|  |  |  |
| --- | --- | --- |
| ITU Logo | INTERNATIONAL TELECOMMUNICATION UNION**TELECOMMUNICATIONSTANDARDIZATION SECTOR**STUDY PERIOD 2017-2020 | FG-AI4H-I-036 |
| **ITU-T Focus Group on AI for Health** |
| **Original: English** |
| **WG(s):** | Plenary | E-meeting, 7-8 May 2020 |
| **DOCUMENT** |
| **Source:** | Editors |
| **Title:** | Updated DEL2.2: Guidelines for AI based medical device: Regulatory requirements (Draft: April 2020) |
| **Purpose:** | Discussion |
| **Contact:** | Luis OalaHHI FraunhoferGermany | Email: luis.oala@hhi.fraunhofer.de |
| **Contact:** | Christian JohnerJohner Institute for Healthcare ITGermany | Email: christian.johner@johner-institut.de |
| **Contact:** | Pradeep BalachandranTechnical Consultant (eHealth) India | Email: pbn.tvm@gmail.com |

|  |  |
| --- | --- |
| **Abstract:** | This document contains the latest draft of the FG-AI4H deliverable DEL2.2 "Guidelines for AI based medical device: Regulatory requirements". This deliverable defines a set of guidelines intended to serve the AI solution developers/‌manufacturers on how to do conduct a comprehensive requirements analysis and to streamline the conformity assessment procedures to ensure regulatory compliance for the AI based Medical Devices (AI-MD). |

|  |  |
| --- | --- |
|  | **International Telecommunication Union** |
|  |  |
|  |  | **ITU-T** | **FG-AI4H Deliverable** |
|  |  | TELECOMMUNICATIONSTANDARDIZATIONSECTOROFITU | (draft 2020-04-14) |
|  |  |
|  | **FG AI4H DEL2.2Guidelines for AI Based Medical Device- Regulatory Requirements** |
|  |  |

Summary

This technical paper defines a set of guidelines intended to serve the AI solution developers/ manufacturers on how to do conduct a comprehensive requirements analysis and to streamline the conformity assessment procedures to ensure regulatory compliance for the AI based Medical Devices (AI-MD). This set of guidelines gives prime priority to the factor of patient safety and focuses on a streamlined process for risk minimization and quality assurance for AI based health solutions. The proposed set of guidelines adopts, extends and leverages the best practices and recommendations provided by international medical device regulatory agencies such as the IMDRF and the FDA. These guidelines are devoid any legally binding or statutory requirements applicable to any specific regulatory framework or specific geographic jurisdiction

Keywords

[Regulatory Checklist, Software-as-a-Medical Device, AI based Medical Devices]

Change Log

This document contains Version 1.0 of the FG-AI4H deliverable on "Guidelines for AI based Medical Devices - Regulatory Requirements" approved at the ITU-T Focus Group on Artificial Intelligence for Health meeting held in Geneva, dd-ddMMMM YYYY.

|  |  |  |
| --- | --- | --- |
| **Editors:** | Luis OalaHHI Fraunhofer, Germany | Email: luis.oala@hhi.fraunhofer.de |
|  | Christian JohnerJohner Institute for Healthcare IT, Germany | Email:christian.johner@johner-institut.de |
|  | PradeepBalachandranTechnical Consultant(eHealth), India | Email: pbn.tvm@gmail.com |

|  |  |  |
| --- | --- | --- |
| **Contributors:** | (in alphabetical order)Alixandro Werneck LeiteMachine Learning Laboratory of Finance and Organizations – LAMFO - BrazilAndrew MurchisonCompany Oxford University Hospitals NHS Foundation Trust, United KingdomPeter. G. Goldschmidt WORLD DEVELOPMENT GROUP, Inc., USA Shan XuChina Academy of Information and Communications Technology (CAICT)ChinaSven PiechottkaMerantix HealthcareGermany | Email: alixandrowerneck@outlook.comEmail: agmurchison@gmail.comEmail: pgg@has.comEmail: xushan@caict.ac.cnEmail: sven.piechottka@merantix.com |
|  |  |  |
|  |  |  |
|  |  |  |

**CONTENTS**

|  Page |
| --- |
| [1 Scope 10](#_Toc39496668)[2 References 10](#_Toc39496669)[3 Terms and definitions 11](#_Toc39496670)[3.1 Terms defined elsewhere 11](#_Toc39496671)[3.2 Terms defined here 12](#_Toc39496672)[4 Abbreviations and acronyms 12](#_Toc39496673)[5 Regulatory guidelines formulation: baseline references 12](#_Toc39496674)[6 Proposed roadmap for regulatory requirements guidelines 13](#_Toc39496675)[6.1 Guidelines: applicability scope 13](#_Toc39496676)[1.1.1 Scope of regulation 13](#_Toc39496677)[1.1.2 Scope of product 13](#_Toc39496678)[1.1.3 Scope of application 13](#_Toc39496679)[1.2 Guidelines: aim & objectives 13](#_Toc39496680)[1.3 Guidelines: target user class / roles 14](#_Toc39496681)[1.4 Requirements assessment model: a product lifecycle development based approach 14](#_Toc39496682)[2 Regulatory guidelines: proposed structure 17](#_Toc39496683)[3 Regulatory guidelines: requirements checklist 19](#_Toc39496684)[3.1 Pre-market requirements 20](#_Toc39496685)[3.1.1 Intended use and stakeholder requirements 22](#_Toc39496686)[3.1.2 Product and software requirements 29](#_Toc39496687)[3.1.3 Data management requirements 34](#_Toc39496688)[3.1.4 Model development requirements 41](#_Toc39496689)[3.1.5 Product development requirements 46](#_Toc39496690)[3.1.6 Product validation requirements 52](#_Toc39496691)[3.1.7 Product release requirements 55](#_Toc39496692)[3.2 Post-market requirements 56](#_Toc39496693)[3.2.1 Production, distribution & installation requirements 56](#_Toc39496694)[3.2.2 Post-market surveillance requirements 58](#_Toc39496695)[Annex A: IMDRF Essential Principles 63](#_Toc39496696)[Annex B: IMDRFSaMD Risk Categorization Framework 70](#_Toc39496697)[B.1 Criteria for determining the SaMDcategory 70](#_Toc39496698)[B.2 Levels of autonomy 71](#_Toc39496699)[Annex C: Johnerregulatory guidelines for AI- for medical devices 72](#_Toc39496700)[Annex D: FG-AI4H data and AI solution quality assessment criteria 73](#_Toc39496701)[Annex E: ITU ML5G high-level requirements mapping to AI for health requirements 80](#_Toc39496702)[Annex F: DIN SPEC 92001-AI devices life cycle processes requirements 83](#_Toc39496703)[Annex G: AI4H Project Deliverable Reference 92](#_Toc39496704) |

**List of Tables**

|  Page |
| --- |
| [Table 1: Process requirements 20](#_Toc39496627)[Table 2: Competencies requirements 21](#_Toc39496628)[Table 3: Intended use requirements 22](#_Toc39496629)[Table 4: Intended users and intended context of use requirements 24](#_Toc39496630)[Table 5: Stakeholder requirements 25](#_Toc39496631)[Table 6: Inputs to risk management and clinical evaluation requirements 27](#_Toc39496632)[Table 7: Functionality and performance requirements 29](#_Toc39496633)[Table 8: Userinterface requirements 31](#_Toc39496634)[Table 9: Additional software requirements 31](#_Toc39496635)[Table 10: Risk management & clinical evaluation requirements 32](#_Toc39496636)[Table 11: Data collection requirements 34](#_Toc39496637)[Table 12: Data annotation requirements 37](#_Toc39496638)[Table 13: Data pre-processing requirements 38](#_Toc39496639)[Table 14: Documentation and version control requirements 40](#_Toc39496640)[Table 15: Model preparation requirements 41](#_Toc39496641)[Table 16: Model training requirements 42](#_Toc39496642)[Table 17:Model evaluation requirements 43](#_Toc39496643)[Table 18: Model documentation requirements 45](#_Toc39496644)[Table 19: Software development requirements 47](#_Toc39496645)[Table 20: Accompanying materials requirements 49](#_Toc39496646)[Table 21: Usability validation requirements 52](#_Toc39496647)[Table 22: Clinical evaluation requirements 53](#_Toc39496648)[Table 23: Product release requirements 55](#_Toc39496649)[Table 24: Production, distribution and installation requirements 56](#_Toc39496650)[Table 25:Post-market surveillance requirements 58](#_Toc39496651)[Table A.1: IMDRFEP 5.1 – General 64](#_Toc39496652)[Table A.2: IMDRFEP 5.2 – Clinical evaluation 66](#_Toc39496653)[Table A.3: IMDRFEP 5.8 – Medical devices and IVD medical devices that incorporate software or are software as a medical device 66](#_Toc39496654)[Table A.4: IMDRFEP 5.10 – Labelling 67](#_Toc39496655)[Table A.5: IMDRFEP 5.12 – Protection against the risks posed by medical devices and IVD medical devices intended by the manufacturer for use by lay users 67](#_Toc39496656)[Table A.6: IMDRFEP7.2 – Performance characteristics 68](#_Toc39496657)[Table B.1: IMDRFSaMD risk categories 70](#_Toc39496658)[Table B.2: IMDRFSaMD risk categories(revised) 71](#_Toc39496659)[Table D.1:FG-AI4H data and AI solution quality assessment criteria 74](#_Toc39496660)[Table G: AI4H Project Deliverable Reference ID 92](#_Toc39496661) |

