-Modified Content	Content that originates from human input but has been altered, enhanced, or adapted by an AI system.
AI-Specific Security Incidents	Events where AI systems are compromised, misused, or malfunction due to vulnerabilities, attacks, or breaches, potentially impacting model integrity, data privacy, or system behaviour.
All user groups	All people including marginalized groups, people with disabilities (e.g., accessible to screen readers, including alt text for images, colour-blind friendly palettes, etc.).
Anonymization	The process of irreversibly transforming data so that an individual can no longer be identified, directly or indirectly.
Architecture	The architecture refers to the structural design of the AI model, including the number, size and properties of network layers. It defines how data flows through the model.
Asset inventory	A detailed list of all resources, components, or items owned or used within a system. It includes physical assets, software, data, and third-party solutions.
Associated Risks	The likelihood and potential impact of adverse effects, such as security vulnerabilities or functional failures, caused by handling non-conforming inputs.
Attack Vectors	Paths or methods leveraged by adversaries to exploit vulnerabilities in systems, such as data poisoning,

Term	Definition
	model manipulation, or injection of malicious triggers or backdoors.
Authenticity	The assurance that the transaction is genuine, unaltered, and can be verified as having been carried out by the AI system as intended, without unauthorized modifications or tampering.
Authenticity of data	The quality of being genuine, unaltered/altered with traces, and verifiably from a trusted source.
Automation bias	Occurs when a human decision-maker prefers recommendations made by an automated decision-making system, even when the system makes errors, over non-automated information.
Availability	Ability of a system, to remain accessible and operational for authorized users whenever needed.
Black Box	A scenario where attackers have no internal knowledge of the system or model and rely on observable input-output behaviour.
Boundary Values	Inputs at the edge of an allowed range or threshold, used to evaluate a system's performance and stability under extreme conditions.
Consistent logging	Systematic recording of events, actions, and decision made by the system.
Conventional IT security	Traditional methods and practices for protecting systems, data, and infrastructure from unauthorized access, misuse, or harm.
Coverage bias	Happens when the population represented in a dataset does not match the population that the machine learning model is making predictions about.
Critical level of bias	A threshold or degree of bias that, if exceeded, poses a significant risk to fairness, functionality, or compliance with ethical and legal standards.
Data acquisition	The process of collecting, measuring or obtaining raw data from various sources to be used for training and validating machine learning models.
Data characteristics	Attributes or properties that describe the nature, structure, quality, and behavior of data. Documenting these characteristics helps ensure that data is well-understood, appropriately used, and suitable for its intended purpose.
Data decommissioning	The process of securely and systematically retiring, archiving or disposing of data that is no longer needed.

Term	Definition
Data Management	The practice of organizing, storing, protecting and maintaining data to ensure its quality, accessibility, and usability throughout its lifecycle.
Data planning	The process of strategically defining how data will be collected, processed, managed, stored, analyzed, and maintained to meet organizational objectives and compliance requirements.
Data Preparation	The process of cleaning, transforming, and organizing raw data into a format suitable for analysis, modeling, or machine learning. This step ensures that data is of high quality and ready for use.
Data provisioning	The process of making data available to users, applications, or systems in a secure, efficient, and controlled manner. This can involve extracting, preparing, and delivering data from various sources to its intended destination.
Data quality	The degree to which data meets the standards for (among others) consistency, informativeness, representativeness, timeliness, validity, reliability, completeness.
Data split ratios	Proportions of data allocated to training, validation, and test sets.
Deep Learning	Deep learning is a subset of machine learning involving neural networks with multiple layers, enabling the model to learn complex patterns and make predictions from large amounts of data.
Disruptions	For example server crashes, network outages, high traffic or load, hardware failures, power failures etc.
End-User License Agreement (EULA)	A legal contract between the provider of software or services and the user, granting the user rights to use the software while outlining restrictions and obligations.
Error Report Evaluation	A systematic process for collecting, analyzing, categorizing, and addressing errors or malfunctions within the AI system to ensure continuous improvement, reliability, and safety.
Evasion Attacks	Methods where attackers modify inputs to trick a system or model into making wrong or unexpected decisions without changing the system itself.
Event of a failure	An incident where a system, component, or process does not perform as intended, leading to disrupted functionality, reduced performance, or complete inoperability. Common types of failures include

