APPENDIX I:
REQUIREMENTS SPECIFICATION

ID	<X.X> (<Obligation ID>)
Description	<Description of the requirement>
Rationale	<Art. X (X); short rationale in own words>
Difficulty	<low, medium, high>
Fit Criterion	<As precise as possible: how will/can the requirement be evaluated?>
Type	<functional, non-functional, process>
Applicability	<All (high-risk AI systems) / Restricted (Details, Ref.)>
Category	<Category or Sub-Category, if applicable>

APPENDIX I.1: RISK MANAGEMENT SYSTEM (ART. 9)

ID	9.1 (O1)
Description	A risk management system shall exist that is maintained and documented.
Rationale	Art. 9 (1); The risks from AI systems need to be understood and controlled
Difficulty	low
Fit Criterion	A risk management system is continually operating and accessible by a user that has access to its documentation.
Type	Functional Requirement
Applicability	All (high-risk AI systems)
Category	Risk Management

ID	9.2 (O2)
Description	The risk management system shall operate through the entire lifetime of the high-risk AI system as a continuous iterative process.
Rationale	Art. 9 (2); The risks from AI systems need to be evaluated continuously
Difficulty	medium
Fit Criterion	After multiple fixed periods of operation of an AI system, respectively, the risk management system accessible by a user is still operating and updated to potentially changed circumstances with respect to the high-risk AI system.
Type	Functional Requirement
Applicability	All (high-risk AI systems), given req. 9.1 is fulfilled.
Category	Risk Management

ID	9.3 (O3)
Description	The risk management system shall have the ability to identify all known and foreseeable risks with respect to the high-risk AI system.
Rationale	Art. 9 (2)(a); The risks from AI systems need to be identified in order to be treated
Difficulty	high
Fit Criterion	The risks with respect to multiple known high-risk AI systems returned to an expert user from the risk management system match at-large the risks of these systems known beforehand.
Type	Functional Requirement
Applicability	All (high-risk AI systems), given req. 9.1 is fulfilled
Category	Risk Management

ID	9.4 (O4)
Description	The risk management system shall have the ability to evaluate and estimate the risks with respect to the high-risk AI system that arise from its purpose-conform use, reasonably foreseeable misuse, or the output from a post-market monitoring system according to requirements 12.4-12.6.
Rationale	Art. 9 (2)(b), (c); The risks from AI systems need to be evaluated and characterised in order to be treated
Difficulty	high
Fit Criterion	The evaluation of risks from ordinary use, foreseeable misuse, and post-market monitoring mechanisms with respect to multiple known high-risk AI systems returned to an expert user match at-large his evaluation of these risks.
Type	Functional Requirement
Applicability	All (high-risk AI systems), given req. 9.1 is fulfilled
Category	Risk Management

ID	9.5 (O5, O6)
Description	The risk management system shall adopt risk management measures that duly consider the effects and possible interactions from the entirety of the requirements defining the high-risk AI system in this Requirements Specification.
Rationale	Art. 9 (2)(d) + Art. 9 (3); Unexpected risks may arise from any AI system established according to a variety of independent requirements
Difficulty	high
Fit Criterion	An expert user is unable to identify any risks from the interactions of the requirements established in this Requirements Specification that define the AI system that were already identified by the risk management system.
Type	Functional Requirement
Applicability	All (high-risk AI systems), given req. 9.1 is fulfilled
Category	Risk Management

ID	9.6 (O5, O7)
Description	The risk management system shall adopt risk management measures that operate according to the industrial standard, for example through harmonised standards or common specification.
Rationale	Art. 9 (2)(d) + Art. 9 (3); Pre-existing standards and common practices in risk management system are applicable and useful to high-risk AI systems
Difficulty	medium
Fit Criterion	A proficient risk management engineer verifies that the risk management system measures conform to the most appropriate standard or common practices, if any.
Type	Functional Requirement
Applicability	Restricted (high-risk AI systems whose risks are applicable to common practices or standards), given req. 9.1 is fulfilled
Category	Risk Management

ID	9.7 (O5, O8)
Description	The risk management system shall adopt risk management measures that ensure that residual risks from a high-risk AI system used according to its purpose or under condition of reasonably foreseeable misuse associated with each hazard and overall is judged acceptable.
Rationale	Art. 9 (2)(d) + Art. 9 (4); A risk management system is only sufficiently effective when the residual, non-treatable risks are acceptable
Difficulty	high
Fit Criterion	None of the evaluations of residual risks returned from the risk management with respect to a high-risk AI system used according to its purpose or under condition of reasonably foreseeable misuse is classified worse than acceptable or some equivalent threshold.
Type	Functional Requirement
Applicability	All (high-risk AI systems), given req. 9.1 is fulfilled
Category	Risk Management

ID	9.8 (O5, O9)
Description	The risk management system shall communicate all residual risks to the user.
Rationale	Art. 9 (2)(d) + Art. 9 (4); Residual risks may only be act upon when communicated to the user of the high-risk AI system
Difficulty	low
Fit Criterion	The system returns all of the identified residual risks according to 9.7 to the user via an appropriate interface.
Type	Functional Requirement
Applicability	All (high-risk AI systems), given req. 9.1 is fulfilled
Category	Risk Management

ID	9.9 (O5, O10)
Description	The risk management system shall adopt risk management measures such that the high-risk AI system's architecture and implementation minimises the risks associated with its purpose-conform use or reasonably foreseeable misuse.
Rationale	Art. 9 (2)(d) + Art. 9 (4)(a); The use of an effective risk management system is intended to lead to the elimination of risks
Difficulty	medium
Fit Criterion	No other risk management measures can be identified by a risk management engineer the use of which would yield a further reduction of risks in the operation of the high-risk AI system.
Type	Functional Requirement
Applicability	All (high-risk AI systems), given req. 9.1 is fulfilled
Category	Risk Management

ID	9.10 (O5, O11)
Description	The risk management system shall adopt risk management measures such that the high-risk AI system's implementation includes adequate mitigation and control measures for residual risks associated with its purpose-conform use or reasonably foreseeable misuse that cannot be eliminated.
Rationale	Art. 9 (2)(d) + Art. 9 (4)(b); The use of an effective risk management system is intended to lead to the control and mitigation of risks
Difficulty	medium
Fit Criterion	No other risk management measures can be identified by a risk management engineer the use of which would yield more effective risk control and mitigation measures within the implementation of the high-risk AI system.
Type	Functional Requirement
Applicability	Restricted (High-risk AI systems with residual risks after application of risk management measures according to 9.9), given req. 9.1 is fulfilled
Category	Risk Management

ID	9.11 (O5, O12)
Description	The risk management system shall adopt risk management measures such that adequate information is provided to users about the risks associated with its purpose-conform use or reasonably foreseeable misuse of the high-risk AI system (see also requirement 13.4).
Rationale	Art. 9 (2)(d) + Art. 9 (4)(c); Risks may only be act upon when communicated to the user of the high-risk AI system
Difficulty	low
Fit Criterion	An expert user is provided with information according to requirements 13.1 through 13.10 about the risks associated with its purpose-conform use or reasonably foreseeable misuse before or shortly after beginning of their use.
Type	Functional Requirement
Applicability	All (High-risk AI systems), given req. 9.1 is fulfilled
Category	Risk Management

ID	9.12 (O5, O13)
Description	The risk management system shall adopt risk management measures such that adequate training, considering requirements 13.1 through 13.10, is provided to users.
Rationale	Art. 9 (2)(d) + Art. 9 (4)(c); Risks may only be act upon when the user of the high-risk AI system is proficient in dealing with them
Difficulty	medium
Fit Criterion	A user is provided with training before or shortly after beginning of their use.
Type	Functional Requirement
Applicability	Restricted (High-risk AI systems with risks for which training is appropriate), given req. 9.1 is fulfilled
Category	Risk Management

ID	9.13 (O5, O14)
Description	The risk management system shall adopt risk management measures such that in eliminating or reducing risks due consideration is given to the technical knowledge, experience, education, training to be expected by the user, and the environment in which the system is intended to be used.
Rationale	Art. 9 (2)(d) + Art. 9 (4); When acting upon risks in a high-risk AI system, the accumulated circumstances of use must be duly considered to allow the most accurate evaluation and the derive the most appropriate counter measures
Difficulty	high
Fit Criterion	The ways to eliminate and reduce risks of a high-risk AI system proposed by the risk management system are different between a target user with more and less technical proficiency.
Type	Process Requirement
Applicability	All (high-risk AI systems), given req. 9.1 is fulfilled
Category	Risk Management

ID	9.14 (O5, O15)
Description	The high-risk AI system shall be tested with the purpose of identifying appropriate risk management measures.
Rationale	Art. 9 (2)(d) + Art. 9 (5); Testing a high-risk AI system reveals the risks associated with its use that are hard to expect or predict
Difficulty	low
Fit Criterion	The risk management measures adopted in the finalised risk management system were informed by the results of a technical testing procedure performed on the high-risk AI system.
Type	Non-Functional Requirement
Applicability	All (high-risk AI systems), given req. 9.1 is fulfilled
Category	Testing

ID	9.15 (O15, O16)
Description	Testing procedures shall assess whether the high-risk AI system performs consistently for their intended purpose.
Rationale	Art. 9 (5); Only consistent performance of the intended objective renders an AI system reliable
Difficulty	medium
Fit Criterion	The test of multiple AI systems known to operate inconsistently showcases to the user that that is the case.
Type	Non-Functional Requirement
Applicability	All (high-risk AI systems), given req. 9.14, 15.2, and 15.3 are fulfilled
Category	Testing

ID	9.16 (O15, O18)
Description	The testing procedures shall be appropriate to the intended purpose of the high-risk AI system.
Rationale	Art. 9 (6); Testing sufficiently fulfils its intent when it relates to the intended purpose of the high-risk AI system
Difficulty	low
Fit Criterion	Each testing procedures corresponds to some aspect of the intended purpose of the high-risk AI system and all aspects of the intended purpose are covered by a test
Type	Non-Functional Requirement
Applicability	All (high-risk AI systems), given req. 9.14 is fulfilled
Category	Testing

ID	9.17 (O15, O19)
Description	The testing procedures and response to their results shall be performed before the high-risk AI system's entry into market or putting into service.
Rationale	Art. 9 (7); Testing only fulfils its intent when it allows to fix shortcomings before the high-risk AI system is used in production and affecting real users
Difficulty	low
Fit Criterion	The high-risk AI system on the market was tested beforehand.
Type	Process Requirement
Applicability	All (high-risk AI systems), given req. 9.14 is fulfilled
Category	Testing

ID	9.18 (O15, O20)
Description	The testing procedures shall be based on preliminarily defined metrics and probabilistic thresholds appropriate to the intended purpose of the high-risk AI system.
Rationale	Art. 9 (7); To ensure comparability and expressibility, testing of a high-risk AI system must be based in recognised metrics and threshold values of these metrics that determine the system's suitability
Difficulty	medium
Fit Criterion	The output of a test is presented in an industry-recognised metrics and a qualitative result associated with it is based on one or multiple threshold values of that metric
Type	Non-Functional Requirement
Applicability	All (high-risk AI systems), given req. 9.14 is fulfilled
Category	Testing

ID	9.19 (O15, O21)
Description	The risk management system shall assess and respond when the high-risk AI system is likely to be accessed by or have an impact on children.
Rationale	Art. 9 (8); A high-risk AI system affecting children imposes special risks on them that are required to be addressed and mitigated accordingly
Difficulty	medium
Fit Criterion	The output and/or behaviour of a risk management system assessing a high-risk AI system impacting children differs from that of assessing the same system without impact on children
Type	Functional Requirement
Applicability	All (high-risk AI systems), given req. 9.14 is fulfilled
Category	Risk Management

ID	9.20 (O15, O22)
Description	The risk management system shall form part of risk management procedures set out in article 74 of Directive 2013/36/EU.
Rationale	Art. 9 (9); The risks from high-risk AI systems add to intrinsic risks in the financial services industry and need to be jointly mitigated
Difficulty	low
Fit Criterion	The high-risk AI system's risk management system is included in the documentation of risk management measures and their output communicated to the authorities
Type	Process Requirement
Applicability	Restricted (high-risk AI systems deployed by credit institutions regulated by Directive 2013/36/EU), given req. 9.14 is fulfilled
Category	Risk Management