**List of Figures**

|  Page |
| --- |
| [Figure 1: AI-MD Classification Criteria & Scope 13](#_Toc39496662)[Figure 2(a): Regulatory roadmap-AI-medical device scope 14](#_Toc39496663)[Figure 2(b): Regulatory roadmap- AI4H project deliverables scope 15](#_Toc39496664)[Figure 3: Product development life-cycle process (V-model) 16](#_Toc39496665)[Figure 4: AI Software life-cycle diagram 16](#_Toc39496666)[Figure 5: Structure of the regulatory guidelines 18](#_Toc39496667) |

ITU-T FG-AI4H Deliverable 2.2

Guidelines for AI based medical devices-regulatory requirements

Summary

Artificial intelligence based technologies find extensive use in medical applications and the proliferation of AI at scale in global health holds great potential to significantly improve accessibility, quality, and cost-effectiveness of healthcare delivery and outcomes. Regulation plays an important role in ensuring the safety of patients and users, and in the commercialization and market acceptance of these AI based medical devices (AI-MD).Therefore, a streamlined and systematic regulatory compliance process can help to expedite the regulatory approval and to reduce the time-to-market for these products. Software is an important aspect of AI based medical devices and as per *The International Medical Device Regulators Forum (IMDRF), a 'Software as a Medical Device (SaMD)' is defined as a software intended to be used for one or more medical purposes that perform these purposes without being part of a hardware medical device, where 'medical purposes' include treatment, diagnosis, cure, mitigation, or prevention of disease or other conditions.*

Since Artificial intelligence based technologies are data driven, they have the ability to learn from real-world feedback data and thus improve and adapt their performance over time. Because of these characteristics, they are identified and classified as belonging to the class of continuous learning systems. The complexity introduced by these changes affects the clinical functionality, clinical performance specifications and safety of the medical device, and may introduce new risks or even lead to modification of existing risks. The fact that many of the AI processes are not transparent (black box phenomenon) also places barriers to acceptance. These characteristics raise important technological, methodological, ethical, privacy, security, and regulatory issues and there is an absolute need for reasonable assurance mechanisms to maintain and/or improve the performance, compatibility, safety and effectiveness of the medical devices.

Apart from these device-oriented issues, there are other challenges that include a lack of universally accepted policies and guidelines for regulation of AI based medicals devices, which create barriers for these type of devices to scale-up at the global level. Many medical devices companies do not have proper awareness of regulatory standards and thus fail to assess the potential implications of safety, ethical and legal risks There is a need for proper guidance mechanisms to educate and train medical device manufactures to work to regulatory guidelines applicable to AI based devices.

Current regulatory requirements are generally designed to evaluate more conventional medical devices, and there is also need for regulatory policies and guidelines to be tailored for AI based medicals devices. Regulatory frameworks need to adopt a full product lifecycle approach that facilitates continual product improvement in an iterative and adaptive manner in conformance to the appropriate standards and regulations. Furthermore, regulators and manufacturers must establish a system to ensure transparency and accountability of all the processes involved in AI medical device development. To aid manufactures meet these requirements, AI4H FG proposes a comprehensive set of guidelines on how to conduct a regulatory review process for pre-market and post-market product performance assessment and reporting. The main aim of these guidelines is to safeguard patient safety as its first priority through a streamlined process for manufacturers that will help ensure that products benefit patients by promoting health and minimizing risk. The proposed set of guidelines adopts, extends and leverages best practices and recommendations provided by the international medical device regulatory agencies such as the IMDRF and the FDA.

# Scope

This document defines a set of guidelines intended to serve the target audience with adequate guidance on how to do conduct a comprehensive requirements analysis and to streamline the conformity assessment procedures to ensure regulatory compliance for the AI for Medical Devices (AI-MD)

Primary and secondary audience…

The scope of AI-MD, in this context, DO include (a) medical devices with or without enforcement of regulations, (b) Software-as-a-Medical Device (SaMD) or Software-in-a-Medical Device (SiMD) and (c) healthcare applications intended to improve medical outcomes or efficiency of healthcare system

The scope of AI-MD, in this context, DOES NOT include software applications for (a)healthcare facility administrative support, (b) for maintaining or encouraging healthy lifestyle, behaviour and wellness

This set of guidelines complements existing quality practices, recommendations and standards developed by national and international medical device regulatory organizations including ISO, IEC, FDA, etc. within the AI-MD regulatory domain and is DOES NOT replace or conflict with their respective regulatory processes and standards

This set of guidelines DOES NOT include any legally binding or statutory requirements applicable to any specific regulatory framework or specific geographic jurisdiction

# References

The following list of reference documents were reviewed as part of broad literature survey towards the design of the proposed regulatory requirements guidelines, considering aspects of regulations, standards, guidelines, best-practices, directives and laws that are relevant in the context of AI-MD

[EU-IVDR], [ EU-MDR (2017/745) ] General-Definition of "medical device", General safety and performance requirements, Technical documentation on design, functions, components, Intended use, Intended users, Clinical evaluation, Post-market surveillance, Vigilance, etc.

[FDA 21 CFR] FDA 21 CFR part 820, Quality System Regulations

[FDA] Guidance "General Principles of Software Validation"

[FDA] FDA's "Proposed Regulatory Framework for Modifications to Artificial Intelligence / Machine Learning (AI/ML) Based Software as Medical Device

[GDPR] European General Data Protection Regulation

[IEC 62304] IEC 62304:2006 + A1:2015, "Medical device software – Software life cycle processes"

[IEC 62366] IEC 62366-1:2015, "Medical devices – Part 1: Application of usability engineering to medical devices"

[IEC 82304] IEC 82304-, “ Health software — Part 1: General requirements for product safety”

[IMDRF/GRRP WG/N47 FINAL: 2018] “Essential Principles of Safety and Performance of Medical Devices and IVD Medical Devices”

[ISO 13485] ISO 13485:2016, "Medical devices — Quality management systems — Requirements for regulatory purposes"

[ISO 14971] ISO 14971:2019, "Medical devices – Application of risk management to medical devices"

[ISO 9241-11] ISO 9241-11, " Ergonomics of human-system interaction —Part 11: Usability: Definitions and concepts"

[ISO 9241-210] ISO 9241-210, " Ergonomics of human-system interaction —Part 210: Human-centred design for interactive systems"

 [ISO 14155] ISO 14155, " Clinical investigation of medical devices for human subjects — Good clinical practice"

[ISO/IEC 27000] ISO/IEC 27000, " Information technology — Security techniques — Information security management systems — Overview and vocabulary"

[ISO/IEC 27002] ISO/IEC 27002, "Information technology — Security techniques — Code of practice for information security controls"

[ISO/IEC 27002:2013] ISO/IEC 27002:2013, "Information technology — Security techniques — Code of practice for information security controls, TECHNICAL CORRIGENDUM 1"

[ISO/IEC 27002:2013] ISO/IEC 27002:2013, "Information technology — Security techniques — Code of practice for information security controls, TECHNICAL CORRIGENDUM 2"

# Terms and definitions

## Terms defined elsewhere

This document uses the following terms defined elsewhere:

**3.1.1 Clinical Evaluation** [GHTF/SG5/N1R8:2007]: The assessment and analysis of clinical data pertaining to a medical device to verify the clinical safety and performance of the device when used as intended by the manufacturer

**3.1.2 In Vitro Diagnostic (IVD) Medical Device** [GHTF/SG1/N71:2012]:A medical device, whether used alone or in combination, intended by the manufacturer for the in-vitro examination of specimens derived from the human body solely or principally to provide information for diagnostic, monitoring or compatibility purposes

**3.1.3 Life-Cycle** [ISO/IEC Guide 51:2014]: All phases in the life of a medical device, from the initial conception to final decommissioning and disposal

**3.1.4 Medical Device** [GHTF/SG1/N71:2012]: Any instrument, apparatus, implement, machine, appliance, implant, reagent for in vitro use, software, material or other similar or related article, intended by the manufacturer to be used, alone or in combination, for human beings, for one or more of the specific medical purpose(s) of: a) diagnosis, prevention, monitoring, treatment or alleviation of disease, b) diagnosis, monitoring, treatment, alleviation of or compensation for an injury, c) investigation, replacement, modification, or support of the anatomy or of a physiological process, d) supporting or sustaining life, e) control of conception, f) disinfection of medical devices, g) providing information by means of in vitro examination of specimens derived from the human body; and does not achieve its primary intended action by pharmacological, immunological or metabolic means, in or on the human body, but which may be assisted in its intended function by such means

**3.1.5 Software as a Medical Device** [IMDRF/SaMD WG/N12FINAL:2014]:Software intended to be used for one or more medical purposes that perform these purposes without being part of a hardware medical device

**3.1.6 Software Validation** [IEEE-STD-610]: The process of evaluating software during or at the end of the development process to determine whether it satisfies specified requirements

**3.1.7 Software Verification** [IEEE-STD-610]: The process of evaluating software to determine whether the products of a given development phase satisfy the conditions imposed at the start of that phase

## Terms defined here

This document defines the following terms:

**3.2.1 term [reference]:** definition

**3.2.2 term [reference]:** definition

# Abbreviations and acronyms

This document uses the following abbreviations and acronyms.

|  |  |
| --- | --- |
| AI | Artificial intelligence |
| AI4H | Artificial intelligence for health |
| AI-MD | Artificial Intelligence based Medical Devices |
| DAISAM | Data and AI Solution Assessment Methods |
| EP | Essential Principle |
| FDA | Food and Drug Administration |
| GDPR | General Data Protection Regulation |
| HIPAA | [Health Insurance Portability and Accountability Act](https://en.wikipedia.org/wiki/Health_Insurance_Portability_and_Accountability_Act) |
| IMDRF | International Medical Device Regulators Forum |
| ITU | International Telecommunication Union |
| IVD | In vitro diagnostics |
| MDD | Medical device directives |
| MDR (2017/745) | Medical device regulation |
| SaMD | Software-as-a-medical device |
| SiMD | Software-in-a-medical device |
| WG | Working group |
| WHO | World Health Organization |

# Regulatory guidelines formulation: baseline references

The proposed guidelines were formulated based on a critical review of existing global regulations and standards for AI related technologies in medical applications. The critical review included identifying the gaps of existing regulatory requirements assessments methods and incorporating a quality risk management approach with necessary monitoring and control parameters for improved safety and efficiency of AI-MDs. A detailed list of regulatory references considered towards the formulation of the new guidelines is included in Annexes A to F

# Proposed roadmap for regulatory requirements guidelines

## Guidelines: applicability scope

For defining the applicability scope of the proposed guidelines, classification criteria based on a) scope of regulation, b) scope of product and c) scope of application are used. The classification criteria & scope of the proposed guidelines is illustrated in Figure 1.