Term	Definition
	hardware failures, software failures, security failures, network failures or data failures.
Explanation of the system	An explanation of a system refers to the information provided to make the system's decisions, outputs, or behavior understandable to different stakeholders. This can include describing how, why, and under what conditions a decision or outcome was made.
Extraction Attacks	Methods where attackers try to obtain sensitive information from a model, such as its training data, parameters, or underlying logic, by analyzing its responses or internal structure.
Fairness	Fairness in AI systems refers to the principle that decisions and predictions should be made without prejudice or discrimination. A fair AI system ensures that all individuals or groups are treated equally, regardless of gender, ethnicity, age, or other protected characteristics. It involves ongoing measures to review, assess, and correct any unfair outcomes to maintain equitable treatment.
Finetuning	Finetuning is the process of tweaking a pre-trained machine learning model on a specific, smaller dataset to optimize its performance and accuracy for a particular task or domain.
Formal incident response procedures	Structured guidelines and processes for identifying, managing, and mitigating security incidents or breaches.
Formal Verification	The use of mathematical techniques to ensure that the system or its components operate as intended by checking against predefined rules and specifications.
General Purpose AI (GPAI)	Artificial intelligence systems designed to perform a wide range of tasks across various domains. These systems are adaptable and can solve diverse problems using the same core model, making them versatile for multiple applications.
Generalize	The ability of an AI system to perform well on new, unseen data. This demonstrates adaptability beyond examples used in training datasets.
Generative AI (GenAI)	A type of artificial intelligence that creates new content, such as text, images, or audio, by learning patterns from existing data.
Generative AI Foundation Model	A Generative AI Foundation Model is a type of AI technology that uses large amounts of data to generate new content, mimicking various forms of

Term	Definition
	human-like output such as text, images, or speech based on learned patterns and examples.
Graduated set of controls	Security measures that were implemented in layers or tiers.
Gray Box	A scenario where attackers have partial knowledge of the system or model, such as access to some parameters, architecture details, or limited input-output behaviour.
Group attribution bias	Occurs when a human assumes that what is true for one individual or object is also true for all individuals or objects in the group.
Group bias	The tendency of an AI system to favor or disadvantage certain groups based on patterns learned from biased data associated with group characteristics such as gender, ethnicity, or age.
Human Control	This refers to a pure assistance system that cannot trigger any reactions without human confirmation and is therefore completely dependent on human operation.
Human Oversight	The process in which humans closely monitor AI systems and intervene when necessary to ensure safe, ethical, and compliant operations.
Human-in-the-loop	This refers to a semi-autonomous AI application that is able to process tasks but cannot complete them independently and is dependent on the operation or confirmation of a human.
Human-on-the-Loop	This means that humans are rarely or not at all involved in decisions under normal conditions, but instead mainly monitor the AI application. It is possible to correct decisions retrospectively, even if immediate intervention is not possible at all times and in all places.
Human-out-of-the-Loop	This refers to the state in which an AI application acts fully autonomously under all conditions, including errors or unforeseen events
Impact analysis	An assessment of how changes or risks affect a system's functionality, security, and performance, helping identify risks and plan mitigations.
Incident	Any unexpected or incorrect behavior, malfunction, bias, or failure in the AI system that impacts performance, accuracy, safety, or user experience.
Individual bias	The tendency of an AI system to make skewed decisions based on patterns learned from biased data related to individual preferences or

Term	Definition
	characteristics, e.g. when a person with a university degree from a less known university applies to a job that is not recognized by an AI system.
Infrastructure	IT-Infrastructure that houses the key components of the AI System such as the database, computational resources used for training and inference and the model itself.
Integrity	In terms of data, integrity ensures that data remains unchanged, uncorrupted, and true to its original form through collection, storage, transmission and processing.
Intended purpose	The specific goal or function that the system is designed to achieve. It defines why something was created or implemented and what outcomes are expected from its use.
Interested Parties	Individuals or entities who have a vested interest in the AI system and require documentation to fulfill their roles or responsibilities. Examples include /- Developers and engineers/- End users/- Compliance officers and regulators/- Business stakeholders/- Supervisory authorities/- Partners
IP (Intellectual property)	Concerns content that is subject to copyright, trademark, or patent protection.
KPIs (Key Performance Indicators)	Metrics to evaluate the performance.
Large Language Model	A Large Language Model (LLM) is an AI-based model trained on vast datasets to understand and generate human-like text, enabling sophisticated tasks such as translation, summarization, and conversation.
Life cycle	The process from design and development through deployment, operation, and eventual decommissioning.
Machine Learning (ML)	A type of AI where systems learn from data to make predictions or decisions without explicit programming. It includes methods like supervised, unsupervised, and reinforcement learning.
Model Theft Attacks	Attempts by attackers to replicate or steal a machine learning model by exploiting access to its outputs, structure, or functionality.
Multiple Models	The practice of training and evaluating different machine learning algorithms or architectures on the same task to compare their performance, robustness, and suitability before selecting the final model.

Term	Definition
Non-Conforming Inputs	Inputs that deviate from defined standards, such as malformed, corrupted, or improperly formatted data, potentially affecting system integrity or performance.
Non-normality bias	Arises when statistical methods are applied to non-normally distributed data.
Non-Specified Users	Individuals who are either untrained or lack proper authorization to access the system.
On Site / Cloud / Hybrid Infrastructure	A piece of IT-Infrastructure is considered to be On Site, if it is physically present in location owned and controlled by the owner of the AI system. It is considered to be in the Cloud if it is located in a datacenter or compute cluster of a cloud provider such as Azure or AWS. An AI System can have parts of its infrastructure in the Cloud and parts on site in this case it is referred to as hybrid.
Output Integrity Attack	An attack that occurs when an adversary manipulates or alters the final results produced by the AI model without interfering with the input data or performing adversarial attacks on the model itself. In this type of attack, the input data remains untouched, but the generated output is modified or tampered with before it reaches the end user or decision-making system.
Overfitting	A situation where a model learns the training data too well, including noise and specific details, resulting in poor performance on new, unseen data due to lack of generalization.
Overreliance	The tendency of users to depend excessively on the AI system's outputs or recommendations, potentially leading to reduced human judgment, critical thinking, or oversight, even in cases where the AI might produce incorrect or incomplete results.
Override the system	Update or modify decisions or completely stop the operation of the system.
Performance	Indicator that shows how effectively and efficiently an AI system accomplishes its tasks.
Performance Requirements	Criteria that define the expected level of functionality and efficiency of the system, often measured through specific metrics like accuracy, speed, reliability, or resource usage, to ensure it meets its intended goals.
Personally Identifiable Data	Any information relating to an identified or identifiable natural person. This can include anything from a name, a photo, an email address,