APPENDIX I.2: DATA AND DATA GOVERNANCE (ART. 10)

ID	10.1 (O3)
Description	Data governance and management practices shall concern relevant design choices (e.g., data features, AI system/data platform architecture).
Rationale	Art. 10 (2a); Design choices impact the quality and safety of the data sets which is needed to prevent attacks (e.g., adversarial examples, social engineering).
Difficulty	high
Fit Criterion	A group of experts identifies that data governance and management practices deal with relevant design choices or an appropriate standard (e.g., ISO/IEC JTC 1/SC 42) is used.
Type	Process Requirement
Applicability	All (high-risk AI systems)
Category	Dataset Properties

ID	10.2 (O4)
Description	Data governance and management practices shall concern the collection of data sets.
Rationale	Art. 10 (2b); The collection of data needs to comply with relevant data governance rules (e.g., GDPR).
Difficulty	low
Fit Criterion	The processes for collecting data comply with defined data governance rules and this is validated by a group of people responsible for data governance and management.
Type	Process Requirement
Applicability	All (high-risk AI systems)
Category	Dataset Properties

ID	10.3 (O5)
Description	Data governance and management practices shall concern relevant data preparation steps.
Rationale	Art. 10 (2c); Data preparation is a crucial step before the data is used in the AI system and all relevant operations on the data need to conform with data governance and management guidelines.
Difficulty	medium
Fit Criterion	The operations performed on the data sets during data preparation are developed and overseen by a group of experts or with use of an appropriate standard (e.g., ISO/IEC JTC 1/SC 42).
Type	Non-Functional Requirement
Applicability	All (high-risk AI systems)
Category	Dataset Properties

ID	10.4 (O6)
Description	Data governance and management practices shall concern assumptions made about the given data sets.
Rationale	Art. 10 (2d); Data governance and management practices ensure that any assumptions made regarding data are consistent over different data sets and use cases.
Difficulty	low
Fit Criterion	Any assumptions made regarding data are performed and overseen by a group of experts. Assumptions are made within the boundaries of the information the given data is supposed to measure and represent.
Type	Process Requirement
Applicability	All (high-risk AI systems)
Category	Dataset Properties

ID	10.5 (O7)
Description	Data governance and management practices shall concern the assessment of quality, availability, and suitability of the required data sets.
Rationale	Art. 10 (2e); Data governance and management practices ensure that any assumptions made regarding data are consistent over different data sets and use cases.
Difficulty	medium
Fit Criterion	All assessments regarding data are performed and overseen by a group of experts or with the use of an appropriate standard (e.g., ISO/IEC JTC 1/SC 42).
Type	Non-Functional Requirement
Applicability	All (high-risk AI systems)
Category	Dataset Properties

ID	10.6 (O8)
Description	Data governance and management practices shall concern the examination of biases in the data sets.
Rationale	Art. 10 (2f); Biases in the used data sets results in biased output of the AI system which can lead to flawed output and potentially discrimination of its users.
Difficulty	high
Fit Criterion	A group of experts verifies that current data governance and management practices can identify biases, or it is identified with the use of an appropriate standard (e.g., ISO/IEC JTC 1/SC 42).
Type	Non-Functional Requirement
Applicability	All (high-risk AI systems)
Category	Dataset Properties

ID	10.7 (O9)
Description	Data governance and management practices shall identify and address gaps and shortcomings in the data.
Rationale	Art. 10 (2g); Errors in the data sets can reduce the quality of the data, lead to biases and result in a flawed output of the system.
Difficulty	high
Fit Criterion	A group of experts verifies if current data governance and management practices can identify and address gaps and shortcomings, or it is identified with the use of an appropriate standard (e.g., ISO/IEC JTC 1/SC 42).
Type	Non-Functional Requirement
Applicability	All (high-risk AI systems)
Category	Dataset Properties

ID	10.8 (O11)
Description	Training, validation, and testing data sets shall be relevant, representative, free of errors and complete.
Rationale	Art. 10 (3); These flaws in the data sets can lead to biases, sampling errors and finally, a flawed output of the system.
Difficulty	high
Fit Criterion	A group of experts determines the correctness of the data sets based on demographic data of the persons the AI systems is used on and based on statistical analysis of the data, or it is identified with the use of an appropriate standard (e.g., ISO/IEC JTC 1/SC 42).
Type	Non-Functional Requirement
Applicability	Restricted (high-risk AI systems that perform model training with data)
Category	Dataset Properties

ID	10.9 (O12)
Description	Training, validation, and testing data sets shall have the appropriate statistical properties as regards users/groups of users.
Rationale	Art. 10 (3); Flaws in the data sets may lead to a flawed output of the system.
Difficulty	high
Fit Criterion	A group of experts performs statistical analysis to confirm that the datasets fulfil the required statistical properties, or it is identified with the use of an appropriate standard (e.g., ISO/IEC JTC 1/SC 42). Properties need to be applicable to the given use case and are only regarding the people it is intended to be used on.
Type	Non-Functional Requirement
Applicability	Restricted (high-risk AI systems that perform model training with data)
Category	Dataset Properties

ID	10.10 (O13)
Description	Training, validation, and testing data sets shall contain characteristics specific to the geographical, behavioural, or functional setting.
Rationale	Art. 10 (4); Data sets that are not representative of the AI systems' training data and specifically the environment in which the system is used in, may lead to a flawed output of the system.
Difficulty	high
Fit Criterion	A group of experts that have knowledge about the given setting the AI system is intended to be used in verify that the data sets fulfil these characteristics.
Type	Non-Functional Requirement
Applicability	Restricted (high-risk AI systems that perform model training with data)
Category	Dataset Properties

APPENDIX I.3: TECHNICAL DOCUMENTATION (ART. 11)

ID	11.1 (O1)
Description	A technical documentation shall exist for/within the high-risk AI system.
Rationale	Art. 11 (1); Authorities must be able to assess the compliance of the system with the help of the technical documentation.
Difficulty	medium
Fit Criterion	A technical documentation was drafted for the system.
Type	Non-functional Requirement
Applicability	All (high-risk AI systems)
Category	Technical Documentation

ID	11.2 (O2)
Description	The technical documentation shall be kept up to date with respect to any change that is introduced to the system.
Rationale	Art. 11 (1); The documentation needs to include every change that was made to the system.
Difficulty	medium
Fit Criterion	The accessible technical documentation contains every recent change.
Type	Process Requirement
Applicability	All (high-risk AI systems)
Category	Technical Documentation

ID	11.3 (O3)
Description	The technical documentation shall contain a general description of the AI system including its intended purpose, the person/s developing the system, the date, and the version of the system.
Rationale	Annex IV (1a); The documentation needs to be able to provide any authority with the needed basic information and complies with the standard of technical documentations regarding high-risk AI systems.
Difficulty	medium
Fit Criterion	The required specification is complete and included in the technical documentation.
Type	Non-functional Requirement
Applicability	All (high-risk AI systems)
Category	Technical Documentation

ID	11.4 (O4)
Description	The technical documentation shall contain how the AI system interacts or can be used to interact with hardware or software that is not part of the AI system itself.
Rationale	Annex IV (1b); The documentation needs to be able to provide any authority with the needed basic information and complies with the standard of technical documentations regarding high-risk AI systems.
Difficulty	medium
Fit Criterion	The required specification is complete and included in the technical documentation.
Type	Non-functional Requirement
Applicability	All (high-risk AI systems)
Category	Technical Documentation

ID	11.5 (O5)
Description	The technical documentation shall contain the versions of relevant software or firmware and any requirement related to version update.
Rationale	Annex IV (1c); The documentation needs to be able to provide any authority with the needed basic information and complies with the standard of technical documentations regarding high-risk AI systems.
Difficulty	medium
Fit Criterion	The required specification is complete and included in the technical documentation.
Type	Non-functional Requirement
Applicability	All (high-risk AI systems)
Category	Technical Documentation

ID	11.6 (O6)
Description	The technical documentation shall contain the description of all forms in which the AI system is placed on the market or put into service.
Rationale	Annex IV (1d); The documentation needs to be able to provide any authority with the needed basic information and complies with the standard of technical documentations regarding high-risk AI systems.
Difficulty	medium
Fit Criterion	The required specification is complete and included in the technical documentation.
Type	Non-Functional Requirement
Applicability	All (high-risk AI systems)
Category	Technical Documentation

ID	11.7 (O7)
Description	The technical documentation shall contain the description of hardware on which the AI system is intended to run.
Rationale	Annex IV (1e); The documentation needs to be able to provide any authority with the needed basic information and complies with the standard of technical documentations regarding high-risk AI systems.
Difficulty	medium
Fit Criterion	The required specification is complete and included in the technical documentation.
Type	Non-Functional Requirement
Applicability	All (high-risk AI systems)
Category	Technical Documentation

ID	11.8 (O8)
Description	The technical documentation shall contain where the AI system is a component of products, photographs or illustrations showing external features, marking and internal layout of those products.
Rationale	Annex IV (1f); The documentation needs to be able to provide any authority with the needed basic information and complies with the standard of technical documentations regarding high-risk AI systems.
Difficulty	medium
Fit Criterion	The required specification is complete and included in the technical documentation.
Type	Non-Functional Requirement
Applicability	All (high-risk AI systems)
Category	Technical Documentation

ID	11.9 (O9)
Description	The technical documentation shall contain instructions of use for the user as defined in the requirements 13.2- 13.10 and installation instructions.
Rationale	Annex IV (1g); The documentation needs to be able to provide any authority with the needed basic information and complies with the standard of technical documentations regarding high-risk AI systems.
Difficulty	medium
Fit Criterion	A user that accesses the technical documentation accessibly finds the instructions.
Type	Non-Functional Requirement
Applicability	All (high-risk AI systems)
Category	Technical Documentation

ID	11.10 (O10)
Description	The technical documentation shall contain a detailed description of the system development process, which needs to include all methods and steps that were performed, and all used pre-trained systems or third-party tools and how they have been used, integrated, or modified.
Rationale	Annex IV (2a); The documentation needs to be able to provide any authority with detailed information and comply with the standard of technical documentations regarding high-risk AI systems.
Difficulty	medium
Fit Criterion	The required specification is complete and included in the technical documentation.
Type	Non-Functional Requirement
Applicability	All (high-risk AI systems)
Category	Technical Documentation

ID	11.11 (O11)
Description	The technical documentation shall contain a detailed description about the design specifications of the system, namely the general logic of the AI system and of the algorithms; the key design choices including the rationale and assumptions made, also in terms of the people the system will be used on; the main classification choices; what the system is designed to optimize for and the relevance of the different parameters; decisions about any possible trade-off made to comply with other the other requirements in this Requirements Specification.
Rationale	Annex IV (2b); The documentation needs to be able to provide any authority with detailed information and comply with the standard of technical documentations regarding high-risk AI systems.
Difficulty	medium
Fit Criterion	The required specification is complete and included in the technical documentation.
Type	Non-Functional Requirement
Applicability	All (high-risk AI systems)
Category	Technical Documentation

ID	11.12 (O12)
Description	The technical documentation shall contain a detailed description of the systems architecture, explaining how software components build on or feed into each other and integrate into the overall processing and the computational resources used to develop, train, test and validate the AI system.
Rationale	Annex IV (2c); The documentation needs to be able to provide any authority with detailed information and comply with the standard of technical documentations regarding high-risk AI systems.
Difficulty	medium
Fit Criterion	The required specification is complete and included in the technical documentation.
Type	Non-Functional Requirement
Applicability	All (high-risk AI systems)
Category	Technical Documentation

ID	11.13 (O13)
Description	The technical documentation shall contain a detailed description about the data requirements in terms of datasheets describing the training methodologies and techniques and the training data sets used, including information about the provenance of those data sets, their scope, and main characteristics; how the data was obtained and selected; labelling procedures (e.g., for supervised learning), data cleaning methodologies (e.g., outlier detection).
Rationale	Annex IV (2d); The documentation needs to be able to provide any authority with detailed information and comply with the standard of technical documentations regarding high-risk AI systems.
Difficulty	medium
Fit Criterion	The required specification is complete and included in the technical documentation.
Type	Non-Functional Requirement
Applicability	All (high-risk AI systems)
Category	Technical Documentation