Figure 1: AI-MD Classification Criteria & Scope

### Scope of regulation

Medical device:

* with enforcement of regulations
* without enforcement of regulations

### Scope of product

* Software is the product ->Standalone Software
* Software is embedded in product

### Scope of application

In healthcare

* To improve medical outcome (e.g. to support diagnosis, treatment, prevention, monitoring and prediction of diseases and injuries)
* To improve efficiency of healthcare system (e.g. streamline processes)

## Guidelines: aim & objectives

Aim

* to help manufacturers develop AI-based products conforming to the law and bring them to market quickly and safely
* to help internal and external auditors test the legal conformity of AI-based medical devices and the associated life-cycle process.

Objectives

The objective of this guideline is to provide target users with instructions and to provide them with a concrete checklist to

* understand what the expectations of the regulatory bodies are,
* to promote step-by-step implementation of safety and effectiveness of Artificial Intelligence / Machine Learning (AI/ML) Based Software as Medical Device,
* to compensate for the lack of a harmonized standard (in the interim) to the greatest extent possible.

## Guidelines: target user class / roles

The following user classes /roles are deemed responsible for using the guidelines

* Quality Assurance Auditors / Managers
* Developers
* Testers
* Data scientists
* Clinical specialists
* Physicians
* Product Managers
* Medical Device Consultants
* Regulatory specialists
* Risk assessment specialists
* Service and support providers

## Requirements assessment model: a product lifecycle development based approach



Figure 2(a): Regulatory roadmap-AI-medical device scope

Figure 2(a) shows the generic as well as the AI specific aspects that need to be considered under the regulatory roadmap of medical devices. From Figure 2(a), it can be inferred that AI-MD, as continuous learning or adaptive systems, are subject to modifications throughout its lifecycle and this result in unforeseen outcomes for the device including change of core device functionality and risk levels. These aspects pose additional challenges to the device manufacturers in terms of managing rapid development cycles, frequent software update and distribution cycles. Hence change management considerations tailored for AI-MDs are expected to have appropriate level of controls to manage these changes



Figure 2(b): Regulatory roadmap- AI4H project deliverables scope

Figure 2(b) shows the relevant AI-MD specific deliverables produced as part of the AI4H FG project. It can be seen that these AI4H deliverables include the necessary product development life-cycle processes that support the regulatory roadmap scope for AI-MDs. Document identifiers of AI4H deliverables are listed in Table-G- AI4H Project Deliverable Reference ID (Annex G) for further reference.

To facilitate continual product improvement in an iterative and adaptive manner with conformance to appropriate standards and regulations, it becomes imperative for any regulatory framework to establish a system that can ensure transparency and accountability of all the life cycle processes involved in AI-MD development. A brief rationale is provided here on the need for a product lifecycle development process-oriented approach that forms the basis of the proposed regulatory requirements guidelines



Figure 3: Product development life-cycle process (V-model)



Figure 4: AI Software life-cycle diagram

Figure 3 shows the V-model, which is widely accepted as the default product development lifecycle model in software engineering practice.

* A V- model based regulatory roadmap is proposed with an aim to maximize the completeness and coverage of various regulatory needs / aspects across the AI-MD life cycle processes -requirements, design, development, testing, deployment, maintenance, etc.
* The V- model supported by the principles of transparency and real-world performance monitoring, conformance assessments can be performed to measure and trace the compliance / deviation of in-house processes with standardized regulatory assessment procedures
* Apart from compliance verification, V-model gives thrust to software process improvement and supports integration of best practice for process improvement to achieve improved software quality, performance, safety, and effectiveness of medical device.

# Regulatory guidelines: proposed structure

Figure 5 shows the proposed guidelines structure. This guideline follows a process-oriented approach, whereby all relevant processes and phases of the AI-MD life cycle are considered such as:

1. Research and development
2. Data management
3. Post-market surveillance

Accordingly, the guideline does not set forth specific requirements for the products, but for the **processes applicable to pre-market evaluation,** which includes clinical evaluation **and post-market evaluation**. The requirements scope includes the following:

1. General requirements
2. Requirements for product development
3. Intended use
4. Software requirement specification
5. Data management
6. Model development
7. Product development
8. Product release

Requirements for phases following development



Figure 5: Structure of the regulatory guidelines

# Regulatory guidelines: requirements checklist

A regulatory requirement assessment checklist is proposed as a standard assessment and reporting tool to aid regulatory auditing/‌review process. Checklist enlists an orderly set of verification and validation procedures on how to conduct a comprehensive review covering all relevant aspects of the quality assurance pipeline.

The **quality criteria for a checklist** item include the following:

* It is atomic (not a combination)
* It can be checked within seconds or maximum a few minutes
* The result is binary i.e. either 'Yes' or 'No'
* It clearly specifies the necessary evidence
* It is understandable and verifiable also for non-experts
* It has to match / prove the requirement

## Pre-market requirements

Table 1: Process requirements

| Requirement(s) | Checklist item(s) | Checklist examples | Product lifecycle phase | Standard(s) / Regulation(s)applicable |
| --- | --- | --- | --- | --- |
| The manufacturers should establish a quality management (QM) System that covers all life cycle phases. | * There is at least one SOP[[1]](#footnote-1) covering the design& development process including verification and validation
* There is /areSOP(s) covering the post-market surveillance and vigilance
* There is a SOP covering risk management
* There is a SOP covering Computerized Systems Validation (CSV)
* There is a SOP covering the data management (process)
* There is / are SOP(s) covering software delivery, service, installation, decommissioning
* There is a SOP covering customer communication including handling of customer complaints
 |  | Full Lifecycle | EU MDR (2017/745) Article 10.9 ISO 13485 e.g. clause7.1ISO 13485 clause4.1.6 |
| The manufacturer should compile all product specific plans as required by respective regulations. | * There is a product specific development plan (including verification and validation)
* There is a product specific post-market surveillance plan
* There is a product specific clinical evaluation plan
* There is a product specific documented risk management plan
 |  | Full Lifecycle | MDR (2017/745) Annex I (3)MDR (2017/745) Annex III (1.3)IEC 62304 clause5.1ISO 14971:2019 (4.2) |

Table 2: Competencies requirements

| Requirement(s) | Checklist item(s) | Checklist examples | Product lifecycle phase | Standard(s) / Regulation(s) applicable |
| --- | --- | --- | --- | --- |
| The manufacturer should identify the roles inside the scope of its QM system that is directly or indirectly concerned with AI. | * There is a list that specifies roles and responsibilities inside the manufacturer's organization involved in its product life cycle activities
* These roles include software developers, software testers, data scientists, experts of clinical evaluations, risk managers, usability engineers, domain experts
 | Examples for domain experts are physicians, clinicians, nurses, lab technicians, pharmacists etc.Additional roles may include the following:* Regulatory affairs and quality managers
* Product managers
* Medical device consultants
* Service technicians e.g. update, upgrade, configuration, installation, capturing audit logs, etc.
* Support staff
 | Intended Use and Stakeholder Requirements Specifications | ISO 13485 clause 5.5.1EU MDR (2017/745)Article 10.9 |
| The manufacturer should ensure the necessary competencies for each role inside the scope of its QM system that is directly or indirectly concerned with AI. | * There are documented competency requirements for each role.
* There is a documented procedure on user role training and allied training materials
* There are records that provide evidence that the competency requirements have been met.
 | Examples of competencies are related to* Education
* Knowledge
* Skills: Capability to perform a particular task

Examples for training records are* (self) tests
* Artefacts that result from practicing a particular skill e.g. documents
 | Intended Use and Stakeholder Requirements Specifications | ISO 13485 clause 6.2. |