Term	Definition
	bank details, posts on social networking websites, medical information, or even a computer IP address. The key aspect is that if the information can directly or indirectly identify a person, it qualifies as personal data.
Plausible	The AI system’s output should be credible, reasonable, and appropriate given the context of the inputs.
Poisoning Attacks	A type of adversarial attack where malicious data is injected into the training dataset to corrupt the model's learning process, leading to degraded performance, biased outputs, or hidden malicious behaviours.
Pre-Trained	A model is considered to be pretrained when it has already been trained on a large, general dataset before being customized or finetuned with specific data for targeted tasks. This allows the pretrained model to have a broad understanding which can then be adapted for particular applications.
Privacy	Privacy in AI systems refers to the protection of personal data and ensuring its confidential treatment throughout the system. A privacy-conscious AI system collects, processes and stores personal data only when necessary and in accordance with relevant data protection laws (such as GDPR). It also incorporates measures such as anonymization or pseudonymization, restricts access by unauthorized users, and ensures that individuals are informed about the use of their data and have given their consent.
Privacy Impact Analysis (PIA)	A systematic process to identify, assess, and mitigate risks to individuals' privacy posed by a system. It evaluates how personal data is collected, stored, processed, and shared, ensuring compliance with data protection laws and safeguarding individuals' rights.
Procedures (exemplary)	Monitoring and Logging, Self-Checking Mechanisms, Error Detection Algorithms, Threshold-Based Alerts, Redundancy Checks, User Feedback Loops
Pseudonymization	The process of replacing identifiable data with pseudonyms or identifiers so that individuals cannot be identified without additional information.
Quality Management System (QMS)	A structured framework that is concerned with the implementation, monitoring, and governance of AI systems.

Term	Definition
Reliability	Reliability in AI Systems refers to the ability of the system to consistently perform its intended functions over time and across varying conditions. A reliable AI system operates without significant downtime or failure, ensuring continuous performance even when subjected to diverse workloads or environmental changes. The goal is to ensure long-term operational stability and dependable outcomes.
Resilience	Resilience in AI systems refers to the ability of the system to maintain its functionality and stability in the face of adverse conditions, such as technical failures or unforeseen disruptions. A resilient AI system can self-monitor, detect failures, and take necessary recovery actions to ensure that it remains operational. This includes both the prevention of external threats and the system's capacity to recover quickly and fully after an incident.
Revocation of consent	The ability of individuals to withdraw their previously granted consent at any time, making further processing or storage of their data unlawful unless another legal basis exists.
Right to be forgotten	The right of individuals to have their personal data erased or deleted when it is no longer necessary for the purposes for which it was collected, or when consent has been withdrawn (GDPR Art. 17).
Right to deletion (Right to be forgotten)	The right of individuals to have their personal data erased or deleted when it is no longer necessary for the purposes for which it was collected, or when consent has been withdrawn (GDPR Art. 17).
Robustness	Robustness in AI systems refers to the ability to function reliably and consistently even under uncertain conditions, such as unexpected inputs or disturbances. A robust AI system delivers consistent and correct results on previously unseen data, even when faced with noisy or manipulated data in adversarial attacks. The goal is to ensure that the system is not easily disrupted by external influences.
Robustness Reviews	Complete evaluations of a model's ability to maintain functionality and performance under adversarial attacks, unexpected inputs, or challenging conditions.
Runtime Environment	The combination of hardware and software configurations required to execute a system or application, including operating systems,

Term	Definition
	frameworks, libraries, and physical devices essential for proper functionality and performance.
Sampling bias	Occurs when data records are not collected randomly from the intended population.
Secure Integrations	The process of safely connecting systems or components to ensure data integrity, confidentiality, and functionality.
Security	Security in AI systems refers to the protection of the underlying infrastructure, data, and communications from cybersecurity threats. A secure AI system employs various security mechanisms, such as encryption, authentication, and access controls, and includes regular security audits to safeguard against unauthorized access and misuse. The aim is to ensure that both data and system functionality are protected from internal and external threats, including potential cyber-attacks and data poisoning.
Security Controls	Measures and mechanisms implemented to safeguard systems, data, and processes against threats, ensuring confidentiality, integrity, and availability as per applicable regulations and standards.
Security Reviews	Systematic assessments to detect and address vulnerabilities in productive models, ensuring they are safeguarded against potential threats, breaches, and unauthorized access.
Self-error detection	Criteria and mechanisms for automatic error detection.
Sensitive Data	Special category of personal data which require more protection due to their sensitivity. This includes information about an individual's racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, genetic data, biometric data for the purpose of uniquely identifying a person, health data, and information concerning a person’s sex life or sexual orientation. Such data can only be processed under strict conditions.
Sensitive features	Features that were identified during the assessment in FAIR-01, e.g., gender, ethnicity, religion etc.
Shallow Learning	Shallow learning refers to machine learning methods that involve simpler, often linear models with fewer layers and complexity compared to deep learning, focusing on tasks that require less hierarchical feature extraction.