ID	11.14 (O14)
Description	The technical documentation shall contain a detailed description about the assessment of the demanded human oversight measures and the necessary technical measures to facilitate the interpretation of the outputs of AI systems by the users.
Rationale	Annex IV (2e); The documentation needs to be able to provide any authority with detailed information and comply with the standard of technical documentations regarding high-risk AI systems.
Difficulty	medium
Fit Criterion	The required specification is complete and included in the technical documentation.
Type	Non-Functional Requirement
Applicability	All (high-risk AI systems)
Category	Technical Documentation

ID	11.15 (O15)
Description	The technical documentation shall contain a detailed description of pre-determined changes to the AI system and its performance, together with all the relevant information related to the technical solutions adopted to ensure continuous compliance of the AI system with the relevant requirements in this Requirements Specification.
Rationale	Annex IV (2f); The documentation needs to be able to provide any authority with detailed information and comply with the standard of technical documentations regarding high-risk AI systems.
Difficulty	medium
Fit Criterion	The required specification is complete and included in the technical documentation.
Type	Non-Functional Requirement
Applicability	All (high-risk AI systems)
Category	Technical Documentation

ID	11.16 (O16)
Description	The technical documentation shall contain the validation and testing procedures used in the development of the system, including information about the validation and testing data used and their main characteristics. This also includes metrics used to measure accuracy, robustness, cybersecurity, and compliance as well as potentially discriminatory impacts. In addition to this, Test logs and all test reports dated and signed by the persons responsible.
Rationale	Annex IV (2g); The documentation needs to be able to provide any authority with detailed information and comply with the standard of technical documentations regarding high-risk AI systems.
Difficulty	medium
Fit Criterion	The required specification is complete and included in the technical documentation.
Type	Non-functional Requirement
Applicability	All (high-risk AI systems)
Category	Technical Documentation

ID	11.17 (O17)
Description	The technical documentation shall contain detailed information about the monitoring, functioning and control of the High-risk AI system, in particular with regard to: its capabilities and limitations in performance, including the degrees of accuracy for specific persons or groups of persons on which the system is intended to be used and the overall expected level of accuracy in relation to its intended purpose, as well as the foreseeable unintended outcomes and sources of risks to health and safety, fundamental rights and discrimination in view of the intended purpose of the AI system; specifications on input data.
Rationale	Annex IV (3); The documentation needs to be able to provide any authority with detailed information and comply with the standard of technical documentations regarding high-risk AI systems.
Difficulty	medium
Fit Criterion	The required specification is complete and included in the technical documentation.
Type	Non-functional Requirement
Applicability	All (high-risk AI systems)
Category	Technical Documentation

ID	11.18 (O18)
Description	The technical documentation shall contain a detailed description of the risk management system in accordance with the requirements 9.1 - 9.14.
Rationale	Annex IV (4); The documentation needs to be able to provide any authority with detailed information and comply with the standard of technical documentations regarding high-risk AI systems.
Difficulty	medium
Fit Criterion	The required specification is complete and included in the technical documentation.
Type	Non-Functional Requirement
Applicability	All (high-risk AI systems)
Category	Technical Documentation

ID	11.19 (O19)
Description	The technical documentation shall contain a description of any change made to the system through its lifecycle.
Rationale	Annex IV (5); The documentation needs to be able to provide any authority with needed basic information and comply with the standard of technical documentations regarding high-risk AI systems.
Difficulty	medium
Fit Criterion	The required specification is complete and included in the technical documentation.
Type	Non-Functional Requirement
Applicability	All (high-risk AI systems)
Category	Technical Documentation

ID	11.20 (O20)
Description	The technical documentation shall contain a list of the harmonized standards applied in full or in part the references of which have been published in the Official Journal of the European Union; where no such harmonized standards have been applied, a detailed description of the solutions adopted to meet the requirements, including a list of other relevant standards and technical specifications applied.
Rationale	Annex IV (6); The documentation needs to be able to provide any authority with detailed information and comply with the standard of technical documentations regarding high-risk AI systems.
Difficulty	medium
Fit Criterion	The required specification is complete and included in the technical documentation.
Type	Non-functional Requirement
Applicability	All (high-risk AI systems)
Category	Technical Documentation

ID	11.21 (O21)
Description	The technical documentation shall contain a copy of the EU declaration of conformity.
Rationale	Annex IV (7); The documentation needs to be able to provide any authority with needed basic information and comply with the standard of technical documentations regarding high-risk AI systems.
Difficulty	low
Fit Criterion	The required specification is complete and included in the technical documentation.
Type	Non-Functional Requirement
Applicability	All (high-risk AI systems)
Category	Technical Documentation

ID	11.22 (O22)
Description	The technical documentation shall contain a detailed description of the system in place to evaluate the AI system performance in the post-market monitoring system.
Rationale	Annex IV (8); The documentation needs to be able to provide any authority with detailed information and comply with the standard of technical documentations regarding high-risk AI systems.
Difficulty	medium
Fit Criterion	The required specification is complete and included in the technical documentation.
Type	Non-Functional Requirement
Applicability	All (high-risk AI systems)
Category	Technical Documentation

ID	11.23 (O3)
Description	The technical documentation of the high-risk AI system is combined with all the other information that is legally required to form one single technical documentation. A high-risk AI system that enters the market and is related to a product, to which the legal acts listed in Annex II, section A apply, only one single technical documentation is needed.
Rationale	Art. 11 (2); By abiding by this requirement, redundancies are avoided.
Difficulty	medium
Fit Criterion	The technical documentation of the high-risk AI system is combined with the one of the related products and can be accessed in one document.
Type	Process Requirement
Applicability	Restricted (high-risk AI systems related to a product for which the EU has already set harmonized standards.)
Category	Technical Documentation

APPENDIX I.4: RECORD KEEPING (ART. 12)

ID	12.1(O1)
Description	The high-risk AI system shall possess automatic event-recording capabilities.
Rationale	Art. 12 (1); The performance of a high-risk AI system must be reviewable in order to be trusted
Difficulty	medium
Fit Criterion	For any point in time during the operation of the high-risk AI system, its records may be accessed by a user
Type	Functional Requirement
Applicability	All (high-risk AI systems)
Category	Record Keeping

ID	12.2 (O2)
Description	All events during the AI system's entire lifecycle operation shall be recorded in a way that ensures traceability with respect to the intended purpose of the system.
Rationale	Art. 12 (2); Depending on the type of AI system, the events governing its decisions and outputs must be reviewable and understandable by an independent party.
Difficulty	low
Fit Criterion	Reviewing all steps and events of a system's operation period in the past allows a third party not present during operation to understand the system's behaviour and decisions during that period.
Type	Functional Requirement
Applicability	All (high-risk AI systems), given req. 12.1 is fulfilled
Category	Record Keeping

ID	12.3 (O1)
Description	The event-recording capability shall create and maintain its records according to an industry-acknowledged standard or common practice.
Rationale	Art. 12 (1); The records need to be interchangeable.
Difficulty	low
Fit Criterion	The records returned by the system fulfil the standard as determined by an expert user.
Type	Functional Requirement
Applicability	All (high-risk AI systems), given req. 12.1 is fulfilled
Category	Record Keeping

ID	12.4 (O4a)
Description	Depending on the type of AI system, the data provided by users or through other sources during operation shall be automatically, and systematically collected and documented such that they can be assessed against the present Requirements Specification.
Rationale	Art. 12 (3) + Art 61 (2); Depending on the type of AI system, the events governing its decisions and outputs must be reviewable and understandable by an independent party that may verify its compliance with applicable regulations.
Difficulty	medium
Fit Criterion	After a period of operation of the system, the automatically recorded, structured respective data provided in that period may be accessed by a competent user through an interface.
Type	Functional Requirement
Applicability	All (high-risk AI systems), given req. 12.1 is fulfilled
Category	Record Keeping

ID	12.5 (O4a)
Description	The data provided by users or through other sources during operation shall be automatically, and systematically analysed.
Rationale	Art. 12 (3) + Art. 61 (2); Depending on the type of AI system, the events governing its decisions and outputs must be automatically reviewed to highlight potential risks and weaknesses.
Difficulty	Medium to high
Fit Criterion	After a period of operation of the system, the automatically created, structured analysis of the respective data provided in that period may be accessed by a competent user through an interface.
Type	Functional Requirement
Applicability	All (high-risk AI systems), given req. 12.1 is fulfilled
Category	Record Keeping

ID	12.6 (O4b)
Description	A post-market monitoring plan shall be established that governs the specifics of 12.4 and 12.5.
Rationale	Art. 12 (3) + Art. 61(3); The monitoring procedure needs to be documented and reviewable to be deemed appropriate and compliant
Difficulty	low
Fit Criterion	A monitoring plan according to the template by the European Commission is included in the technical documentation of the system and adhered to.
Type	Process Requirement
Applicability	All (high-risk AI systems), given req. 12.1, 12.4, 12.5 are fulfilled
Category	Record Keeping

ID	12.7 (O3)
Description	The records of the logging capability are appropriate to monitor situations where the system a) may impose a risk to health or safety or the fundamental rights of persons or b) may lead to a substantial modification of itself.
Rationale	Art. 12 (3); Detailed review of operation periods of an AI system that are critical in the sense of previous legislation must be possible.
Difficulty	low
Fit Criterion	After occurrence of a relevant situation, the detailed records may be examined by a user through an interface
Type	Functional Requirement
Applicability	All (high-risk AI systems)
Category	Record Keeping

ID	12.8 (O5)
Description	The logging records shall include the period of each use.
Rationale	Art. 12 (4)(a); Detailed review of operation of a high-risk AI system dealing with biomedical data of human beings is critical
Difficulty	low
Fit Criterion	The periods of all past usages of the system may be examined by a user through an interface.
Type	Functional Requirement
Applicability	Restricted (high-risk AI systems intended to be used for the ‘real-time’ and ‘post’ remote biometric identification of natural persons)
Category	Record Keeping

ID	12.9 (O6)
Description	The logging records shall include the database against which the input to the model is assessed.
Rationale	Art. 12 (4)(b); Detailed review of operation of a high-risk AI system dealing with biomedical data of human beings is critical
Difficulty	low
Fit Criterion	The of the reference databases in all past usages of the system may be examined by a user through an interface.
Type	Functional Requirement
Applicability	Restricted (high-risk AI systems intended to be used for the ‘real-time’ and ‘post’ remote biometric identification of natural persons)
Category	Record Keeping

ID	12.10 (O7)
Description	The logging records shall include the input data for which the model found determined a search match.
Rationale	Art. 12 (4)(c); Detailed review of operation of a high-risk AI system dealing with biomedical data of human beings is critical
Difficulty	low
Fit Criterion	The input data points in all past biometrical identification processes carried out in the system may be examined by a user through an interface.
Type	Functional Requirement
Applicability	Restricted (high-risk AI systems intended to be used for the ‘real-time’ and ‘post’ remote biometric identification of natural persons)
Category	Record Keeping

ID	12.11 (O8)
Description	The logging records shall include the identification of the human overseer accountable according to requirements 14.5 to 14.9 during the operation of the system shall be recorded.
Rationale	Art. 12 (4)(d); Detailed review of operation of a high-risk AI system dealing with biomedical data of human beings is critical
Difficulty	low
Fit Criterion	The identification of the human overseer in all past usages of the system may be examined by a user through an interface.
Type	Functional Requirement
Applicability	Restricted (high-risk AI systems intended to be used for the 'real-time' and 'post' remote biometric identification of natural persons)
Category	Record Keeping

APPENDIX I.5: TRANSPARENCY AND PROVISION OF INFORMATION TO USERS (ART. 13)

ID	13.1 (O1)
Description	The operations executed by the AI system shall be sufficiently transparent for users to be able to interpret and appropriately use the system output.
Rationale	Art. 13 (1); The users must be able to work productively with the system outputs, and for this it is essential that they are able to trace the creation of these outputs.
Difficulty	medium
Fit Criterion	The system operations are transparent to a degree that the user can comprehend the system output.
Type	Functional Requirement
Applicability	All (high-risk AI systems)
Category	Explainability