### Intended use and stakeholder requirements

Table 3: Intended use requirements

| Requirement(s) | Checklist item(s) | Checklist examples | Product lifecycle phase | Standard(s) / Regulation(s) applicable |
| --- | --- | --- | --- | --- |
| The manufacturer should determine the medical purpose of the medical device. | There is documented specification of* Indication including disease or injury or physiological state
* Goal: e.g. diagnosis, treatment, monitoring, prevention, elevation and / or prognosis
 | The disease or injury is specified using ICD-10 codes (at least 3 digits)Increasing adherence is an example for improving treatment. | Intended use and stakeholder requirements specifications | MDR (2017/745) Annex II (1.1)ISO 14971:2019 clause 5.2 |
| The manufacturer should specify other positive impacts on health care |  | * Faster patient care e.g. treatment, diagnosis
* Reductions inworkload
* Reductions in costs of healthcare
 | Intended use and stakeholder requirements specifications | MEDDEV 2.7/1 rev. 4MDR (2017/745) Annex I (23.4) |
| The manufacturer should specify the target patients | There is a documented specification of* Demographics (e.g. age, sex)
* Contraindications
* Co-morbidities
 |  | Intended use and stakeholder requirements specifications | MDR (2017/745) Annex I (23.4)MDR (2017/745) Annex II (1.1)IEC 62366-1clause 5.1 |
| The manufacturer should specify the intended part of body or type of tissue the medical device shall interact with |  |  | Intended use and stakeholder requirements specifications | IEC 62366-1clause 5.1 |
| The manufacturer should specify the operating principle  | There is a description of the task the ML-model shall perform.There is a specification of the type machine learning | Typical tasks include* Segmentation
* Detection
* Decision support
* Recommendation
* Process automation
* Search (e.g. similarities)

Typical dimensions include:* Type of learning (supervised, unsupervised, semi-supervised, reinforcement)
* Time and type of learning (before placing on the market, during use, globally, per product instance, per hospital)
* Technical task (classification, regression, clustering, control)
 | Intended use and stakeholder requirements specifications | IEC 62366-1clause 5.1MDR (2017/745) Annex II (1.1) |

Table 4: Intended users and intended context of use requirements

| Requirement(s) | Checklist item(s) | Checklist examples | Product lifecycle phase | Standard(s) / Regulation(s) applicable |
| --- | --- | --- | --- | --- |
| The manufacturer should characterize theintended users. | * There is a list of intended primary and secondary users
* The characteristics and prerequisites that each user group has to fulfil are specified
 | User characteristics my include:* Education
* Experience in medical domain
* Technical skill knowledge
* Training to be accomplished
* Physical prerequisites and limitations (height, sight, disabilities)
* Intellectual and mental prerequisites and limitations
* Language skills
* Experience with product type or technology
* Cultural and social background
 | Intended use and stakeholder requirements specifications | MDR (2017/745) Annex I (5)MDR (2017/745) Annex II (1.1)IEC 62366-1clause 5.1 |
| The manufacturer should characterize the intended use environment | There is a documented specification of the* physical use environment
* social use environment
* work environment
 | The physical environment might include:* Brightness,
* Loudness e.g. alarms
* Temperature
* Contamination
* Visibility
* Humidity, moisture

The social environment may include:* Stress, mental workload
* Shift operation
* Number of people and frequently changing colleagues

The work environment may include* Typical dress code
* Wearing of gloves or other personal protection equipment
* Usage of tools
* Physical stress
 | Intended use and stakeholder requirements specifications | MDR (2017/745) Annex I (5)IEC 62366-1clause 5.1 |

Table 5: Stakeholder requirements

| Requirement(s) | Checklist item(s) | Checklist examples | Product lifecycle phase | Standard(s) / Regulation(s) applicable |
| --- | --- | --- | --- | --- |
| The manufacturer should operationalize the goals listed in the intended use with quantitative values for product | * There are documented user requirements
* There are documented quantitative performance requirements
 | Examples of user requirements* 95% of radiologists working with system detect the cancer

Examples of performance requirements:* The system shall have a sensitivity of 97%
* The system must be able to detect coronary artery plaques of at least 0.2 mm diameter
 | Intended Use and Stakeholder Requirements Specifications | MDR (2017/745)Annex I (23.4)MDR (2017/745) Annex III (1.1) |
| The manufacturer should specify the runtime environment of the product regarding hardware and software | * The minimum hardware requirements are specified
* The minimum software requirements are specified
 | Hardware requirements may include* Screen size, resolution and orientation
* Physical storage
* Network connectivity e.g. bandwidth, latency, reliability
* Required peripherals such as printers, scanners, input devices

Software requirements may include* Operating system (incl. version)
* Browser (type, version)
* Virtualization (e.g. Java Runtime Environment,.NET, Docker, virtual machines)
 | Intended Use and Stakeholder Requirements Specifications | ISO 13485 clauses 6.3 (infrastructure) and 7.5.1 (verification of infrastructure requirements for deployment)EU MDR (2017/745) Annex 1, 17.4 IEC 62304 clause 5.2 |
| The manufacturer should identify and specify the data interfaces. | * There is a list of data interfaces (can be specified in a context diagram as well)
* The protocols are specified
* The formats are specified
* The semantic standards are specified
 | Protocols might include* OSI-Protocols such as TCP/IP, HTTPS
* Bus-Systems such as CAN, USB
* Physical hardware connections

Format might include* File formats (XML, JSON, PDF, docx, CSV, DICOM)
* Image formats (size, resolution, colour coding)

Semantic standards might include* Taxonomies e.g. ICD-10, ATC
* Nomenclatures e.g. LOINC
 | Intended Use and Stakeholder Requirements Specifications | IEC 62304 clause 5.2 |
| The manufacturer should specify the requirements for input data for each inbound data interface. | There is a specification of input data  | Input data specifications may include:* Sensor requirements
* Type of data capturing device
* Precision of data
* Size / quantity of data
* Type and technical parameters of recording procedure (e.g. strength of magnetic field, number of electrodes,
 | Intended Use and Stakeholder Requirements Specifications | IEC 62304 clause 5.2ISO 14971:2019 clause 5.3 |
| The manufacturer should determine the regulatory requirements. | * There is a list of countries / markets that the product shall be place in.
* There is a list of laws, standards, regulations, directives, guidance
 | The list might include documents such as* FDA guidance documents
* Standards (e.g. IEC 62304, ISO 13485)
* Laws and regulations e.g. MDR (2017/745), IVDR
 | Intended Use and Stakeholder Requirements Specifications | MDR Annex IX (2.2)ISO 13485 (clauses 5.2 and 7.2.1) |

Table 6: Inputs to risk management and clinical evaluation requirements

| Requirement(s) | Checklist item(s) | Checklist examples | Product lifecycle phase | Standard(s) / Regulation(s) applicable |
| --- | --- | --- | --- | --- |
| The manufacturer should evaluate alternatives to the given product (e.g. other products, procedures, technologies).  | * There is a clinical evaluation
* The clinical evaluation lists alternative products, technologies and / or procedures
* There is a search protocol revealing how the manufacturer was searching for alternatives
* The clinical evaluation assessing alternatives with respect to clinical benefits, safety / risks, performance
* The alternatives include non-ML based technologies
* There is a statement confirming that the product reflects the state-of-the-art.
 | Alternative technologies might include* Other ML models
* Non-ML methods e.g. classical algorithms

Performance parameters might include* Accuracy
* Specificity
* Sensitivity
* Response times
* Robustness
 | Risk Management and Clinical Evaluation | MEDDEV 2.7/1 MDR (2017/745)ISO 14971:2019 clauses 4.2 and 10.9 |
| The manufacturer should compile a list of risks specifically associated with the use of the method of machine learning. | * The risk management file contains an analysis of hazards and related harms with related probabilities and severities resulting from ML models not meeting the requirements.
* There is a FMEA that analyses the effects of ML models that do not meet the performance requirements
 | Performance requirements might include* Accuracy
* Specificity
* Sensitivity
* Response times
* Robustness
 | Risk Management and Clinical Evaluation | ISO 14971:2019 clauses5.4 and 5.5EU MDR (2017/745) Annex 1 (3) |
| The manufacturer should analyse the reasonably foreseeable risks. | The risk management file analyses risks associated with* non-specified user type / category
* non-specified use environment
* application of product for patients other than those specified
* reasonably foreseeable misuse
 | Non-specified users* other profession e.g. nurse instead of physician
* missing training

Other patients* Different age, sex, race
* Other co-morbidities
* Different severity of disease or injury
 | Risk Management and Clinical Evaluation | ISO 14971:2019 clause 5.2 |

### Product and software requirements

Table 7: Functionality and performance requirements

| Requirement(s) | Checklist item(s) | Checklist examples | Product lifecycle phase | Standard(s) / Regulation(s) applicable |
| --- | --- | --- | --- | --- |
| The manufacturer should derive from the stakeholder requirements traceable quantitative quality criteria and requirements for the software and/or the algorithm from the intended use | There is a specification of quantitative minimum 'quality criteria'There is 'traceability matrix' that links the intended use with quantitative quality product requirements. | 'Quantitative quality criteria' may include the following:* for classification problems
* accuracy (mean or balanced accuracy)
* positive predictive value (precision)
* specificity and sensitivity
* F1 Score [Area under the ROC curve (AUC)](https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc)
* for regression problems
* mean absolute error
* mean square error