Term	Definition
Simpson’s paradox	Manifests when a trend indicated in individual groups of data reverses when the groups of data are combined.
Societal Impact	Societal Impact in AI systems refers to the broader ethical and social consequences of deploying AI technologies. A system designed with societal impact in mind ensures it contributes positively to society and avoids causing harm to communities, individuals, or the environment on all timescales. The goal is to align the AI system with societal values, ensuring its use promotes the common good and is both socially and environmentally sustainable.
Stability	In the context of data, it refers to the degree to which data remains consistent, reliable and unchanges over time or under varying conditions. It is critical for maintaining trust in data-driven processes, ensuring consistency, and supporting robust decision-making.
Stratification method	Criteria and approach for data stratification.
Supply chain	The network of entities and processes involved in creating, delivering, and maintaining a system.
System description	A detailed document ensuring transparency and ease of understanding.
System interfaces	Points of interaction where a system communicates with other systems, applications, or users.
Tailored Explanation	A specific, context-aware description provided to users, detailing the reasoning, factors, or data that led the AI system to produce a particular output or decision. The explanation is adapted to the user’s needs, expertise, and the complexity of the output.
Tampering	The unauthorized alteration, manipulation, or interference with data, inputs, or outputs to compromise the integrity, functionality, or security of the system.
Terms and Conditions	The legal agreement that defines the rules, guidelines, and expectations for users interacting with a service or product. It specifies user rights, responsibilities, and liabilities.
The Cloud Computing Compliance Criteria Catalogue	The Cloud Computing Compliance Criteria Catalogue (C5) by BSI outlines minimum security requirements for cloud applications.
Third Party	An external entity (not directly affiliated the system's owner), responsible for providing training data, models, or other services.

Term	Definition
Third-Party Model	A model that was developed and trained by a third party, i.e. not trained and developed by the user of the model.
Traceability	Refers to the ability to track, document, and verify the origin, history and transformation of data throughout its lifecycle.
Training Data	Training data is a dataset used to teach machine learning models how to recognize patterns and make decisions, forming the foundational knowledge that guides their predictions and actions.
Transaction	Any action or event initiated or completed by the AI system that results in a change of state, data exchange, or processing outcome.
Transparency	Transparency in AI systems refers to the clarity and openness with which the system's operations, decision-making processes, and data handling practices are communicated. A transparent AI system ensures that stakeholders, including users and regulators, can understand how decisions are made and how data is used. This involves clear documentation of algorithms, data sources, and decision criteria, as well as providing explanations for the system's outputs.
Uncertainty Assessment	The process of evaluating and quantifying the confidence or reliability of an AI system's predictions or outputs, often by analyzing how variations in inputs or conditions impact its performance and decision-making.
Underfitting	A situation where a model fails to capture the underlying patterns in the training data, leading to poor performance both on the training data and unseen data.
Uniquely Identifiable	Each transaction has a distinct identifier (e.g., a transaction ID or hash) that makes it distinguishable from all other transactions.
Unspecified Usage Environments	Physical or operational settings that are not explicitly defined.
User Awareness	Ensuring users clearly understand that they are interacting with an AI system rather than a human.
User Feedback	Any input, complaint, suggestion, or observation provided by users or stakeholders regarding their experience with the system.
White-Box	Scenarios where attackers have full access to the internal structure, parameters, and algorithms of a

Term	Definition
	system or model, allowing detailed analysis and exploitation.

18 Appendix A – Additional Regulation relevant to an IRB Use Case by Criterion

Criterion ID	Finance Regulation (IRB Scoring Use Case)
AISR-01	IRBA-RTS Art. 75
AISR-02	IRBA-RTS Art. 75
AISR-03	CRR Art. 144 EBA/GL/2020/06 4.3.4
AISR-04	IRBA-RTS Art. 75
AISR-05	IRBA-RTS Art. 75
AISR-06	EBA/GL/2020/06 4.3.3, 4.3.4 MaRisk AT 4.3.4
AISR-07	IRBA-RTS Art. 75
AISR-08	IRBA-RTS Art. 75
AISR-09	IRBA-RTS Art. 75
AISR-10	IRBA-RTS Art. 75
AISR-11	IRBA-RTS Art. 75
AISR-12	IRBA-RTS Art. 75
AISR-13	IRBA-RTS Art. 75
AISR-14	IRBA-RTS Art. 75
AISR-15	IRBA-RTS Art. 75
AISR-16	IRBA-RTS Art. 75
AISR-17	IRBA-RTS Art. 75
AISR-18	IRBA-RTS Art. 75
AISR-19	IRBA-RTS Art. 75
DEV-01	CRR Art. 144, 174, 175, 179 IRBA-RTS Art. 3, 31, 35, 40 EBA/GL/2020/06 4.3.4 EGIM Overreaching 2, 9.7 MaRisk AT 6
DEV-02	CRR Art. 144, 174, 175, 179 IRBA-RTS Art. 3, 31, 35 , 40 EBA/GL/2020/06 4.3.4 EGIM Overreaching 2, 9.7 MaRisk AT 6