ID	13.2 (O2)
Description	The High-Risk AI System shall be accompanied by instructions for use, in an appropriate digital format or otherwise that include concise, complete, correct, and clear information.
Rationale	Art. 13 (2); Users need relevant, accessible, and comprehensible instructions when interacting with the system.
Difficulty	medium
Fit Criterion	Every kind of instruction that is required follows this quality standard. Once the system is available to the market, every user can access a guide of instructions in which he does not miss any information he deems relevant and in which nothing is contained he deems superfluous.
Type	Process Requirement
Applicability	All (high-risk AI systems)
Category	Explainability

ID	13.3 (O3a)
Description	There shall be instructions about the identity and the contact details of the provider and its authorised representative, given there is one.
Rationale	Art. 13 (3) (a); The user should be given the opportunity to reach out to a contact person, whether for technical or legal questions.
Difficulty	low
Fit Criterion	The required instructions can be obtained and meet the quality standard for instructions from requirement 13.2.
Type	Process Requirement
Applicability	All (high-risk AI systems)
Category	Explainability

ID	13.4 (O3b, O3d)
Description	There shall be instructions about the intended purpose of the high-risk AI System and any known or foreseeable circumstances which may lead to risks to health and safety or fundamental rights when the system is used as intended or misused.
Rationale	Art. 13 (3) (b) (i) & (iii); The user should know about the scope and non-scope of the system and be informed about possible hazardous situations.
Difficulty	medium
Fit Criterion	The required instructions can be obtained and meet the quality standard for instructions from requirement 13.2.
Type	Process Requirement
Applicability	All (high-risk AI systems)
Category	Explainability

ID	13.5 (O3c, O3d)
Description	There shall be instructions about the tested and validated level of accuracy, robustness, and cybersecurity and any known or foreseeable circumstances which could impact these levels.
Rationale	Art. 13 (3) (b) (ii); The levels are intended to show the user how susceptible the system could be to errors.
Difficulty	medium
Fit Criterion	The required instructions can be obtained and meet the quality standard for instructions from requirement 13.2.
Type	Process Requirement
Applicability	All (high-risk AI systems)
Category	Explainability

ID	13.6 (O3e)
Description	There shall be instructions about the performance of the High-Risk AI system as regards its intended use cases.
Rationale	Art. 13 (3) (b) (iv); Users are informed of the default system performance for intended use.
Difficulty	low
Fit Criterion	The required instructions can be obtained and meet the quality standard for instructions from requirement 13.2.
Type	Process Requirement
Applicability	All (high-risk AI systems)
Category	Explainability

ID	13.7 (O3f)
Description	There shall be instructions about specifications for the input data, or any other relevant information in terms of the training, validation and testing data sets used.
Rationale	Art. 13 (3) (b) (v); Transparency about which data sets are processed for which purpose.
Difficulty	low
Fit Criterion	The required instructions can be obtained and meet the quality standard for instructions from requirement 13.2.
Type	Process Requirement
Applicability	Restricted (high-risk AI systems using input data or data sets when operating according to their intended use)
Category	Explainability

ID	13.8 (O3g)
Description	There shall be instructions about changes to the High-Risk AI system and its performance that were made after the initial conformity assessment.
Rationale	Art. 13 (3) (c); Timeliness and completeness of the other instruction requirements.
Difficulty	low
Fit Criterion	The required instructions can be obtained and meet the quality standard for instructions from requirement 13.2.
Type	Process Requirement
Applicability	All (high-risk AI systems)
Category	Explainability

ID	13.9 (O3h)
Description	There shall be instructions about the human oversight measures, including the applied technical measures.
Rationale	Art. 13 (3) (d); In article 14, human oversight measures are introduced, as necessary. By communicating the taken measures to the user, he may be able to better understand the system output.
Difficulty	medium
Fit Criterion	The required instructions can be obtained and meet the quality standard for instructions from requirement 13.2.
Type	Process Requirement
Applicability	All (high-risk AI systems)
Category	Explainability

ID	13.10 (O3i)
Description	There shall be instructions about the expected lifetime and any measures to ensure proper functioning.
Rationale	Art. 13 (3) (e); The user should be shown that appropriate steps are being taken to maintain the system until the end of its life cycle.
Difficulty	low
Fit Criterion	The required instructions can be obtained and meet the quality standard for instructions from requirement 13.2.
Type	Process Requirement
Applicability	All (high-risk AI systems)
Category	Explainability

APPENDIX I.6: HUMAN OVERSIGHT (ART. 14)

ID	14.1 (O1, O5)
Description	High-risk AI systems shall operate such that they can be effectively overseen by a natural person.
Rationale	Art. 14 (1); Accountability and sensitivity of the context requires the possibility of human intervention.
Difficulty	medium
Fit Criterion	Human oversight is ensured and included by design in the high-risk AI system and a human-machine interface tool can be used, the level of implementation is confirmed by an expert group.
Type	Non-Functional Requirement
Applicability	All (high-risk AI systems)
Category	Human Oversight

ID	14.2 (O2)
Description	The high-risk AI system shall integrate human oversight with the aim of preventing or minimising the risks to health, safety or fundamental rights caused by the active high-risk AI system, within its boundaries of intended purpose and under conditions of foreseeable misuse.
Rationale	Art. 14(2); Protection of human health from potential harm caused by AI systems.
Difficulty	high
Fit Criterion	The individual responsible for human oversight is able to prevent or mitigate foreseeable misuse and risk of high-risk AI systems within the scope of its intended purpose, with respect to its consequence on preservation of health, safety, or fundamental rights.
Type	Process Requirement
Applicability	All (high-risk AI systems)
Category	Human Oversight

ID	14.3 (O3)
Description	High-risk AI systems shall integrate human oversight interfaces before they are placed on or used in the market.
Rationale	Art. 14 (3a, 3b); Interfaces provide easy access to non-technical experts and allow more direct control over the AI systems with respect to interpretation and stopping mechanisms.
Difficulty	medium
Fit Criterion	Human oversight is implemented in high-risk AI systems at the point when it is ready to enter the market or put into productive service or are accompanied and outlined by the provider via instructions, so that users must implement and perform the oversight themselves.
Type	Process Requirement
Applicability	All (high-risk AI systems)
Category	Human Oversight

ID	14.4 (O4)
Description	High-risk AI systems shall operate such that the limitations and capacities of the system are clearly outlined and understood by the individuals responsible for human oversight, deviations must be detected, investigated, and properly addressed.
Rationale	Art. 14 (4a); The user must know in which scenarios, with what data and how to use the high AI system, so that misuse can be prevented.
Difficulty	high
Fit Criterion	The individual responsible for human oversight confirms their understanding of the limitations and capacities of the high-risk AI system and their ability to respond to anomalies, dysfunctions, and unexpected performance.
Type	Process Requirement
Applicability	All (high-risk AI systems)
Category	Human Oversight

ID	14.5 (O6)
Description	The high-risk AI systems shall operate such that the individuals responsible for human oversight are not over-relying on the system (automation bias) with respect to predictions and other decisions made by the system.
Rationale	Art. 14 (4b); Overreliance and consequent inattention regarding the produced output of the AI system may lead to wrong decisions as the output of the system can be flawed.
Difficulty	high
Fit Criterion	A group of experts determines that the functions and the mode of operation of the high-risk AI system sufficiently prevents its users from over-relying on its output, for instance through provision of information and warning.
Type	Non-Functional Requirement
Applicability	All (high-risk AI systems)
Category	Human Oversight

ID	14.6 (O7)
Description	The high-risk AI system shall operate such that the produced outputs and the systems' logic are transparent and can be interpreted via tools and methods by the individuals responsible for human oversight.
Rationale	Art. 14 (4c); The supervising user needs to understand how the inputs map to the outputs to prevent using a "black-box" system.
Difficulty	high
Fit Criterion	A group of experts of experts determines that the characteristics of the system are transparent, and the corresponding interpretation tools and methods are understood by the individuals responsible for human oversight.
Type	Non-Functional Requirement
Applicability	All (high-risk AI systems)
Category	Human Oversight

ID	14.7 (O8)
Description	The decisions of the high-risk AI system shall be such that they can be disregarded, overwritten, and reversed in any situation by the individuals responsible for human oversight.
Rationale	Art. 14(4d); The possibility of intervention must be guaranteed due to potential erroneous decisions arising from the results produced by the AI system.
Difficulty	low
Fit Criterion	A group of experts determines that the implementation is satisfactory regarding the ability to disregard, overwrite and reverse the decision of the high-risk AI system.
Type	Functional Requirement
Applicability	All (high-risk AI systems)
Category	Human Oversight

ID	14.8 (O9)
Description	The high-risk AI system shall operate such that the individuals responsible for human oversight can at any point interrupt or halt the program with a single procedure.
Rationale	Art. 14 (4e); The possibility of intervention must be guaranteed due to potential erroneous decisions arising from the results produced by the AI system and concomitant harm that could be caused.
Difficulty	low
Fit Criterion	A group of experts determines that the implementation is satisfactory regarding the ability to immediately stop the high-risk AI system.
Type	Functional Requirement
Applicability	All (high-risk AI systems)
Category	Human Oversight

ID	14.9 (O10)
Description	The high-risk AI system shall operate such that its decisions with respect to identification, assignment, and assessment of natural persons in educational and vocational training institutions, are confirmed by at least two natural persons.
Rationale	Art. 14 (5); Over-reliance on decisions made by AI systems in critical environments must be confirmed by natural persons, to mitigate bias and ensure an equal and fair treatment of natural persons.
Difficulty	low
Fit Criterion	At least two natural persons confirm the decisions of high-risk AI systems in the context of educational and vocational training institutions, this includes determining access of natural persons to educational and vocational training institutions or assigning natural persons thereto, assessing students in test and assessing participants in test commonly required for admission to educational institutions.
Type	Functional Requirement
Applicability	Restricted (AI systems intended to be used for the 'real-time' and 'post' remote biometric identification of natural persons.)
Category	Human Oversight

APPENDIX I.7: ACCURACY, ROBUSTNESS AND CYBERSECURITY (ART. 15)

ID	15.1
Description	Appropriate levels and metrics for the high-risk AI system's accuracy, robustness and cybersecurity shall be defined, tested, and validated.
Rationale	Art. 13 (3); Art. 15 (1); Users and maintainers need individually defined levels to verify appropriate accuracy, robustness, and cybersecurity
Difficulty	high
Fit Criterion	Clear and appropriate levels and metrics regarding the system's accuracy, robustness and cybersecurity are defined, tested, and validated based on the individual context or a commonly recognized standard.
Type	Process Requirement
Applicability	All (high-risk AI systems)
Category	Accuracy, Robustness, Cybersecurity

ID	15.2 (O1, O2)
Description	The high-risk AI system shall operate with the defined (see 15.2), consistent level of accuracy throughout its lifecycle, appropriate to its intended purpose.
Rationale	Art. 15 (1); A low or inconsistent level of accuracy poses a risk to the quality of the systems output.
Difficulty	high
Fit Criterion	The level complies with the defined specifications deemed appropriate by a subject matter expert.
Type	Non-Functional Requirement
Applicability	All (high-risk AI systems)
Category	Accuracy, Robustness, Cybersecurity

ID	15.3 (O1, O2)
Description	The high-risk AI system shall operate with the defined (see 15.2), consistent level of robustness throughout its lifecycle, appropriate to its intended purpose.
Rationale	Art. 15 (1); A low or inconsistent level of robustness poses a risk to the performance and reliability of the system.
Difficulty	high
Fit Criterion	The level complies with the defined specifications deemed appropriate by a subject matter expert.
Type	Non-Functional Requirement
Applicability	All (high-risk AI systems)
Category	Accuracy, Robustness, Cybersecurity

ID	15.4 (O1, O2)
Description	The high-risk AI system shall operate with the defined (see 15.2), consistent level of cybersecurity throughout its lifecycle, appropriate to its intended purpose.
Rationale	Art. 15 (1); A low or inconsistent level of cybersecurity poses a risk to the integrity and safety of the system.
Difficulty	high
Fit Criterion	The level complies with the defined specifications deemed appropriate by a subject matter expert.
Type	Non-Functional Requirement
Applicability	All (high-risk AI systems)
Category	Accuracy, Robustness, Cybersecurity