 **Example 1**: The stakeholder requirement states that 95% of radiologists must be able to detect a cancer with the product. The requirement of the algorithm states that it must display a sensitivity of 97%. **Example 2**: The stakeholder requirements state that arterial calcification must be able to be detected at a sensitivity of 92%. The requirements of the algorithm state that it must be able to exactly predict the strength of the plaques in the blood to 0.2 mm. | Product and Software Requirements Specifications | IEC 62304 clause 5.2ISO 13485 7.3.3 |
| The manufacturer should derive non-functional requirements from the intended use. | There is a specification of non-functional requirements such as* Repeatability / Reproducibility
* Response times
* Availability
 | Self-tests can be a mean to verify the repeatability of a system.The specification of response times might depend on number of users, number of transactions, frequency and amount of input data etc.Availability can be expressed as percentage of time, percentage of usages or as meantime between failure. | Product and Software Requirements Specifications | ISO 13485 clause 7.2.1MDR Annex I (17) |
| The manufacturer should determine how the system behaves if the inputs do not meet the specified requirements | There is a specification that describes how the system reacts on* incomplete data sets
* lack of data sets
* wrong data format
* excessive data quantities
* data outside of specified value ranges
* wrong temporal sequence of data, etc.
 |  | Product and Software Requirements Specifications | ISO 25010IEC 62304 clause 5.2ISO 14971:2019 clause 5.4 |

Table 8: Userinterface requirements

| Requirement(s) | Checklist item(s) | Checklist examples | Product lifecycle phase | Standard(s) / Regulation(s) applicable |
| --- | --- | --- | --- | --- |
| The manufacturer should specify what the user interface must display in case of errors. if the inputs do not meet the specified requirements | There is specification of the user interface in case of* incomplete data sets
* Internal errors
 | See previous checklist item.UI output display modes may include the following:* warning
* alert
* caution,
* Meantime between failure, etc.
 | Product and Software Requirements Specifications | IEC 62304 clause 5.2 |
| The manufacturer should determine whether there is a need for instructions for use and training materials. | Either there is an Instructions-for-Use (IFU) or the user riskanalysis reveals no risks that can be further mitigated by an IFU |  | Product and Software Requirements Specifications | MDR (2017/745) Annex I (23) |

Table 9: Additional software requirements

| Requirement(s) | Checklist item(s) | Checklist examples | Product lifecycle phase | Standard(s) / Regulation(s) applicable |
| --- | --- | --- | --- | --- |
| The manufacturer should set forth requirements to detect internal errors | The risk analysis considers risk that are caused by internal errors.The device specification specifies how manufacturers or service technicians can gain access to internal errors. | Examples of interfaces include:* Data and user interfaces to audit logs
* Monitoring ports

Examples of internal errors are* Runtime errors such as null pointer exception
* Resource overload such as out of memory errors
* Lack of access to resources such as databases
* Compromised integrity of data and program code
 | Product and Software Requirements Specifications | MDR (2017/745) Annex I (17, 18, 23.4)IEC 62304 clauses 5.2, 5.3 and 7.1ISO 149781:2019 clause 5.4 |
| The manufacturer should justify if the device takes decisions exclusively based on automatic data processing | * There are records of processing activities.
* There is a data protection impact assessment
 |  | Product and Software Requirements Specifications | Art. 22 of the GDPR.Exceptions of Art. 22 section 2 may apply. |

Table 10: Risk management & clinical evaluation requirements

| Requirement(s) | Checklist item(s) | Checklist examples | Product lifecycle phase | Standard(s) / Regulation(s) applicable |
| --- | --- | --- | --- | --- |
| The manufacturer should assess the risks arising if the inputs do not meet the specified requirements | The risk analysis assesses the risks for wrong inputs at each data interface.The risk analysis considers all relevant types of wrong inputs. | Invalid/ non-compliant input conditions may include the following:* incomplete data sets
* lack of data sets
* wrong data format
* excessive data quantities
* data outside of specified value ranges
 | Product and Software Requirements Specifications | ISO 14971:2019 clause 5.4IEC 62304 clause 7.1 |
| The manufacturer should set the gold standard against which the quality criteria can be reviewed, and justify their choice | The clinical evaluation lists alternatives.The clinical evaluation compares these alternatives with respect to specified quality criteria.There is a documented justification for the selected ground truth. | The gold standard is not the same as alternatives. E.g. the gold standard to determine the blood pressure is an invasive measurement. But this is not the alternative. | Product and Software Requirements Specifications |  |
| The manufacturer should analyse the risks arising if the outputs do not meet the specified quality criteria | There is risk assessment report / risk table that specifies risks in case outputs do not meet the specified 'quantitative quality criteria' |  | Product and Software Requirements Specifications | ISO 14971:2019 clause 5.4IEC 62304 clause 7.1 |
| The manufacturer should assess the consequences if the system provides socially unacceptable / discriminatory outputs | There is of outputs that an assessment report on consequences / implications of socially unacceptable outputs.Assessment report includes:* Cost estimation for wrong clinical decision making
* AI autonomy level assignment and associated risk acceptance criteria based on criticality of the clinical use case and environment
 |  | Product and Software Requirements Specifications | TODO (EU framework on AI) |
| The manufacturer should assess the risk arising if the system does not meet the specified non-functional requirements | The risk analysis assesses risk arising from* Lack of availability / robustness
* Slow response times
* Interoperability problems
 |  | Product and Software Requirements Specifications | IEC 62304 clause 7.1 |
| With Continuous Learning Systems, the manufacturer should mitigate risks that are specific to continuously learning systems.  | The risk analysis assesses risks that a specific to continuous learning systemsThe risk management file specifies the respective risk mitigation. | Examples of risk mitigation:* option to reset the systems
* self-tests
 | Product and Software Requirements Specifications | MDR (2017/745) Annex I (17) |
| With Continuous Learning Systems, the manufacturer should show quantitatively why the risk-benefit analysis is better than for non-continuously learning systems | Analysis report showing a Positive Risk-Benefit Ratio compared to the state-of-the art. The clinical evaluation compares benefits for continuously learning and non-continuously learning systems. |  | Product and Software Requirements Specifications | ISO 14971:2019 clause 6 |

### Data management requirements

Table 11: Data collection requirements

| Requirement(s) | Checklist item(s) | Checklist examples | Product lifecycle phase | Standards / Regulations applicable |
| --- | --- | --- | --- | --- |
| The manufacturer should specify the number of test data sets. | * There is a specification of number of data sets.
* There is a rationale for this number.
 |  |  | ISO 13485 clause 7.3.7 |
| The manufacturer should specify the inclusion and exclusion criteria for individual data sets | * There is a specification of technical requirements
* There is a specification of patient attributes that have to be met to include a data set.
 | Technical inclusion / exclusion criteria may include for each attribute* Data ranges
* Data type (numeric (float, integer etc.), ordinal, categorical, String / text, date / time, image / binary)
* Data formats (e.g. date and number formats)
* Unit of measure
* Precision of numbers
* Attributes values
* File formats / types
* Sampling rates
* Image parameters such as compression, images size, resolution, colour coding, zoom

Inclusion / Exclusion criteria of patient data may include the following attributes:* demographic data (age, gender)
* physical parameters (height, weight)
* diseases
* vital parameters
* lab parameters
* presence of additional tests
* case history
* Special conditions (e.g. patients having heart pacemaker or lung surgery)
 | Data Management | TODO: Extract from AI/ML standards |
| The manufacturer should specify the data source requirements | * There is a list of allowed / expected data sources
* There is a specification of data source requirements
* There is a validation of surveys (justify the selection of the surveys, the time of survey and possibly the method for their assessment, in particular if no standardized survey exists.)
* Survey methods may include the type of questions, the types of answers, the decision to have open or closed questions etc
 | Data sources may include* Medical devices
* In-vitro diagnostic devices
* Questionnaires
* Cameras
* Electronic patient records

Examples for input requirements:* with or without contrast agent (MRT, CT)
* number of electrodes (ECG)
* Voltage (X-Ray, CT)
* Position of patient
 |  | TODO: Extract from AI/ML standards |
| The manufacturer should specify the distribution of input data | * There is an analysis of factors causing a bias
* There is a specification of the distribution of relevant patient characteristics
 | Factors causing biases include:* Non representative patient population
* Attributes that are irrelevant for the expected output
* Confusion of correlation and causation
* Specific data sources, type and location of data collection

Even if all individual data sets meet the specification, still the distribution of data might not be representative and/or cause a bias. |  | TODO: Extract from AI/ML standards |
| The manufacturer should validate that the test and training data meet the specified criteria | * There is a description how it is ensured that data sets that do not meet the inclusion criteria are actually excluded
* there is a descriptive statistic
* there is a justification that the data are representative for the target population
* there is an analysis of a potential "label leakage"
 | Descriptive statistic may include the following:* calculation of distributions (histograms),
* mean / average values,
* quartiles,
* joint distribution of features, correlation, etc.