Criterion ID	Finance Regulation (IRB Scoring Use Case)
DEV-03	CRR Art. 144, 174, 175, 179 IRBA-RTS Art. 3, 31, 35, 40 EBA/GL/2020/06 4.3.4 EGIM Overreaching 2, 9.7 MaRisk AT 6
DEV-04	IRBA-RTS Art. 75
DEV-05	N/A
DEV-06	CRR Art. 174, 185 IRBA-RTS Art. 40 EBA/GL/2020/06 4.3.3, 4.3.4 MaRisk AT 4.3.5
DQM-01	CRR Art. 175 IRBA-RTS Art. 3, 31, 69, Chapter 12 EBA/GL/2020/06 4.3.4 EBA/GL/2017/16 4.2 EGIM Overreaching 9.5, Credit risk 8 MaRisk AT 4.3.4, 4.3.5
DQM-02	CRR Art. 175 IRBA-RTS Art. 3, 31, 69, Chapter 12 EBA/GL/2020/06 4.3.4 EBA/GL/2017/16 4.2 EGIM Overreaching 9.5, Credit risk 8 MaRisk AT 4.3.4, 4.3.5
DQM-03	CRR Art. 175 IRBA-RTS Art. 3, 31, 69 and Chapter 12 EBA/GL/2020/06 4.3.4 EBA/GL/2017/16 4.2 EGIM Overreaching 9.5, Credit risk 8 MaRisk AT 4.3.4, 4.3.5
DQM-04	CRR Art. 175 IRBA-RTS Art. 3, 31, 69 and Chapter 12 EBA/GL/2020/06 4.3.4 EBA/GL/2017/16 4.2 EGIM Overreaching 9.5 Credit risk 8 MaRisk AT 4.3.4, 4.3.5
DQM-05	CRR Art. 175 IRBA-RTS Art. 3, 31, 69 and Chapter 12 EBA/GL/2020/06 4.3.4 EBA/GL/2017/16 4.2 EGIM Overreaching 9.5, Credit risk 8 MaRisk AT 4.3.4, 4.3.5

Criterion ID	Finance Regulation (IRB Scoring Use Case)
DQM-06	CRR Art. 175 IRBA-RTS Art. 3, 31, 69 and Chapter 12 EBA/GL/2020/06 4.3.4 EBA/GL/2017/16 4.2 EGIM Overreaching 9.5, Credit risk 8 MaRisk AT 4.3.4, 4.3.5
DQM-07	CRR Art. 175 IRBA-RTS Art. 3, 31, 69 and Chapter 12 EBA/GL/2020/06 4.3.4 EBA/GL/2017/16 4.2 EGIM Overreaching 9.5, Credit risk 8 MaRisk AT 4.3.4, 4.3.5
DQM-08	N/A
DQM-09	N/A
DQM-10	N/A
DQM-11	CRR Art. 175 IRBA-RTS Art. 3, 31
DQM-12	CRR Art. 175 IRBA-RTS Art. 3, 31, 69 and Chapter 12 EBA/GL/2020/06 4.3.4 EBA/GL/2017/16 4.2 EGIM Overreaching 9.5, Credit risk 8 MaRisk AT 4.3.4, 4.3.5
DQM-13	CRR Art. 175 IRBA-RTS Art. 3, 31, 69 and Chapter 12 EBA/GL/2020/06 4.3.4 EBA/GL/2017/16 4.2 EGIM Overreaching 9.5, Credit risk 8 MaRisk AT 4.3.4, 4.3.5
DQM-14	CRR Art. 175 IRBA-RTS Art. 3, 31, 69 and Chapter 12 EBA/GL/2020/06 4.3.4 EBA/GL/2017/16 4.2 EGIM Overreaching 9.5, Credit risk 8 MaRisk AT 4.3.4, 4.3.5
FAIR-01	CRR Art. 174, 175, 179, 180 IRBA-RTS Art. 3, 31, 35, 40, 42 EBA/GL/2020/06 4.3.3, 4.3.4 EBA/GL/2017/16 4.2
FAIR-02	CRR Art. 174, 175, 179, 180 IRBA-RTS Art. 3, 42 EBA/GL/2020/06 4.3.3, 4.3.4 EBA/GL/2017/16 4.2