ID	15.5 (O3)
Description	The levels of accuracy and the respective metrics shall be declared in the accompanying instructions of use.
Rationale	Art. 15 (2); As regards quality assurance and control, users need to understand what levels of accuracy are considered acceptable.
Difficulty	low
Fit Criterion	Test-users confirm to understand all relevant metrics and levels of accuracy by consulting the instructions of use.
Type	Process Requirement
Applicability	All (high-risk AI systems)
Category	Accuracy, Robustness, Cybersecurity

ID	15.6 (O4)
Description	The high-risk AI system shall identify and mitigate or prevent errors, faults or inconsistencies within the system or its operational environment. This may be achieved through technical redundancy solutions.
Rationale	Art. 15 (3); Errors, faults or inconsistencies can pose a threat in particular towards interacting natural persons or other systems.
Difficulty	high
Fit Criterion	Testing metrics prove a high resiliency towards errors, faults, or inconsistencies.
Type	Non-Functional Requirement
Applicability	All (high-risk AI systems)
Category	Accuracy, Robustness, Cybersecurity

ID	15.7 (O5)
Description	The high-risk AI system shall duly address possibly biased outputs through ‘feedback loops’ with appropriate mitigation measures.
Rationale	Art. 15 (3); The quality and functioning of a system can be compromised by feedback loops creating biased outputs.
Difficulty	medium
Fit Criterion	Testing metrics prove a low susceptibility to biased outputs and feedback loops.
Type	Non-functional Requirement
Applicability	Restricted (high-risk AI systems that continue to learn in production)
Category	Accuracy, Robustness, Cybersecurity

ID	15.8 (O6)
Description	The high-risk AI system shall identify and mitigate or prevent attempts by unauthorised third parties to alter their use or performance.
Rationale	Art. 15 (4); Malevolent third parties can cause great damage by manipulating AI systems.
Difficulty	medium
Fit Criterion	A sufficient level of attacks performed for testing purposes is successfully identified and mitigated or prevented by the system.
Type	Non-Functional Requirement
Applicability	All (high-risk AI systems)
Category	Accuracy, Robustness, Cybersecurity

ID	15.9 (O7)
Description	The technical solutions aimed at ensuring the cybersecurity of high-risk AI systems shall be appropriate to the relevant circumstances and risks.
Rationale	Art. 15 (4); With respect to the cost of risk, the measures must be chosen appropriately.
Difficulty	medium
Fit Criterion	The measures comply with the risk metrics defined by the risk management system.
Type	Process Requirement
Applicability	All (high-risk AI systems)
Category	Accuracy, Robustness, Cybersecurity

ID	15.10 (O)
Description	The technical solutions addressing AI specific vulnerabilities shall include measures to prevent and control for attacks trying to manipulate the training dataset (‘data poisoning’).
Rationale	Art. 15 (4); Manipulation of training datasets is a common way to interfere with an AI system.
Difficulty	medium
Fit Criterion	A sufficient level of ‘data poisoning’ attacks performed for testing purposes is successfully identified and mitigated or prevented by the system.
Type	Non-Functional Requirement
Applicability	Restricted (high-risk AI systems exposed to AI specific vulnerabilities)
Category	Accuracy, Robustness, Cybersecurity

ID	15.11 (O9)
Description	The technical solutions addressing AI specific vulnerabilities shall include measures to prevent and control for inputs designed to cause the model to make a mistake (‘adversarial examples’).
Rationale	Art. 15 (4); Malicious inputs are a common way to interfere with an AI system.
Difficulty	medium
Fit Criterion	A sufficient level of ‘adversarial example’ attacks performed for testing purposes is successfully identified and mitigated or prevented by the system.
Type	Non-Functional Requirement
Applicability	Restricted (high-risk AI systems exposed to AI specific vulnerabilities)
Category	Accuracy, Robustness, Cybersecurity

ID	15.12 (O10)
Description	The technical solutions addressing AI specific vulnerabilities shall include measures to prevent and control for model flaws.
Rationale	Art. 15 (4); Unidentified and uncontrolled model flaws can significantly compromise the systems quality.
Difficulty	medium
Fit Criterion	The high-risk AI system is continuously controlled for model flaws and shows a low rate of such.
Type	Non-Functional Requirement
Applicability	Restricted (high-risk AI systems exposed to AI specific vulnerabilities)
Category	Accuracy, Robustness, Cybersecurity

Appendix II.1: Identified Software Solutions - Testing

Framework	Publisher	Category	Original Paper	Scholar Results	Alternative Source
Azure Machine Learning	Microsoft	Testing	N/A	3.110	https://docs.microsoft.com/en-us/azure/architecture/data-science-process/deploy-models-in-production#ab-testing df
Amazon SageMaker	Amazon	Testing	N/A	1.210	https://docs.aws.amazon.com/sagemaker/latest/dg/model-ab-testing.html https://aws.amazon.com/de/blogs/machine-learning/a-b-testing-ml-models-in-production-using-amazon-sagemaker/
IBM Watson OpenScale	IBM	Testing	N/A	36	https://dataplatform.cloud.ibm.com/docs/content/wsj/model/getting-started.html

Appendix II.2: Identified Software Solutions - Dataset Properties

Framework	Publisher	Category	Original Paper	Scholar Results	Alternative Source
IBM SPSS Modeler	IBM	Dataset Properties	N/A	4.840	https://www.ibm.com/products/spss-modeler
SAP Data Services	SAP	Dataset Properties	N/A	142	https://www.sap.com/products/data-services.html https://www.altexsoft.com/blog/data-quality-management-and-tools/
Informatica Data Quality	Informatica	Dataset Properties	N/A	93	https://www.informatica.com/products/data-quality/informatica-data-quality.html https://www.altexsoft.com/blog/data-quality-management-and-tools/
SAS Data Quality	SAS	Dataset Properties	N/A	82	https://www.sas.com/en_us/software/data-quality.html https://www.researchgate.net/publication/220102796_A_Survey_of_Data_Quality_Tools
TensorFlow Data Validation (TFDV)	Google	Dataset Properties	E. Caveness, P. S. GC, Z. Peng, N. Polyzotis, S. Roy, and M. Zinkevich,"Tensorflow data validation: Data analysis and validation in continuousml pipelines," inProceedings of the 2020 ACM SIGMOD InternationalConference on Management of Data, 2020, pp. 2793–2796.	39	https://www.tensorflow.org/tfx/guide/tfdv
IBM Infosphere Information Server for Data Quality	IBM	Dataset Properties	N/A	26	https://www.ibm.com/products/infosphere-info-server-for-datamgmt https://www.altexsoft.com/blog/data-quality-management-and-tools/
IBM Watson Studio AutoAI	IBM	Dataset Properties	D. Wang, P. Ram, D. K. I. Weidele, S. Liu, M. Muller, J. D. Weisz,A. Valente, A. Chaudhary, D. Torres, H. Samulowitzet al., "Autoai:Automating the end-to-end ai lifecycle with humans-in-the-loop," inProceedings of the 25th International Conference on Intelligent UserInterfaces Companion, 2020, pp. 77–78.	13	https://www.ibm.com/cloud/watson-studio/autoai
Trillium Quality	Precisely	Dataset Properties	N/A	6	https://www.precisely.com/product/precisely-trillium/trillium-quality https://www.researchgate.net/publication/220102796_A_Survey_of_Data_Quality_Tools
Amazon SageMaker Data Wrangler	Amazon	Dataset Properties	N/A	3	https://aws.amazon.com/sagemaker/data-wrangler/
Amazon Web Services Glue DataBrew	Amazon	Dataset Properties	N/A	1	https://aws.amazon.com/glue/features/databrew/
IBM InfoSphere Advanced Data Preparation	IBM	Dataset Properties	N/A	0	https://www.ibm.com/de-de/products/infosphere-advanced-data-preparation

Appendix II.3: Identified Software Solutions - Record Keeping

Framework	Publisher	Category	Original Paper	Scholar Results	Alternative Source
TensorBoard	Google	Record Keeping	N/A	3.800	https://www.tensorflow.org/tensorboard https://colab.research.google.com/github/tensorflow/tensorboard/blob/master/docs/scalars_and_keras.ipynb
Amazon CloudWatch	Amazon	Record Keeping	N/A	1.670	https://geekflare.com/ai-frameworks/
Datadog	Datadog	Record Keeping	N/A	658	https://www.datadoghq.com/
SolarWinds Loggly	Loggly	Record Keeping	N/A	14	https://sematext.com/blog/log-analysis-tools/
Amazon SageMaker Model Monitor	Amazon	Record Keeping	N/A	13	https://docs.aws.amazon.com/sagemaker/latest/dg/model-monitor.html
Sematext Logs	Sematext	Record Keeping	N/A	1	https://sematext.com/blog/log-analysis-tools/
Neptune.ai	Neptune	Record Keeping	N/A	0	https://neptune.ai/product

Appendix II.4: Identified Software Solutions - Explainability

Framework	Publisher	Category	Original Paper	Scholar Results	Alternative Source
SHapley Additive exPlanations (SHAP)	Lundberg, S. and Lee, S.	Explainability	S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting modelpredictions," inProceedings of the 31st international conference onneural information processing systems, 2017, pp. 4768–4777.	9.080	N/A
Local Interpretable Model-Agnostic Explanation (LIME)	Ribero, M. and Singh, S. and Guestrin C.	Explainability	M. T. Ribeiro, S. Singh, and C. Guestrin, "" why should i trust you?"explaining the predictions of any classifier," inProceedings of the 22ndACM SIGKDD international conference on knowledge discovery anddata mining, 2016, pp. 1135–1144.	1.780	N/A
AIX360 Toolkit	IBM	Explainability	V. Arya, R. K. Bellamy, P.-Y. Chen, A. Dhurandhar, M. Hind, S. C.Hoffman, S. Houde, Q. V. Liao, R. Luss, A. Mojsilović et al., "Oneexplanation does not fit all: A toolkit and taxonomy of ai explainabilitytechniques,"arXiv preprint arXiv:1909.03012, 2019.	85	N/A
RuleX AI	Rulex	Explainability	N/A	18	https://www.rulex.ai/
IBM Research Uncertainty Quantification 360	IBM	Explainability	S. Ghosh, Q. V. Liao, K. N. Ramamurthy, J. Navratil, P. Sattigeri, K. R.Varshney, and Y. Zhang, "Uncertainty quantification 360: A holistictoolkit for quantifying and communicating the uncertainty of ai,"arXivpreprint arXiv:2106.01410, 2021.	3	http://uq360.mybluemix.net/

Appendix II.5: Identified Software Solutions - Human Oversight

Framework	Publisher	Category	Original Paper	Scholar Results	Alternative Source
SHapley Additive exPlanations (SHAP)	Lundberg, S. and Lee, S.	Human Oversight	S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting modelpredictions," inProceedings of the 31st international conference onneural information processing systems, 2017, pp. 4768–4777.	9.080	N/A
Local Interpretable Model-Agnostic Explanation (LIME)	Ribero, M. and Singh, S. and Guestrin C.	Human Oversight	M. T. Ribeiro, S. Singh, and C. Guestrin, "" why should i trust you?"explaining the predictions of any classifier," inProceedings of the 22ndACM SIGKDD international conference on knowledge discovery anddata mining, 2016, pp. 1135–1144.	1.780	N/A
MLflow	MLflow	Human Oversight	N/A	409	https://github.com/mlflow/mlflow
Kubeflow	Kubeflow	Human Oversight	N/A	250	https://www.kubeflow.org/docs/about/kubeflow/
IBM Research AI Explainability 360	IBM	Human Oversight	Arya, R. K. Bellamy, P.-Y. Chen, A. Dhurandhar, M. Hind, S. C.Hoffman, S. Houde, Q. V. Liao, R. Luss, A. Mojsilovicet al., "Aiexplainability 360: An extensible toolkit for understanding data andmachine learning models."J. Mach. Learn. Res., vol. 21, no. 130, pp.1–6, 2020.	85	N/A
RuleX AI	Rulex	Human Oversight	N/A	18	https://www.rulex.ai/
Amazon SageMaker Edge Manager	Amazon	Human Oversight	N/A	3	https://aws.amazon.com/de/sagemaker/
Neptune.ai	Neptune	Human Oversight	N/A	0	https://neptune.ai/product
Tensorflow Responsible AI	Google	Human Oversight	J. Wexler, M. Pushkarna, T. Bolukbasi, M. Wattenberg, F. Vi´egas, andJ. Wilson, "The what-if tool: Interactive probing of machine learningmodels,"IEEE Transactions on Visualization and Computer Graphics,vol. 26, no. 1, pp. 56–65, 2020.	0	N/A