Label leakage examples include:* in the sorting (e.g. first the data of healthy persons, then of ill persons),
* in the hospital (e.g. if the severe cases originate from just one institution),
* in images (e.g. for skin cancer, one must always see a ruler)
 |  | TODO: Extract from AI/ML standards |
| The manufacturer should ensure data protection.  | * There is a documented patient data protection policy
* There should be a documented procedure for data anonymization / pseudonymization
* Data scientists do not have access to protected data
* There is a data protection officer
 | Data could be derived from machine-to-machine (M2M) communication as well |  | GDPR |

Table 12: Data annotation requirements

| Requirement(s) | Checklist item(s) | Checklist examples | Product lifecycle phase | Standard(s) / Regulation(s) applicable |
| --- | --- | --- | --- | --- |
| The manufacturer using "supervised learning" should derive the labels from the intended use and justified this selection | There is specification for "Label"selection criteria in case of "supervised learning "based machine learning task |  | Data Management |  |
| The manufacturer using "supervised learning" should has a procedure to ensure correct labelling | * The procedure describes how the ground truth is derived
* The procedure specifies quantitative classification / segmentation criteria for labelling
* There is a justification of these criteria
* The procedure specifies how and how frequently the correctness of labelling is monitored
* The procedure specifies how to deal with inconsistency of data annotation from multi-annotators
 | If, for example, patients have to be classified as healthy and sick, the manufacturer must derive the criteria specifically for the intended use, when a patient is to be classified as healthy and when as sick. |  | ISO 13485 clause 4.1 |
| The manufacturer should ensure the competency of persons responsible for labelling | * There is specification for the number of people recruited for "labelling" task
* There is description of the training to be given to persons responsible for 'labelling'
* There is specification for the competency level of persons responsible for 'labelling'
* There is a procedure for assessing the success of training success and of the competency for persons responsible for 'labelling'
* There are respective records
 | The results of the monitoring of the labelling can be used to continuously verify the fitness of persons responsible for labelling | Data Management | ISO 13485 clause 6.2 and 7.3.2 |

Table 13: Data pre-processing requirements

| Requirement(s) | Checklist item(s) | Checklist examples | Product lifecycle phase | Standard(s) / Regulation(s) applicable |
| --- | --- | --- | --- | --- |
| The manufacturer should set a procedure that describes the pre-processing of the data before data is used to train or test the model | There is documented procedure for data pre-processing:* This procedure describes how the correctness of the interim steps and the final results are assessed through risk-based evaluations
* This procedure specifies how values with various measurement scales or units are detected and processed
* This procedure specifies how values are detected and processed that have been collected with various measurement methods
* This procedure specifies how missing values within data sets are detected and processed
* This procedure specifies how unusable data sets are detected and handled as per the data inclusion and exclusion criteria
 | Data pre-processing stepsmay include the following:* conversion,
* transformation,
* aggregation,
* normalization,
* format conversion,
* calculation of feature,
* conversion of numerical data into categories, etc.

"Missing value"problem includes"missing at random" and "missing not at random""Missing value"processing techniques include:* deleting the data set
* replacement by the average value of other data sets
* new value "missing" (for categorical values), etc.

"Outliers" processing techniques include* deleting the data set
* correcting the value
* setting the value to a set value (min/max),etc.

Examples of unusable datasets may include:* x-rays of poor quality as specified in the technical exclusion criteria or patients/‌persons who do not meet the patient inclusion criteria. etc.
 | Data Management | ISO 13485:2016 clause4.1 |

Table 14: Documentation and version control requirements

| Requirement(s) | Checklist item(s) | Checklist examples | Product lifecycle phase | Standard(s) / Regulation(s) applicable |
| --- | --- | --- | --- | --- |
| The manufacturer should trace data back to data source.  | * There is a list of data sources
* There is a document describing the data processing steps
* There is a specification of rules for data inclusion and exclusion
* There is a rationale if additional data have been excluded or if data have been kept despite meeting the specification
 | The description of data sources might include* Location (e.g. clinic)
* Capture device
 | Data Management |  |
| The manufacturer should document all software for data processing  | * There is a list of all software applications
* All applications are clearly identified
* It is identifiable if the software is off-the shelf or individually developed
 | Means to identify a software are* Manufacturer
* Name of software
* Version of software
 | Data Management | ISO 13485 clause 4.1 |
| The manufacturer should put all software under version control | * There is a policy (e.g. SOP) specifying the configuration and version control process
* There are records demonstrating that the software actually is under version control
* The software libraries and frameworks are identified and under version control
 |  |  | IEC 62304 clause 8 |

### Model development requirements

Table 15: Model preparation requirements

| Requirement(s) | Checklist item(s) | Checklist examples | Product lifecycle phase | Standard(s) / Regulation(s) applicable |
| --- | --- | --- | --- | --- |
| The manufacturer should deliberately select the features for training | * There is a list of features
* There is a rationale as to why a feature is taken into account
* There is a description of the dependency of the features among each other
 | Dependencies can be visualized e.g. with a directed Acyclic Graph (DAG)  | Model Development | TODO: Extract from AI/ML standards |
| The manufacturer should deliberately divide the data into training, validation and test data | * There is justification for the ratio of training, validation and test data
* There is a documented stratification for dividing up the data in to training, validation and test data
* There is documentation that reveals how multiple data sets for an object are in the same "bucket" (training, validation and test data).
* There is a justification if data are not distributed at random
 | E.g. for data with rare features or labels, it may be necessary to distribute the data not just at randomExample for an object can be a CT scan. The images of one series should not be distributed into the three different "buckets". |  | TODO: Extract from AI/ML standards |
| The manufacturer should document how it ensures that the development team has no access to the test data | * There is a role-based policy for data access
* There is a description how the development team is prevented from gaining access to test data
 |  | Model Development | TODO: Extract from AI/ML standards |

Table 16: Model training requirements

| Requirement(s) | Checklist item(s) | Checklist examples | Product lifecycle phase | Standard(s) / Regulation(s) applicable |
| --- | --- | --- | --- | --- |
| The manufacturer should document model specific data processing | * There is a document that describes which feature have been recoded specifically for a model or technology
 | Examples of this are normalization, selection of class labels (e.g. 0 or 1), selection of column names, distribution of categorical values over multiple columns. |  | ISO 13485 clause 4.1 |
| If there are several quality metrics, the manufacturer should document the quality metrics for the model to which it wants to optimize the model and justified it based on the intended use. | * There are one or more quality metrics identified and respective target values specified
* There is a documented rationale how these quality metrics relate to the intended use.
 |  |  | TODO: Extract from AI/ML standards |
| The manufacturer should avoid overfitting | * There is a policy forbidding the use of test data to optimize the model (only training and test data may be used)
* There is a justification of the choice of the hyperparameters
* There is a justification of the choice of epochs
 | Visualization (e.g. learning curves) might be helpful for justification and to illustrate the impact of hyperparameter and epochs on quality metrics |  | TODO: Extract from AI/ML standards |

Table 17:Model evaluation requirements

| Requirement(s) | Checklist item(s) | Checklist examples | Product lifecycle phase | Standards / Regulations applicable |
| --- | --- | --- | --- | --- |
| The manufacturer should evaluate the correctness and robustness of the model | * There is model validation report
* There is a test specification and test results for the final evaluation of the model with new test data
* There are documented values for specified quality metrics
* There is an analysis of datasets that have been predicted particularly well and not well.
* There is an analysis of datasets that have been predicted securely and particularly insecurely
* There is an evaluation on how and how strongly individual features had to change for the model to come to another prediction.
* For individual data sets there MAY be an evaluation of the feature that the model particularly determined in the decision
* There MAY be an analysis/visualization of the dependency (strength, direction) of the prediction of the feature values
* There MAY be a synthetization of data sets that activate the model particularly strong
* There MAY be an approximation of the model using a simplified surrogate model
 | A residual analysis in which the errors are listed via the feature values.For classification tasks, the model is particularly insecure with probabilities around 0.5.This is referred to as "Counterfactuals". This, however, depends on the ML method and cannot be demanded as a general best practice.Approaches include LIME (Local Interpretable Model-agnostic Explanations), Beta (Black Box Explanations through Transparent Approximations), LRP (Layer-wise Relevance Propagation) and Feature Summary Statistics (incl. Feature Importance and Feature Interaction). This, however, depends on the ML method and cannot be demanded as a general best practice.Examples of Sharpley-Values, ICE-Plots, Partial Dependency Plots (PDP). This, however, depends on the ML method and cannot be demanded as a general best practice.Examples of synthetization can be found here <http://yosinski.com/deepvis>. This, however, depends on the ML method and cannot be demanded as a general best practice.A decision tree is an example for a surrogate model. This, however, depends on the ML method and cannot be demanded as a general best practice. |  | MDR (2017/745) Annex I (17), Annex II (6.1)IEC 62304 clauses 5.5ff.ISO 13485 clauses 7.3.4 ff. |
| The manufacturer should justify its ultimate selection of the model (architecture, hyperparameter) | There is a documentation of various models that have been compared.There is a comparison of these models (architectures)There is a comparison of different sets of hyperparameterThe comparison includes quality metricsThe comparison not only assesses the quality metrics globally, but also separately for various features.There is a risk-benefit assessment that discusses interpretability, performance (e.g. quality metrics, efficiency) and robustness | Example for quality metrics see above. |  |  |

Table 18: Model documentation requirements

| Requirement(s) | Checklist item(s) | Checklist examples | Product lifecycle phase | Standards / Regulations applicable |
| --- | --- | --- | --- | --- |
| The manufacturer should document the model | * There is a documentation of the model (architecture)
* There is a documentation of the selected hyperparameters
* There is a documentation of used software libraries and frameworks (also SOUPs)
* There is a documentation of the quality metrics and of the evaluation results e.g. of performance and robustness as specified in Table 17: Model evaluation requirements
 | A way to document models are the 'model card / sheet'that includes:* model version
* assumptions, constraints, dependencies on the algorithm used
* current performance figures
* expected / optimal performance
* major risk conditions