Criterion ID	Finance Regulation (IRB Scoring Use Case)
FAIR-03	CRR Art. 174, 175, 179, 180 IRBA-RTS Art. 3, 42 EBA/GL/2020/06 4.3.3, 4.3.4 EBA/GL/2017/16 4.2
FAIR-04	N/A
GOV-01	CRR Art. 175 IRBA-RTS Art. 3, 31 EBA/GL/2020/06 4.3.4 EGIM Overreaching 2, 9.2 MaRisk AT 6 Dokumentation
GOV-02	The term QMS, which is commonly used for product regulation, is not a term used in financial sector regulation. Comparable elements can be found here: CRR Art. 185, 190 MaRisk AT 4.3, 4.4, 5, 6
GOV-03	N/A
GOV-04	N/A
GOV-05	N/A
GOV-06	CRR Art. 175 IRBA-RTS Art. 3, 31 EBA/GL/2020/06 4.3.4 EGIM Overreaching 2, 9.2 MaRisk AT 6
GOV-07	CRR Art. 175 IRBA-RTS Art. 3, 31 EBA/GL/2020/06 4.3.4 EGIM Overreaching 2, 9.2 MaRisk AT 6
GOV-08	IRBA-RTS Art. 38
GOV-09	N/A
GOV-10	N/A
GOV-11	CRR Art. 175, 189 IRBA-RTS Art. 3, 14, 15, 16, 31 EGIM Overreaching 9.2, Credit Risk 3
GOV-12	CRR Art. 175, 189 IRBA-RTS Art. 3, 14, 15, 16, 31 EGIM Overreaching 9.2, Credit Risk 3
GOV-13	CRR Art. 174, 185, 189, 191 EGIM Overreaching 9.2, Credit Risk 3 MaRisk AT 7.1

Criterion ID	Finance Regulation (IRB Scoring Use Case)
HO-01	CRR Art. 174, 180 IRBA-RTS Art. 39 EBA/GL/2020/06 4.3.4 EGIM Overreaching 9.8, Credit risk 6 MaRisk AT 4.3.5
HO-02	CRR Art. 174, 180 IRBA-RTS Art. 39 EBA/GL/2020/06 4.3.4 EGIM Overreaching 9.8, Credit risk 6 MaRisk AT 4.3.5
HO-03	CRR Art. 174, 175, 180 IRBA-RTS Art. 3, 31, 39 EBA/GL/2020/06 4.3.4 EGIM Overreaching 9.8, Credit risk 6 MaRisk AT 4.3.5, 5, 6
HO-04	CRR Art. 175, 189, 190 IRBA-RTS Art. 3, 14, 15, 16, 31 MaRisk AT 5, 6 EGIM Overreaching 9.2, Credit Risk 3
ITSEC-01	MaRisk AT 5
ITSEC-02	IRBA-RTS Art. 75
ITSEC-03	IRBA-RTS Art. 75
ITSEC-04	IRBA-RTS Art. 75
ITSEC-05	IRBA-RTS Art. 75
ITSEC-06	IRBA-RTS Art. 75
ITSEC-07	IRBA-RTS Art. 75
ITSEC-08	IRBA-RTS Art. 75
ITSEC-09	EGIM Credit risk 8
ITSEC-10	IRBA-RTS Art. 75
ITSEC-11	IRBA-RTS Art. 75
ITSEC-12	IRBA-RTS Art. 75
MON-01	CRR Art. 174, 175, 185 IRBA-RTS Art. 3, 31, 40 EBA/GL/2020/06 4.3.3, 4.3.4 EGIM Overreaching 6, 9.3, Credit risk 4 MaRisk AT 4.3.5
MON-02	N/A
MON-03	N/A
MON-04	N/A

Criterion ID	Finance Regulation (IRB Scoring Use Case)
PERF-01	CRR Art. 174, 175, 185 IRBA-RTS Art. 3, 31, 40 EBA/GL/2020/06 4.3.3, 4.3.4 EGIM Overreaching 6, 9.3, Credit risk 4 MaRisk AT 4.3.5
PERF-02	CRR Art. 174, 175, 185 IRBA-RTS Art. 3, 31, 40 EBA/GL/2020/06 4.3.3, 4.3.4 EGIM Overreaching 6, 9.3, Credit risk 4 MaRisk AT 4.3.5
PERF-03	CRR Art. 174, 175, 185 IRBA-RTS Art. 3, 31, 40 EBA/GL/2020/06 4.3.3, 4.3.4 EGIM Overreaching 6, 9.3, Credit risk 4 MaRisk AT 4.3.5
PERF-04	CRR Art. 174, 175, 185 IRBA-RTS Art. 3, 31, 40 EBA/GL/2020/06 4.3.3, 4.3.4 EGIM Overreaching 6, 9.3, Credit risk 4 MaRisk AT 4.3.5
PERF-05	CRR Art. 174, 175, 185 IRBA-RTS Art. 3, 31, 40 EBA/GL/2020/06 4.3.3, 4.3.4 EGIM Overreaching 6, 9.3, Credit risk 4 MaRisk AT 4.3.5
PERF-06	CRR Art. 174, 175, 185 IRBA-RTS Art. 3, 31, 40 EBA/GL/2020/06 4.3.3, 4.3.4 EGIM Overreaching 6, 9.3, Credit risk 4 MaRisk AT 4.3.5
PERF-07	IRBA-RTS Art. 75
TR-01	CRR Art. 174, 175, 179, 185, 189, 190 IRBA-RTS Art. 3, 31, 39 EBA/GL/2020/06 4.3.3, 4.3.4 MaRisk AT 4.3.5
TR-02	N/A
TR-03	CRR Art. 174, 179, 185, 189 IRBA-RTS Art. 39 EBA/GL/2020/06 4.3.3, 4.3.4 MaRisk AT 4.3.5
TR-04	CRR Art. 144, 174 IRBA-RTS Art. 39 EGIM Overreaching 9.8, Credit risk 6 MaRisk AT 4.3.5