Appendix II.6: Identified Software Solutions - Accuracy, Robustness, Cybersecurity

Framework	Publisher	Category	Original Paper	Scholar Results	Alternative Source
Foolbox Native	Rauber, J.	Accuracy, Robustness, Cybersecurity	J. Rauber, R. Zimmermann, M. Bethge, and W. Brendel, "Foolboxnative: Fast adversarial attacks to benchmark the robustness of machinelearning models in pytorch, tensorflow, and jax,"Journal of Open SourceSoftware, vol. 5, no. 53, p. 2607, 2020.	498	https://github.com/bethgelab/foolbox
IBM Adversarial Robustness Toolbox	IBM	Accuracy, Robustness, Cybersecurity	M.-I. Nicolae, M. Sinn, M. N. Tran, B. Buesser, A. Rawat, M. Wistuba,V. Zantedeschi, N. Baracaldo, B. Chen, H. Ludwiget al., "Adversarialrobustness toolbox v1. 0.0,"arXiv preprint arXiv:1807.01069, 2018.	305	https://github.com/Trusted-AI/adversarial-robustness-toolbox
IBM CNN-Cert	IBM	Accuracy, Robustness, Cybersecurity	A. Boopathy, T.-W. Weng, P.-Y. Chen, S. Liu, and L. Daniel, "Cnn-cert: An efficient framework for certifying robustness of convolutionalneural networks," inProceedings of the AAAI Conference on ArtificialIntelligence, vol. 33, no. 01, 2019, pp. 3240–3247.	85	https://github.com/IBM/CNN-Cert
CORTEX CERTIFAI	CognitiveScale	Accuracy, Robustness, Cybersecurity	S. Sharma, J. Henderson, and J. Ghosh, "Certifai: Counterfactual explanations for robustness, transparency, interpretability, and fairness ofartificial intelligence models,"arXiv preprint arXiv:1905.07857, 2019.	59	https://www.cognitivescale.com/certifai/
IBM Research AI Fairness 360 Toolkit	IBM	Accuracy, Robustness, Cybersecurity	R. K. E. Bellamy, K. Dey, M. Hind, S. C. Hoffman, S. Houde,K. Kannan, P. Lohia, J. Martino, S. Mehta, A. Mojsilović, S. Nagar,K. N. Ramamurthy, J. Richards, D. Saha, P. Sattigeri, M. Singh, K. R.Varshney, and Y. Zhang, "Ai fairness 360: An extensible toolkit fordetecting and mitigating algorithmic bias,"IBM Journal of Researchand Development, vol. 63, no. 4/5, pp. 4:1–4:15, 2019.	48	http://aif360.mybluemix.net/?_ga=2.219990425.1482784964.1625572737-971823234.1625572737

Appendix III.1: Evaluation of Software Solutions - Testing

Req. ID	Framework	Evaluation	Comment	Source
9.14	Amazon Sage Maker	1	Amazon Sage Maker can be used to test the AI system with A/B testing and record its performance. Insights about performance differences might be used to design risk systems (with considerable effort).	https://docs.aws.amazon.com/sagemaker/latest/dg/model-ab-testing.html
9.15	Amazon Sage Maker	1	Amazon Sage Maker can be used to record performance but not to explicitly/automatically test for consistent metric values. This could be automized manually.	https://docs.aws.amazon.com/sagemaker/latest/dg/model-ab-testing.html
9.16	Amazon Sage Maker	0	Amazon Sage Maker does not provide purpose-specific testing abilities.	https://docs.aws.amazon.com/sagemaker/latest/dg/model-ab-testing.html
9.17	Amazon Sage Maker	N/A	Process requirement	N/A
9.18	Amazon Sage Maker	2	Amazon Sage Maker provides industry standardized ML model metrics. Thresholds could be applied to these metrics manually. However, no automatic check for exceeding these thresholds is provided.	https://docs.aws.amazon.com/sagemaker/latest/dg/model-ab-testing.html
9.14	Watson OpenScale	2	Monitoring of the model and reporting of customary metrics including with respect to fairness can be used to manually design an adequate risk management system. Test and assessment functions especially targeted at risk management.	https://dataplatform.cloud.ibm.com/docs/content/wsj/model/wos-insight-timechart.html https://dataplatform.cloud.ibm.com/docs/content/wsj/model/cloud-risk-wos-only.html
9.15	Watson OpenScale	1	Ability to record performance and related metrics but not explicitly/automatically test for consistent metric values; this could be automized manually	https://dataplatform.cloud.ibm.com/docs/content/wsj/model/getting-started.html
9.16	Watson OpenScale	2	Watson OpenScale provides model fairness (and quality) metrics depending on different use cases where bias is prevalent.	https://dataplatform.cloud.ibm.com/docs/content/wsj/model/getting-started.html https://dataplatform.cloud.ibm.com/docs/content/wsj/model/wos-fairness-group.html
9.17	Watson OpenScale	N/A	Process requirement	N/A
9.18	Watson OpenScale	2	Watson OpenScale provides industry standardized ML model metrics. Thresholds could be applied to these metrics manually. However, no automatic check for exceeding these thresholds is provided. Provides ability to create custom metrics.	https://dataplatform.cloud.ibm.com/docs/content/wsj/model/getting-started.html
9.14	Azure ML	1	Azure ML can be used to test the AI system with A/B testing and record its performance. Insights about performance differences might be used to design risk systems (with considerable effort).	https://docs.microsoft.com/en-us/azure/machine-learning/how-to-deploy-azure-kubernetes-service?tabs=python#deploy-models-to-aks-using-controlled-rollout-preview
9.15	Azure ML	1	Azure ML can be used to record performance but not to explicitly/automatically test for consistent metric values. This could be automized manually.	https://docs.microsoft.com/en-us/azure/machine-learning/how-to-deploy-azure-kubernetes-service?tabs=python#deploy-models-to-aks-using-controlled-rollout-preview; https://docs.microsoft.com/en-us/azure/machine-learning/how-to-enable-data-collection
9.16	Azure ML	0	Azure ML does not provide purpose-specific testing abilities.	https://docs.microsoft.com/en-us/azure/machine-learning/how-to-deploy-azure-kubernetes-service?tabs=python#deploy-models-to-aks-using-controlled-rollout-preview
9.17	Azure ML	N/A	Process requirement	N/A
9.18	Azure ML	2	Azure ML provides industry standardized ML model metrics. Thresholds could be applied to these metrics manually. However, no automatic check for exceeding these thresholds is provided.	https://docs.microsoft.com/en-us/azure/machine-learning/how-to-deploy-azure-kubernetes-service?tabs=python#deploy-models-to-aks-using-controlled-rollout-preview

Appendix III.2: Evaluation of Software Solutions - Dataset Properties 1/2

Req. ID	Framework	Evaluation	Comment	Source
10.1	IBM SPSS Modeler	N/A	Process requirement	N/A
10.2	IBM SPSS Modeler	N/A	Process requirement	N/A
10.3	IBM SPSS Modeler	3	SPSS Modeler is explicitly build around CRISP-DM process. Data Preparation is one step in the CRISP-DM process.	https://www.ibm.com/docs/en/spss-modeler/18.1.1?topic=preparation-data-overview
10.4	IBM SPSS Modeler	N/A	Process requirement	N/A
10.5	IBM SPSS Modeler	2	SPSS Modeler only provides tools for verifying data quality. Assessing availability and suitability of data sets may require additional expert knowledge.	https://www.ibm.com/docs/en/spss-modeler/18.1.1?topic=understanding-verifying-data-quality
10.6	IBM SPSS Modeler	0	No explicit functionality for detecting biases in the data sets.	N/A
10.7	IBM SPSS Modeler	3	SPSS Modeler provides tools for verifying data quality which includes detecting and addressing missing data with various methods.	https://www.ibm.com/docs/en/spss-modeler/18.1.1?topic=understanding-verifying-data-quality, https://www.ibm.com/docs/en/wsd?topic=values-handling-missing
10. 8	IBM SPSS Modeler	1	SPSS Modeler provides tools for verifying data quality which includes detecting errors in the data. Assessing whether it is relevant and representative will require additional expert knowledge.	https://www.ibm.com/docs/en/spss-modeler/18.0.0?topic=nodes-data-audit-node
10.9	IBM SPSS Modeler	1	General data exploration to analyze statistical properties of the data possible. Does not assess whether data sets have statistical properties with regard to users/groups of users.	https://www.ibm.com/docs/en/spss-modeler/18.1.1?topic=understanding-exploring-data
10.10	IBM SPSS Modeler	0	Does not provide functionality to assess whether datasets have characteristics specific to the geographical, behavioral or functional setting.	N/A
10.1	SAP Data Services	N/A	Process requirement	N/A
10.2	SAP Data Services	N/A	Process requirement	N/A
10.3	SAP Data Services	3	Provides data quality functionality that includes cleansing, enhancing, matching and consolidating data elements which is part of data preparation.	https://help.sap.com/viewer/ec06fadc50b64b6184f835e4f0e1f52f/4.2.14/en-US/572548f96d6d1014b3fc9283b0e91070.html
10.4	SAP Data Services	N/A	Process requirement	N/A
10.5	SAP Data Services	2	SAP Data Services provide data assessment tools to determine the quality of the data. Assessing availability and suitability of data sets may require additional expert knowledge.	https://help.sap.com/viewer/ec06fadc50b64b6184f835e4f0e1f52f/4.2.14/en-US/5724c17c6d6d1014b3fc9283b0e91070.html
10.6	SAP Data Services	0	No functionality for detecting biases in the data sets.	N/A
10.7	SAP Data Services	1	Detecting and filtering missing or bad values is possible, but no specific methods for handling these values are available.	https://help.sap.com/viewer/e54136ab6a4a43e6a370265bf0a2d744/4.2.14/en-US/5739b9f56d6d1014b3fc9283b0e91070.html
10. 8	SAP Data Services	1	Provides functionality as part of Data Assessment tool to identify and fix errors in the data. Assessing whether it is relevant and representative will require additional expert knowledge.	https://help.sap.com/viewer/e54136ab6a4a43e6a370265bf0a2d744/4.2.7/en-US/5724c17c6d6d1014b3fc9283b0e91070.html
10.9	SAP Data Services	1	Provides functionality to obtain statistics about the cleansing and assignment processes as part of data quality. Does not assess whether data sets have statistical properties with regard to users/groups of users.	https://help.sap.com/viewer/e54136ab6a4a43e6a370265bf0a2d744/4.2.7/en-US/24cad7e5f0fb44a8865e3451b1eff56c.html?q=data%20statistics
10.10	SAP Data Services	0	Does not provide functionality to assess whether datasets have characteristics specific to the geographical, behavioral or functional setting.	N/A