ML models included* Linear Regression
* Logistic Regression
* k-nearest neighbours
* Decision Trees
* Random Forest
* Gradient Boosting Machines
* XGBoost
* Support Vector Machines (SVM)
* Neural Network
* k means clustering
* Hierarchical clustering
* Neural Network including Convolutional Neural Network (CNN), Recurrent Neural Networks (RNNs) and Long Short-Term Memory Networks (LSTMs)
* Apriori algorithm
* Eclat algorithm
* Stacked Auto-Encoders
* Deep Boltzmann Machine (DBM)
* Deep Belief Networks (DBN), etc.
 |  | TODO: Extract from AI/ML standards |
| The manufacturer should apply version and configuration control to development artefacts. | * There is a SOP for document respectively version and configuration control
* The following artifacts are (additionally to software code and libraries) under version control:
* Configuration files, hyperparameters
* Test and evaluation results (including quality metrics)
* Software libraries and frameworks
 | e.g. trained models can be serialized | Model Development | ISO 13485 clauses 7.3.10, 7.5.9.1 |

### Product development requirements

Table 19: Software development requirements

| Requirement(s) | Checklist item(s) | Checklist examples | Product lifecycle phase | Standard(s) / Regulation(s) applicable |
| --- | --- | --- | --- | --- |
| The manufacturer should perform and document the required activities pursuant to IEC 62304. | * There is software development plan
* If the model is implemented in another programming language or for another runtime environment, the plan defines which activities of model development have to be repeated
* There is a verification plan that requires software system tests
* The software safety class (alternatively Level of Concern) is determined
* There is a software requirement specification (SRS)
* The SRS specifies user interface related requirements
* There is a documented software architecture
 | * adhere to the normal best practices such as adherence to coding guidelines,
* review of code by code reviews using defined criteria,
* testing to code with unit tests with a defined coverage, etc.
 | Product Development | IEC 62304 |
| The manufacturer should test software on the target hardware | * The test hardware is specified
* The test hardware is representative for the target hardware
* The tests verify whether the specified performance requirements are met
 | Performance may include:* response times,
* resource consumption

Hardware may include* browser,
* mobile device, etc.
 | Product Development | MDR (2017/745) Annex II 6.1 |
| The manufacturer should identify and verify all SOUP (respectively OTS) components | * there is a list of all SOUP / OTS components
* Each SOUP / OTS component is uniquely identified
* Each SOUP / OTS component is under version control
* The requirements for each SOUP / OTS component are specified
* The is a documented trace between these requirements and respective tests
* The prerequisites for each SOUP /OTS component are specified
 | Components can be uniquely identified by* Manufacturer
* Name of component
* Version of component

Traces can be documented using ALM tools or tablesExamples for prerequisites are* Hardware (e.g. processor architecture, RAM)
* Software (e.g. operating system, run-time environments e.g. .NET, Browser)
 | Product Development | IEC 62304 clauses 5.3 and 8.1.2 |

Table 20: Accompanying materials requirements

| Requirement(s) | Checklist item(s) | Checklist examples | Product lifecycle phase | Standard(s) / Regulation(s) applicable |
| --- | --- | --- | --- | --- |
| The manufacturer should provide instructions for use | * There are instructions for use
* The instructions for use clearly identify the version of the product
* There is a procedure specifying how to develop and verify instructions for use
* The instructions for use and under version control
 | The identification of the product should be achieved by the product’s UDI-DI |  | MDR (2017/745) Annex I (23.4) |
| The instructions for use should describe the intended purpose and intended use | * The instructions for use specify the intended medical purpose and medical benefit
* The instructions for use specify the intended patient population including indications, contraindications and if relevant other parameters
* The instructions for use specify the patients / data / use case for which the product may not be used.
* The instructions for use specify the requirements of the input data
* The instructions for use specify the intended primary and secondary users pursuant to intended use.
* The instructions for use describe the other conditions applicable to the product (e.g. runtime environment, use environment).
* The instructions for use describe how to update the product
 | The medical purpose and benefit typically are related to diagnosis, treatment, prognosis and monitoring of certain diseases or injuries.The patient population can be characterized by age, gender or accompanying diseasesExamples for input data requirements are* Formats
* Resolutions
* value ranges, etc.
 |  | MDR (2017/745) Annex I (23.4) |
| The instructions for use should specific the performance of the product | * The instructions for use specify the quality metrics.
 | Examples of quality metrics are specificity, sensitivity, precision |  | MDR (2017/745) Annex I (23.4) |
| The instructions for use should explain the product and its inner workings. | * The instructions for use indicate the data with which the model was trained.
* The instructions for use describe the model and algorithms.
* The instructions for use specify whether the product is further trained during use.
 |  |  | MDR (2017/745) Annex I (23.4)?TODO: Reference to AI/ML standards |
| The instructions for use should reveal residual risks | * The instructions for use list the factors that could have a negative effect on the product’s performance
* The instructions explain risks arising from a product not meeting the performance requirements
* The instructions for use list possible ethical problems.
 | Example for negative factors are* Patient population deviating from specified population
* Data not meeting the specified criteria (e.g. formats, value ranges)
 | Product Development | EU MDR (2017/745) Annex 1 23.4 ISO 14971:2019 clause 8 |
| The instructions for use should further information that is legally required | * The instructions for use identify the manufacturer
* The instructions for use lists channels for posing questions
* The instructions for use contain references to licensing rights.
* The instructions for use contain the URL under which the most current versions of the instruction of use can be found
 |  |  | EU MDR (2017/745) Annex 1 23.4EU Regulation 207/2012 |

### Product validation requirements

Table 21: Usability validation requirements

| Requirement(s) | Checklist item(s) | Checklist examples | Product lifecycle phase | Standard(s) / Regulation(s) applicable |
| --- | --- | --- | --- | --- |
| The manufacturer should identify risk arising from a lack of usability | * The risk management file lists risks, that arise from misunderstanding, overlooking or ignoring the product’s visual output
* The risk management file lists risks, that arise from users blindly trusting or mistrusting the product.
 | The product’s visual output includes* Displaying information at the user interface
* Reports, printouts

The manufacturer could evaluate how obvious the systems output is before users become suspicious. |  | IEC 62366-1 clause 4.1 |
| The manufacturer should assess whether the users understand the instructions for use. | * The risk management file lists the risks, that have to be mitigated by instructing users e.g. by training or accompanying materials.
* The plan of the summative evaluation describes how the effectiveness of these measures is validated.
* The usability evaluation report reveals whether the instructions for use areadequate to mitigate risks.
 |  | Product Validation | IEC 62366-1 (instructions for use are considered to be part of accompanying documentation that is considered to be part of the user interface) |
| The manufacturer should evaluate all safety relevant use scenarios. | * There is a list of use scenarios.
* There is an assessment of safety relevance for each use scenario
* The use scenarios included in the summative evaluation cover all safety relevant use scenarios.
* The summative evaluation evaluates the effectiveness cover all risk mitigation measure.
 |  |  | IEC 62366-1 clause 5.4 |

Table 22: Clinical evaluation requirements

| Requirement(s) | Checklist item(s) | Checklist examples | Product lifecycle phase | Standard(s) / Regulation(s) applicable |
| --- | --- | --- | --- | --- |
| The manufacturer should assess whether the promised medical benefit is achieved with the quality parameters. | * The clinical evaluation contains the medical benefits the manufacturer claims.
* The clinical evaluation lists the data (sources) that have been evaluated and that support and that contradict the hypothesis, that the benefits have been achieved.
* If the data have been collected from other products, than the clinical evaluation discusses the clinical and technical equivalence of the other products.
* The clinical evaluation evaluates the impact of quality parameters on the achievement of the medical benefit.
 | The data are typically clinical dataThe technical equivalence has to consider the software algorithms. | Product Validation | EU MDR (2017/745) Article 61 |
| The manufacturer should assess whether the promised medical benefit is achieved is consistent with the state of the art. | * The clinical evaluation lists alternative methods, technologies or procedures
* The clinical evaluation compares the risks and benefits of these alternatives
 | Alternative approaches include * Not continuously learning model in comparison to a continuously learning model
* Classic algorithm in comparison to a machine learning model.
 | Product Validation | EU MDR (2017/745) Article 61 |

### Product release requirements

Table 23: Product release requirements

| Requirement(s) | Checklist item(s) | Checklist examples | Product lifecycle phase | Standard(s) / Regulation(s) applicable |
| --- | --- | --- | --- | --- |
| The manufacturer should verify the completeness of documentation. | * There is a risk management report concluding that all risk management related activities have been performed according to risk management plan and that residual risks are acceptable.
* There is a usability evaluation report concluding that all activities to formative and summative evaluation plan have been performed.
* There is a documentation of the model.
 | The documentation of the model should at least cover all aspects that have been mention in chapter “instructions of use”. |  | MDR (2017/745) Annexes I and II |
| If the manufacturer plans to market its product in the US market it should compile the respective documentation. | * There is a "Software as a Medical Device Pre-Specifications “(SPS) that anticipates changes to the product.
* There is an “Algorithm Change Protocol (ACP)”that specifies howthese changes for systems will be performed
 |  | Product Release | FDA: Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) |