Criterion ID	Finance Regulation (IRB Scoring Use Case)
TR-05	N/A
TR-06	N/A
TR-07	CRR Art. 174, 179, 185, 189 IRBA-RTS Art. 39 EBA/GL/2020/06 4.3.3, 4.3.4 MaRisk AT 4.3.5
TR-08	CRR Art. 175 IRBA-RTS Art. 3, 31 EBA/GL/2020/06 4.3.4 EGIM Overreaching 2, 9.7 MaRisk AT 6
TR-09	N/A
TR-10	N/A
TR-11	N/A

19 Appendix B – Reevaluation triggers mapped to testing criteria

Trigger / change	Abbreviation	Related criteria
Advanced tools	AT	AISR-02, AISR-04, AISR-05, AISR-07, AISR-08, AISR-09, AISR-10, AISR-11, AISR-12, AISR-13, AISR-14, AISR-15, AISR-16, AISR-17, AISR-18, AISR-19 GOV-03, GOV-09, GOV-10 HO-02 ITSEC-07 PERF-02, PERF-03, PERF-04, PERF-05, PERF-06 TR-03, TR-06, TR-07
Application context	AC	AISR-02, AISR-05, AISR-09, AISR-10 DEV-01, DEV-02 FAIR-04 GOV-06, GOV-07 HO-01, HO-02 ITSEC-03, ITSEC-09, ITSEC-12 TR-01, TR-04, TR-05, TR-06, TR-07, TR-08, TR-09
Error messages	EM	AISR-03, AISR-05 DEV-04, DEV-05 GOV-02, GOV-11, GOV-13 HO-01, HO-02 ITSEC-05, ITSEC-07, ITSEC-08, ITSEC-09 MON-02, MON-04 PERF-07 TR-1

Trigger / change	Abbreviation	Related criteria
Ethical guidelines	EG	AISR-04 DQM-06, DQM-07, DQM-08, DQM-09, DQM-10 FAIR-01, FAIR-02, FAIR-03 GOV-09, GOV-10 ITSEC-01, ITSEC-10, ITSEC-11 TR-04, TR-05, TR-06, TR-07, TR-09
Internal guidelines	IG	AISR-01, AISR-04, AISR-05, AISR-07, AISR-08, AISR-11, AISR-12 DEV-02, DEV-03, DEV-05 DQM-01, DQM-02, DQM-03, DQM-04, DQM-05, DQM-06, DQM-07, DQM-08, DQM-09, DQM-10, DQM-11, DQM-12, DQM-13, DQM-14 FAIR-01, FAIR-02, FAIR-03, FAIR-04 GOV-01, GOV-02, GOV-04, GOV-05, GOV-08, GOV-09, GOV-10, GOV-11, GOV-12, GOV-13 HO-01, HO-02, HO-03, HO-04 ITSEC-07, ITSEC-10, ITSEC-12 MON-03 PERF-01, PERF-07 TR-02, TR-10

Trigger / change	Abbreviation	Related criteria
Legal requirements	LR	AISR-01, AISR-03, AISR-04, AISR-05, AISR-06, AISR-07, AISR-08, AISR-09, AISR-10, AISR-11, AISR-12, AISR-13, AISR-14, AISR-15, AISR-16, AISR-17, AISR-18, AISR-19 DEV-02, DEV-03, DEV-05 DQM-01, DQM-02, DQM-03, DQM-04, DQM-05, DQM-06, DQM-07, DQM-08, DQM-09, DQM-10, DQM-11, DQM-12, DQM-13, DQM-14 FAIR-01, FAIR-02, FAIR-03, FAIR-04 GOV-02, GOV-03, GOV-04, GOV-05, GOV-08, GOV-09, GOV-10, GOV-11, GOV-12, GOV-13 HO-01, HO-02, HO-03, HO-04 ITSEC-04, ITSEC-05, ITSEC-06, ITSEC-07, ITSEC-10, ITSEC-11, ITSEC-12 MON-01, MON-02, MON-03, MON-04 PERF-01, PERF-07 TR-01, TR-02, TR-04, TR-05, TR-06, TR-07, TR-09, TR-10, TR-11
Model architecture	MA	AISR-02, AISR-03, AISR-05, AISR-06, AISR-07, AISR-08, AISR-09, AISR-10, AISR-11, AISR-12, AISR-14, AISR-15, AISR-16, AISR-17, AISR-18, AISR-19 DEV-01, DEV-02, DEV-03, DEV-05, DEV-06 GOV-03, GOV-06, GOV-07, GOV-09, GOV-11 HO-04 PERF-02, PERF-03, PERF-04, PERF-06 TR-01, TR-02, TR-03, TR-08