Appendix III.2: Evaluation of Software Solutions - Dataset Properties 2/2

Req. ID	Framework	Evaluation	Comment	Source
10.1	Informatica Data Quality	N/A	Process requirement	N/A
10.2	Informatica Data Quality	N/A	Process requirement	N/A
10.3	Informatica Data Quality	2	Provides data quality and profiling functionality that includes some data preparation steps.	https://docs.informatica.com/data-quality-and-governance/data-quality/10-5/data-quality-getting-started-guide/getting-started-overview/informatica-developer-overview/data-quality-and-profiling.html
10.4	Informatica Data Quality	N/A	Process requirement	N/A
10.5	Informatica Data Quality	2	Provides comprehensive "data quality capabilities". Assessing availability and suitability of data sets may require additional expert knowledge.	https://docs.informatica.com/data-quality-and-governance/data-quality/10-5/data-quality-getting-started-guide/getting-started-overview/informatica-developer-overview/data-quality-and-profiling.html
10.6	Informatica Data Quality	0	No explicit functionality for detecting biases in the data sets.	N/A
10.7	Informatica Data Quality	0	No explicit functionality mentioned in the documentation that detects and addresses missing data (gaps).	N/A
10.8	Informatica Data Quality	1	Provides functionality to remove errors as part of data standardization. Assessing whether it is relevant and representative will require additional expert knowledge.	https://docs.informatica.com/data-quality-and-governance/data-quality/10-5/data-quality-getting-started-guide/getting-started-overview/informatica-developer-overview/data-quality-and-profiling.html
10.9	Informatica Data Quality	0	Does not assess whether data sets have statistical properties with regard to users/groups of users.	N/A
10.10	Informatica Data Quality	0	Does not provide functionality to assess whether datasets have characteristics specific to the geographical, behavioral or functional setting.	N/A

Appendix III.3: Evaluation of Software Solutions - Record Keeping 1/2

Req. ID	Framework	Evaluation	Comment	Source
12.1	TensorBoard	2	TensorBoard provides callback functionalities, which ensures that logs are created and stored at a central customizable location, hence automatic event recording capabilities can be recorded and accessed.	https://www.tensorflow.org/tensorboard/get_started
12.2	TensorBoard	1	TensorBoard only provides logging and visualization of training data and therefore does not facilitate record keeping during the system's entire lifecycle.	https://pubs.rsna.org/doi/pdf/10.1148/ryai.2020200012; https://github.com/tensorflow/tensorboard/blob/master/docs/get_started.ipynb
12.3	TensorBoard	0	Insufficient information on whether TensorBoard's record keeping complies with an industry-acknowledged standard or common practice.	https://pubs.rsna.org/doi/pdf/10.1148/ryai.2020200012; https://www.tensorflow.org/tensorboard/scalars_and_keras
12.4	TensorBoard	1	TensorBoard is inherently not able to save input data. However, the closely affiliated TensorFlow framework can be used to store entire data-stream inputs and subsequently access specific groups of them directly to TensorBoard.	https://stackoverflow.com/questions/42355122/can-i-export-a-tensorflow-summary-to-csv
12.5	TensorBoard	1	TensorBoard's visualization capabilities main purpose is facilitating the manual analysis of logged training data, rather than a comprehensive, automatic data analysis.	https://pubs.rsna.org/doi/pdf/10.1148/ryai.2020200012; https://www.tensorflow.org/tensorboard/scalars_and_keras
12.6	TensorBoard	N/A	Process requirement	N/A
12.7	TensorBoard	0	Insufficient information available.	N/A
12.8	TensorBoard	2	TensorBoard's visualized datasets include timestamps, which allows to retrace all periods of use.	https://pubs.rsna.org/doi/pdf/10.1148/ryai.2020200012; https://www.tensorflow.org/tensorboard/scalars_and_keras
12.9	TensorBoard	0	Insufficient information available.	N/A
12.10	TensorBoard	0	Insufficient information available.	N/A
12.11	TensorBoard	0	Insufficient information available.	N/A
12.1	Amazon CloudWatch	3	CloudWatch provides functionality for visibility of metrics and logs data, data retention, and the ability to perform calculations on metrics.	https://s3.cn-north-1.amazonaws.com.cn/aws-dam-prod/china/pdf/acw-dg.pdf; https://aws.amazon.com/cloudwatch/features/?nc1=h_ls
12.2	Amazon CloudWatch	2	CloudWatch's comprehensive metric and logging capabilities should ensure a high level of trackability, but there is insufficient information on whether it covers an AI system's entire lifecycle.	https://s3.cn-north-1.amazonaws.com.cn/aws-dam-prod/china/pdf/acw-dg.pdf; https://aws.amazon.com/cloudwatch/features/?nc1=h_ls
12.3	Amazon CloudWatch	0	Insufficient information on whether CloudWatch's record keeping complies with an industry-acknowledged standard or common practice.	https://aws.amazon.com/cloudwatch/features/?nc1=h_ls
12.4	Amazon CloudWatch	2	The Amazon CloudWatch Logs service provides functionality to collect and store logs from resources, applications, and services in near real-time.	https://s3.cn-north-1.amazonaws.com.cn/aws-dam-prod/china/pdf/acw-dg.pdf; https://aws.amazon.com/cloudwatch/features/?nc1=h_ls
12.5	Amazon CloudWatch	2	CloudWatch custom metrics are automatically extracted from the ingested logs. Further analysis is provided by CloudWatch Logs Insights' advanced query language.	https://s3.cn-north-1.amazonaws.com.cn/aws-dam-prod/china/pdf/acw-dg.pdf; https://aws.amazon.com/cloudwatch/features/?nc1=h_ls

Appendix III.3: Evaluation of Software Solutions - Record Keeping 2/2

Req. ID	Framework	Evaluation	Comment	Source
12.6	Amazon CloudWatch	N/A	Process requirement	N/A
12.7	Amazon CloudWatch	0	Insufficient information available.	N/A
12.8	Amazon CloudWatch	2	Amazon CloudWatch provides functionality to collect custom metrics which may include timestamps or "user activity", enabling the monitoring of periods of use.	https://s3.cn-north-1.amazonaws.com.cn/aws-dam-prod/china/pdf/acw-dg.pdf; https://aws.amazon.com/cloudwatch/features/?nc1=h_ls
12.9	Amazon CloudWatch	0	Insufficient information available.	N/A
12.10	Amazon CloudWatch	0	Insufficient information available.	N/A
12.11	Amazon CloudWatch	0	Insufficient information available.	N/A
12.1	DataDog	1	DataDog provides various application monitoring functionality, building upon logs generated by the AI system, but does not facilitate the creation of logs itself.	https://skemman.is/bitstream/1946/28745/1/Project%20Report.pdf; https://www.datadoghq.com/product/log-management/
12.2	DataDog	2	DataDog's application monitoring capabilities support auditing or investigations on logs and therefore the traceability of the system.	https://skemman.is/bitstream/1946/28745/1/Project%20Report.pdf; https://www.datadoghq.com/product/log-management/
12.3	DataDog	0	Insufficient information on whether DataDog's record keeping complies with an industry-acknowledged standard or common practice.	https://skemman.is/bitstream/1946/28745/1/Project%20Report.pdf; https://www.datadoghq.com/product/log-management/
12.4	DataDog	0	Insufficient information available.	N/A
12.5	DataDog	0	Insufficient information available.	N/A
12.6	DataDog	N/A	Process requirement	N/A
12.7	DataDog	0	Insufficient information available.	N/A
12.8	DataDog	0	Insufficient information available.	N/A
12.9	DataDog	0	Insufficient information available.	N/A
12.10	DataDog	0	Insufficient information available.	N/A
12.11	DataDog	0	Insufficient information available.	N/A

Appendix III.4: Evaluation of Software Solutions - Explainability

Req. ID	Framework	Evaluation	Comment	Source
13.1	SHapley Additive exPlanations (SHAP)	2	SHAP computes Shapley values from game theory and provides feature explanations. It explains the Machine Learning model prediction of a data instance by computing the contribution (= importance) of each feature to that prediction. Makes the model more transparent as it tries to explain its output.	https://arxiv.org/pdf/1705.07874.pdf
13.1	Local Interpretable Model-Agnostic Explanation (LIME)	2	LIME is a model-agnostic explainability method that explains a complex Machine Learning model by approximating it locally with a simpler model that is in itself explainable. LIME has been mostly applied for image and text data. It makes the model more transparent as it tries to explain its output.	https://arxiv.org/pdf/1602.04938v3.pdf
13.1	AIX360 Toolkit	3	AIX360 Toolkit provides 8 state-of-the-art explanation algorithms that can be selected depending on the type of explanation needed. All algorithms may provide a high level of explainability.	https://arxiv.org/pdf/1909.03012.pdf

Appendix III.5: Evaluation of Software Solutions - Human Oversight 1/2

Req. Id	Framework	Evaluation	Comment	Source
14.1	SHapley Additive exPlanations (SHAP)	0	Out of Scope. A human-machine interface tool is not provided to monitor the continuous deployment of the model during operation. SHAP focuses on explaining the prediction (output) of a Machine Learning model.	N/A
14.2	SHapley Additive exPlanations (SHAP)	N/A	Process requirement	N/A
14.3	SHapley Additive exPlanations (SHAP)	N/A	Process requirement	N/A
14.4	SHapley Additive exPlanations (SHAP)	N/A	Process requirement	N/A
14.5	SHapley Additive exPlanations (SHAP)	1	SHAP is inherently not able to prevent automation-bias. However, it makes ML models more transparent by explaining the contribution of each feature to the model output. It could help preventing automation bias if subject matter experts are required to evaluate the produced explanations. If explanations do not make sense or are not as expected, this could be a warning that the output is flawed and hence, could help preventing over-reliance on the system.	https://github.com/slundberg/shap
14.6	SHapley Additive exPlanations (SHAP)	3	SHAP computes Shapley values from game theory and provides feature explanations. It explains the Machine Learning model prediction of a data instance by computing the contribution (= importance) of each feature to that prediction. Makes the model more transparent as it tries to explain its output.	https://stats.stackexchange.com/questions/379744/comparison-between-shap-shapley-additive-explanation-and-lime-local-interpret
14.7	SHapley Additive exPlanations (SHAP)	0	Out of Scope. A human-machine interface tool is not provided to monitor the continuous deployment of the model during operation. SHAP focuses on explaining the prediction (output) of a Machine Learning model.	N/A
14.8	SHapley Additive exPlanations (SHAP)	0	Out of Scope. A human-machine interface tool is not provided to monitor the continuous deployment of the model during operation. SHAP focuses on explaining the prediction (output) of a Machine Learning model.	N/A
14.9	SHapley Additive exPlanations (SHAP)	0	Out of Scope. SHAP cannot be used to check if the high-risk AI system is operating such that its decisions with respect to identification, assignment and assessment of natural persons in educational and vocational training institutions, are confirmed by at least two natural persons. SHAP focuses on explaining the prediction (output) of a Machine Learning model.	https://github.com/slundberg/shap
14.1	Local Interpretable Model-Agnostic Explanation (LIME)	0	Out of Scope. A human-machine interface tool is not provided to monitor the continuous deployment of the model during operation. SHAP focuses on explaining the prediction (output) of a Machine Learning model.	N/A
14.2	Local Interpretable Model-Agnostic Explanation (LIME)	N/A	Process requirement	N/A
14.3	Local Interpretable Model-Agnostic Explanation (LIME)	N/A	Process requirement	N/A
14.4	Local Interpretable Model-Agnostic Explanation (LIME)	N/A	Process requirement	N/A

Appendix III.5: Evaluation of Software Solutions - Human Oversight 2/2

Req. ID	Framework	Evaluation	Comment	Source
14.5	Local Interpretable Model-Agnostic Explanation (LIME)	1	LIME does not provide this functionality inherently. However, it could potentially be added manually via code not native to the framework.	https://github.com/marcotcr/lime
14.6	Local Interpretable Model-Agnostic Explanation (LIME)	2	LIME is a model-agnostic explainability method that explains a complex Machine Learning model by approximating it locally with a simpler model that is in itself explainable. LIME has been mostly applied for image and text data. It makes the model more transparent as it tries to explain its output.	https://github.com/marcotcr/lime
14.7	Local Interpretable Model-Agnostic Explanation (LIME)	0	Out of Scope. A human-machine interface tool is not provided to monitor the continuous deployment of the model during operation. LIME focuses on explaining the prediction (output) of a Machine Learning model.	N/A
14.8	Local Interpretable Model-Agnostic Explanation (LIME)	0	Out of Scope. A human-machine interface tool is not provided to monitor the continuous deployment of the model during operation. LIME focuses on explaining the prediction (output) of a Machine Learning model.	N/A
14.9	Local Interpretable Model-Agnostic Explanation (LIME)	0	Out of Scope. LIME cannot be used to check if the high-risk AI system is operating such that its decisions with respect to identification, assignment and assessment of natural persons in educational and vocational training institutions, are confirmed by at least two natural persons. LIME focuses on explaining the prediction (output) of a Machine Learning model.	https://github.com/marcotcr/lime/blob/master/doc/notebooks/Tutorial%20-%20Faces%20and%20GradBoost.ipynb https://github.com/marcotcr/lime/blob/master/lime/lime_text.py
14.1	MLflow	3	The MLflow Tracking component is an API and UI for logging parameters, code versions, metrics, and output files when running a Machine Learning model. It also allows to later visualize the results. As a result, it provides broad oversight capabilities.	https://mlflow.org/docs/latest/tracking.html#concepts
14.2	MLflow	N/A	Process requirement	N/A
14.3	MLflow	N/A	Process requirement	N/A
14.4	MLflow	N/A	Process requirement	N/A
14.5	MLflow	0	Out of Scope. MLflow does not provide functionality that controls or stops users from over relying on the AI-system.	N/A
14.6	MLflow	2	MLflow provides tools for general interpretation via various metrics and visualizes them in real-time in a dashboard. However, in terms of transparency, the tools available in the MLflow framework are lacking. Nonetheless, integration with other explainability frameworks such as SHAP or LIME could be possible. Using additional frameworks could help providing transparency for the high-risk AI system.	https://mlflow.org/docs/latest/tracking.html#visualizing-metrics
14.7	MLflow	0	Out of Scope. As an AI framework for managing rather than designing AI systems, there is no direct support for such functionality. However, MLflow could potentially be used in synergy with other frameworks to achieve the required objective.	N/A
14.8	MLflow	2	Partly realized via command line interruption directly within the server with Ctrl C or by using 'pkill -f gunicorn' on the server.	https://stackoverflow.com/questions/60531166/how-to-safely-shutdown-mlflow-ui https://github.com/mlflow/mlflow/issues/1856
14.9	MLflow	0	Out of Scope. As an AI framework for managing rather than designing AI systems, there is no direct support for such functionality. However, MLflow could potentially be used in synergy with other frameworks to achieve the required objective.	N/A

Appendix III.6: Evaluation of Software Solutions - Accuracy, Robustness, Cybersecurity 1/2

Req. ID	Framework	Evaluation	Comment	Source
15.1	Foolbox Native	N/A	Process requirement	N/A
15.2	Foolbox Native	1	Foolbox provides attack models for adversarial training. There is a trade-off between robustness ("robust accuracy") and accuracy ("standard accuracy"). A consistent level of robustness through should lead to a consistent level of accuracy.	https://foolbox.jonasrauber.de/guide/getting-started.html#robust-accuracy https://arxiv.org/abs/1805.12152
15.3	Foolbox Native	3	Foolbox provides a variety of "adversarial attacks to benchmark the robustness of machine learning models".	https://foolbox.readthedocs.io/en/stable/modules/attacks.html
15.4	Foolbox Native	2	Foolbox provides adversarial training, which helps mitigating adversarial attacks, but is not sufficient to achieve "cybersecurity" as a whole.	https://publications.jrc.ec.europa.eu/repository/bitstream/JRC119336/dpad_report.pdf
15.5	Foolbox Native	N/A	Process requirement	N/A
15.6	Foolbox Native	0	Foolbox provides adversarial training, but does not address technical redundancy or fault prevention.	N/A
15.7	Foolbox Native	0	Foolbox provides adversarial training, but does not address biased outputs through 'feedback loops'.	N/A
15.8	Foolbox Native	2	Adversarial training mitigates adversarial attacks, being a popular way of AI-System manipulation, but does not generally prevent unauthorized access by third parties.	https://foolbox.jonasrauber.de
15.9	Foolbox Native	N/A	Process requirement	N/A
15.10	Foolbox Native	2	Data poisoning is considered a specific strategy of adversarial attacks, which are implicitly addressed by the framework.	https://foolbox.readthedocs.io/en/stable/modules/attacks.html
15.11	Foolbox Native	3	Adversarial examples are considered a specific strategy of adversarial attacks, which are explicitly addressed by the framework.	https://foolbox.readthedocs.io/en/stable/modules/attacks.html
15.12	Foolbox Native	2	Model flaw exploitation is considered a specific strategy of adversarial attacks, which are implicitly addressed by the framework.	https://foolbox.readthedocs.io/en/stable/modules/attacks.html
15.1	IBM Adversarial Robustness Toolbox	N/A	Process requirement	N/A
15.2	IBM Adversarial Robustness Toolbox	1	Adversarial Robustness Toolbox provides attack models for adversarial training. There is a trade-off between robustness ("robust accuracy") and accuracy ("standard accuracy"). A consistent level of robustness through should lead to a consistent level of accuracy.	https://adversarial-robustness-toolbox.readthedocs.io/en/latest/ https://arxiv.org/abs/1805.1
15.3	IBM Adversarial Robustness Toolbox	3	Adversarial Robustness Toolbox includes certifying and verifying model robustness and model hardening, which ensures a consistent level of robustness.	https://adversarial-robustness-toolbox.readthedocs.io/en/latest/modules/metrics.html?highlight=robustness#empirical-robustness https://arxiv.org/pdf/1807.01069.pdf
15.4	IBM Adversarial Robustness Toolbox	2	Adversarial Robustness Toolbox provides adversarial training, which helps mitigating adversarial attacks, but is not sufficient to achieve "cybersecurity" as a whole.	https://publications.jrc.ec.europa.eu/repository/bitstream/JRC119336/dpad_report.pdf
15.5	IBM Adversarial Robustness Toolbox	N/A	Process requirement	N/A
15.6	IBM Adversarial Robustness Toolbox	0	Adversarial Robustness Toolbox provides adversarial training, but does not address technical redundancy or fault prevention.	N/A
15.7	IBM Adversarial Robustness Toolbox	0	Adversarial Robustness Toolbox provides adversarial training, but does not address biased outputs through 'feedback loops'.	N/A
15.8	IBM Adversarial Robustness Toolbox	2	Adversarial training mitigates adversarial attacks, being a popular way of AI-System manipulation, but does not generally prevent unauthorized access by third parties.	https://adversarial-robustness-toolbox.readthedocs.io/en/latest/

Appendix III.6: Evaluation of Software Solutions - Accuracy, Robustness, Cybersecurity 2/2

Req. ID	Framework	Evaluation	Comment	Source
15.9	IBM Adversarial Robustness Toolbox	N/A	Process requirement	N/A
15.10	IBM Adversarial Robustness Toolbox	2	Data poisoning is considered a specific strategy of adversarial attacks, which are addressed by the framework through a "module providing poisoning attacks under a common interface."	https://adversarial-robustness-toolbox.readthedocs.io/en/latest/modules/attacks/poisoning.html?highlight=poisoning
15.11	IBM Adversarial Robustness Toolbox	3	Adversarial examples are considered a specific strategy of adversarial attacks, which are addressed by the framework. "The attacks implemented in ART allow creating adversarial attacks against Machine Learning models which is required to test defenses with state-of-the-art threat models."	https://adversarial-robustness-toolbox.readthedocs.io/en/latest/modules/attacks.html; https://arxiv.org/pdf/1807.01069.pdf
15.12	IBM Adversarial Robustness Toolbox	2	Model flaw exploitation is considered a specific strategy of adversarial attacks, which are addressed by the framework.	https://adversarial-robustness-toolbox.readthedocs.io/en/latest/modules/attacks.html
15.1	CNN-Cert	N/A	Process requirement	N/A
15.2	CNN-Cert	1	CNN-Cert provides a mechanism for robustness certification. There is a trade-off between robustness ("robust accuracy") and accuracy ("standard accuracy"). Verification of robustness may indirectly support verification of accuracy.	https://arxiv.org/abs/1805.1
15.3	CNN-Cert	2	CNN-Cert provides Robustness verification, which passively supports ensuring a consistent level of robustness.	https://medium.com/@MITIBMLab/cnn-cert-a-certified-measure-of-robustness-for-convolutional-neural-networks-fd2ff44c6807
15.4	CNN-Cert	1	CNN-Cert provides Robustness verification, which does only assist in identifying an AI-System's potential cybersecurity risk.	https://medium.com/@MITIBMLab/cnn-cert-a-certified-measure-of-robustness-for-convolutional-neural-networks-fd2ff44c6807
15.5	CNN-Cert	N/A	Process requirement	N/A
15.6	CNN-Cert	0	CNN-Cert provides a mechanism for robustness certification, but does not address technical redundancy or fault prevention.	N/A
15.7	CNN-Cert	0	CNN-Cert provides a mechanism for robustness certification, but does not address biased outputs through 'feedback loops'.	N/A
15.8	CNN-Cert	0	Verifying the robustness of an AI-System does only passively assist in identifying a low level of robustness as a potential point of vantage for unauthorized third parties.	N/A
15.9	CNN-Cert	N/A	Process requirement	N/A
15.10	CNN-Cert	1	Robustness verification does not sufficiently prove and actively test protection against data poisoning.	https://arxiv.org/pdf/1811.12395.pdf
15.11	CNN-Cert	1	Robustness verification does not sufficiently prove and actively test protection against adversarial examples.	https://arxiv.org/pdf/1811.12395.pdf
15.12	CNN-Cert	1	Robustness verification does not sufficiently prove and actively test protection against model flaw exploitation.	https://arxiv.org/pdf/1811.12395.pdf

APPENDIX IV:
AMBIGUOUS AND VAGUE WORDINGS AND PHRASINGS

AI Act Article	Derived Req. ID	Content	Comment
9 (4c)	9.12	"adequate […] training to users"	It is unclear whether training is related to the high-risk AI system or to the risk response from users? If the former, it is not defined how any such training shall be designed or carried out.
9 (5)	9.15	"perform consistently for their intended purpose"	It is unclear what consistent performance of a purpose means. For instance, shall the same inputs passed to a high-risk AI system lead to the same outputs? In addition, intended purpose is only vaguely defined.
9 (7)	9.18	"preliminarily defined metrics and probabilistic thresholds"	It is not defined which metrics are considered suitable, in which way thresholds should be defined, and what are appropriate levels for thresholds regarding specific high-risk AI system purposes for customary metrics.
10 (2a)	10.1	"[...] relevant design choices."	Design choices regarding data features, AI system/data platform architecture? No sufficient information given, what design choices are about.
10 (2d)	10.4	"[...] relevant assumptions" (about the given data sets)	No examples are given for assumptions. What are possible assumptions?
10 (2e)	10.5	"suitability of the data sets that are needed."	What data sets classify as suitable?
10 (3)	10.9	"appropriate statistical properties […] as regards the persons or groups of persons […]"	What are statistical properties regarding users/groups of users are deemed appropriate?
10 (4)	10.10	"[...] characteristics or elements that are particular to the specific geographical, behavioural or functional setting."	What are characteristics that are specific to the mentioned settings? No examples are given.
11 (1)	11.1	"[...] all the necessary information to assess the compliance of the AI system [...]"	What information is considered necessary?
Annex IV (2d)	11.13	"[...] where relevant, the data requirements in terms of datasheets [...]"	What is considered relevant?
61 (2)	12.5	"systematically […] analyse relevant data"	It is not defined what a systematic analysis of data provided by users comprises of (e.g., aspects of input to consider, output) or how it should be carried out (e.g., tools, frequency, output storage) within the post-market monitoring system.
12 (4b)	12.9	"shall provide […] the reference database against which input data has been checked"	It is unclear whether the inclusion of a database in logging records refers to storing a reference to the database (e.g., ID, hyperlink), metadata about the database, or the database itself with the last option carrying the highest cost and being the least technically feasible
13 (1)	13.1	"their operation is sufficiently transparent to enable users to interpret the system's output"	It is unclear which level of transparency is required and how it should be achieved since the ability of users to interpret and use results is a vague objective.
15 (1)	15.1	"appropriate level of accuracy, robustness and cybersecurity"	What level of accuracy, robustness and cybersecurity is considered appropriate? In what metric are these levels measured?
15 (2)	15.5	"relevant accuracy metrics"	What accuracy metrics are considered relevant?
15 (4)	15.9	"appropriate to the relevant circumstances and the risks"	How can this appropriateness be specifically measured?