Trigger / change	Abbreviation	Related criteria
Model parameters	MP	AISR-03, AISR-05, AISR-06, AISR-07, AISR-08, AISR-09, AISR-10, AISR-11, AISR-12, AISR-14, AISR-15, AISR-16, AISR-17, AISR-18, AISR-19 DEV-01, DEV-05 GOV-06, GOV-07, GOV-11 HO-04 PERF-02, PERF-03, PERF-04, PERF-06 TR-01, TR-02, TR-03
New vulnerability	NV	AISR-02, AISR-04, AISR-05, AISR-07, AISR-08, AISR-09, AISR-10, AISR-11, AISR-12, AISR-14, AISR-15, AISR-16, AISR-17, AISR-18, AISR-19 GOV-03, GOV-09, GOV-10 ITSEC-01
Sociatel impact	SI	FAIR-01, FAIR-02, FAIR-03
Standards, norms, or best practices	SN	AISR-01, AISR-04, AISR-05, AISR-06, AISR-07, AISR-08, AISR-09, AISR-10, AISR-11, AISR-12, AISR-13, AISR-14, AISR-15, AISR-16, AISR-17, AISR-18, AISR-19 DEV-06 DQM-03, DQM-04, DQM-06, DQM-07, DQM-09, DQM-11, DQM-12, DQM-13, DQM-14 FAIR-02, FAIR-03 GOV-03, GOV-04, GOV-09, GOV-10 HO-01 ITSEC-01, ITSEC-07, ITSEC-08, ITSEC-09, ITSEC-10 MON-01, MON-02, MON-03, MON-04 TR-04, TR-05, TR-06, TR-07, TR-11

Trigger / change	Abbreviation	Related criteria
System expansion	SE	AISR-01, AISR-13 DEV-01, DEV-02, DEV-03, DEV-04 GOV-01, GOV-02, GOV-04, GOV-06, GOV-07, GOV-11, GOV-12, GOV-13 HO-04 ITSEC-02, ITSEC-04, ITSEC-06 MON-01, MON-02, MON-03, MON-04 TR-08
System reduction	SR	AISR-01, AISR-13 DEV-01, DEV-04 GOV-01, GOV-02, GOV-04, GOV-06, GOV-07, GOV-11, GOV-12, GOV-13 HO-04 ITSEC-02, ITSEC-04, ITSEC-06 MON-01, MON-02, MON-03, MON-04 TR-08
Technical standards	TS	AISR-02, AISR-04, AISR-05, AISR-06, AISR-07, AISR-08, AISR-09, AISR-10, AISR-11, AISR-12, AISR-13, AISR-14, AISR-15, AISR-16, AISR-17, AISR-18, AISR-19 DEV-06 DQM-03, DQM-04, DQM-06, DQM-07, DQM-09, DQM-11, DQM-12, DQM-13, DQM-14 GOV-03, GOV-09, GOV-10 HO-02 ITSEC-01, ITSEC-07, ITSEC-08, ITSEC-09 PERF-02, PERF-03, PERF-04, PERF-05, PERF-06 TR-03, TR-06, TR-07, TR-11

Trigger / change	Abbreviation	Related criteria
Training data	TD	AISR-03, AISR-05, AISR-06, AISR-07, AISR-08, AISR-09, AISR-10, AISR-11, AISR-12, AISR-14, AISR-15, AISR-16, AISR-17, AISR-18, AISR-19 DEV-01, DEV-05 DQM-01, DQM-02, DQM-03, DQM-04, DQM-05, DQM-06, DQM-07, DQM-10, DQM-11, DQM-12, DQM-13, DQM-14 FAIR-01, FAIR-02 GOV-06, GOV-07, GOV-11 HO-04 ITSEC-10, ITSEC-11 PERF-02, PERF-03, PERF-04, PERF-06 TR-01, TR-02, TR-03
Unexpected results	UR	AISR-03 DEV-04, DEV-05, DEV-06 FAIR-03 GOV-02, GOV-11 HO-01, HO-02 ITSEC-09 MON-01 PERF-06
Unstable performance	UP	AISR-03, AISR-05 DEV-04, DEV-05, DEV-06 FAIR-03 GOV-02, GOV-11 ITSEC-09 MON-01 PERF-01, PERF-03, PERF-04, PERF-05

Trigger / change	Abbreviation	Related criteria
User group or permissions	UG	AISR-07, AISR-08, AISR-09, AISR-10, AISR-11, AISR-12 DEV-01 DQM-08 FAIR-01, FAIR-02, FAIR-03, FAIR-04 GOV-06, GOV-07, GOV-11 HO-02, HO-03 ITSEC-03, ITSEC-09, ITSEC-10, ITSEC-12 TR-04, TR-05, TR-07, TR-08, TR-09