## Post-market requirements

### Production, distribution & installation requirements

Table 24: Production, distribution and installation requirements

| Requirement(s) | Checklist item(s) | Checklist examples | Product lifecycle phase | Standard(s) / Regulation(s) applicable |
| --- | --- | --- | --- | --- |
| The manufacturer should apply version- and configuration control | * There is a SOP or work instruction that specifies how the manufacturer identifies artefacts and how it ensures how the correct artefacts in the respective version are delivered
* Version and configuration applies to the software as well to accompanying materials such as instructions for installation and use
* There is a bill of materials
* There is a unique identification (ID) of the product
 | * The bill of material also contains all SOUP/OTS Software
* In EU and in the US, there is typically the need for a UID-DI and UDI-PI
 |  | IEC 62304 clause 8FDA Cybersecurity GuidanceISO 13485:7.5.8 |
| The manufacturer should ensure the design transfer | * There is a SOP or work instruction that specifies how the persons responsible for installation know which is the most current version and how mistakes in installation can be ruled out
* There is are instructions for installation, update and decommissioning
* These instructions specify the runtime environment
* There instructions specify how the correct installation can be verified
 | The specification of the production runtime environment can include* Hardware (CPU, RAM)
* Monitors, displays (size, resolution, orientation)
* Operating system
 |  | ISO 13485 clause 7.3.8 |
| The manufacturer should ensure effective and efficient communication with operators and users.  | * There is a SOP covering customer communication including handling of customer complaints
* There is a website that contains information about latest product releases
* The website provides the means to download the software
* The instructions for use reference this website
* The instructions for use and the website reveal contact information e.g. e-mail, phone number, and/or a contact form
 |  | Production, Distribution, Installation | ISO 13485 clauses 5.2 and 7.2MDR (2017/745) Annex I (23.1) |

### Post-market surveillance requirements

Table 25:Post-market surveillance requirements

| Requirement(s) | Checklist item(s) | Checklist examples | Product lifecycle phase | Standard(s) / Regulation(s) applicable |
| --- | --- | --- | --- | --- |
| When determining threshold values the manufacturer should analyse how the application of the product might impact feature (input values)  | * There is an analysis, whether feedback loops can influence input values
* There is an analysis, whether self-fulfilling prophecies can influence input values.
* There is a specification of threshold values in the post-market surveillance plan
 | * Example for feedback loops: A travel recommendation app sends targeted advertising depending on feature (last trip). This influences travel behavior.
* Example for feedback loop: An algorithm provides prognoses. Therefore, the physician will treat the patients better or earlier...
* Example for self-fulfilling prophecies: an algorithm for predicting date and location of crimes will cause a higher surveillance by police. This will cause an increased number of detected crimes
 | Post-Market Surveillance | MDR (2017/745) Annex III (1.1) |
| The manufacturer should compile a Post-Market Surveillance Plan | * There is a SOP specifying how to compile post-market surveillance plans
* There is a post-market surveillance plan specifically for the product
* The plan lists all data sources to be monitored
* The data sources include scientific literature, customer communication (e.g. complaints), IT security databases, bug reports and release notes for SOUP / OTS, databases of authorities (e.g. FDA)
* The plan describes for each data source how, how often and by whom data are collected
* The plan specifies how data has to be analysed
* The plan requires that quality metrics such as sensitivity and specificity are monitored
* The plan specifies the data to be collected to be able to analyze whether the data in the field is consistent with the expected data or training data
* The plan lists threshold values that trigger actions
* The threshold values include quality metrics
* These threshold values include features
* The plan specifies the frequency and content of compiling post-market surveillance reports
* The plan is approved
 | * “By whom” not only persons / roles, but also systems can be listed
* Examples for additional quality metrics see above. Also, the variance of these quality metrics over time might be a quality metric (This allows visualization or quantification in particular for non-normally distributed data over the comparison of histograms or core density estimations)
* Actions might include re-evaluation of risk-benefit analysis, re-training of algorithm, product recall
 | Product Release | MDR (2017/745) Article 83 EU MDR Annex III (1.1) |
| The manufacturer should perform post-market surveillance and compile reports, both according to the post-market surveillance plan. | * There is a post-market surveillance report for each product respectively product type
* The post-market surveillance reports clearly identifies the respective products via its UDI
* The post-market surveillance reports identify the post-market data and conclude whether activities are required
 |  |  | MDR (2017/745) Article 85 f. |
| The manufacturer should establish a post-market risk management system | * There is a specification how, how often and by whom the state of the art is monitored and re-assessed
* The state-of-the-art assessment takes latest algorithms for machine learning and for improving interpretability into account
* The state-of-the-art assessment takes alternatives for the “ground-truth” respectively the gold standard
* There is a specification how, how often and by whom post-market data are evaluated for new or changed hazards, hazardous situations, and risks
* The post-market risk analysis searches for (adverse) behavioural changes or (foreseeable) misuse
* For products that have been placed on the market for more than one year post-market risk management activities are documented
 | * It is possible to combine post-market risk management and post-market surveillance
* The interpretability includes transparency and explainability
* The forseeable misuse may include radiologists that rely on the software and don't look at the images anymore, so they overlook finding
* The foreseeable misuse can include users or operators not updating the software or using the product after communicated end of life.
 |  | ISO 14971 clause 10 |
| The manufacturer, should assess the design change before deciding whether notified bodies respectively authorities have to be informed | * For products marketed in the US there is an Algorithm Change Protocol (ACP) and a "SaMD Pre-Specifications" (SPS)
* There is a description of design changes
* There is an impact analysis for these design changes
 | Descriptions of design changes take into account changes to* Intended use
* ML architecture
* Software architecture
* Use of 3rd party libraries (SOUP, OTS)
* Programing language
* User Interface including warning
* Data interfaces
 |  | MDR (2017/745) Article 87 ff.ISO 13485 clause 7.3.9FDA's "Proposed Regulatory Framework for Modifications to Artificial Intelligence / Machine Learning (AI/ML) Based Software as Medical Device |

Annex A:
IMDRF Essential Principles

IMDRF- Essential principles provide broad, high-level, criteria for design, production, and postproduction throughout the life-cycle of all medical devices and IVD medical devices, ensuring their safety and performance

IMDRF Essential Principles (EPs) were evaluated to cover aspects considered applicable to the regulation of SaMDs. Main IMDRF references include the following:

1. "Essential Principles of Safety and Performance of Medical Devices and IVD Medical Devices", IMDRF Good Regulatory Review Practices Group, IMDRF GRRP WG/N47 FINAL, 31 October 2018 (http://www.IMDRF.org/docs/IMDRF/final/technical/IMDRF-tech-181031-grrp-essential-principles-n47.pdf)
2. Table for use in mapping IMDRF Essential Principles (31 October 2018) to controls for Artificial Intelligence and Machine Learning Algorithms utilized in Medical Technology

The scope of Epsapplicable to AI-MDs cover the following:

1. Safety and Performance of Medical Devices – General Essential Principles
2. IMDRF Essential Principles Applicable to all Medical Devices and IVD Medical Devices
* General
* Clinical Evaluation
* Medical Devices and IVD Medical Devices that Incorporate Software or are Software as a Medical Device
* Medical Devices and IVD Medical Devices with a Diagnostic or Measuring Function
* Labelling
* Protection against the Risks posed by Medical Devices and IVD Medical Devices intended by the Manufacturer for use by Lay Users
1. Essential Principles Applicable to IVD Medical Devices
* Performance characteristics

Details on the Essential Principles and their mapping to AI4 concepts are given below.

NOTE: EP#: refers to original section numbers in the document- "Essential Principles of Safety and Performance of Medical Devices and IVD Medical Devices", IMDRF Good Regulatory Review Practices Group, IMDRF GRRP WG/N47 FINAL, 31 October 2018.

Table A.1: IMDRFEP 5.1 – General

| EP # | EP requirements | EP key concepts |
| --- | --- | --- |
| 5.1.1  | Medical devices and IVD medical devices should achieve the performance intended by their manufacturer and should be designed and manufactured in such a way that, during intended conditions of use, they are suitable for their intended purpose. They should be safe and perform as intended, should have risks that are acceptable when weighed against the benefits to the patient, and should not compromise the clinical condition or the safety of patients, or the safety and health of users or, where applicable, other persons. | Performance; Intended conditions of use; Safety; Perform as intended; Acceptable risks; Patient benefits; Health |
| 5.1.2  | Manufacturers should establish, implement, document and maintain a risk management system to ensure the ongoing quality, safety and performance of the medical device and IVD medical device. Risk management should be understood as a continuous iterative process throughout the entire lifecycle of a medical device and IVD medical device, requiring regular systematic updating. In carrying out risk management manufacturers should: | Risk management system; Quality; Safety; Performance; Continuous, iterative risk management; MD life cycle |
| 5.1.2  | a) establish and document a risk management plan covering each medical device and IVD medical device; | Risk management plan |
| 5.1.2  | b) identify and analyse the known and foreseeable hazards associated with each medical device and IVD medical device; | Identify and analyse hazards |
| 5.1.2  | c) estimate and evaluate the risks associated with, and occurring during, the intended use and during reasonably foreseeable misuse; | Risk; Intended use; Foreseeable misuse |
| 5.1.2  | d) eliminate or control the risks referred to in point (c) in accordance with the requirements of points 5.1.3 and 5.1.4 below; | Risk elimination; Risk control |
| 5.1.2  | e) evaluate the impact of information from the production and postproduction phases, on the overall risk, benefit-risk determination and risk acceptability. This evaluation should include the impact of the presence of previously unrecognized hazards or hazardous situations, the acceptability of the estimated risk(s) arising from a hazardous situation, and changes to the generally acknowledged state of the art. | Continuous, iterative risk management |
| 5.1.2  | f) based on the evaluation of the impact of the information referred to in point (e), if necessary, amend control measures in line with the requirements of points 5.1.3 and 5.1.4 below. | Continuous, iterative risk management; Update control measures |
| 5.1.3  | Risk control measures adopted by manufacturers for the design and manufacture of the medical device and IVD medical device should conform to safety principles, taking account of the generally acknowledged state of the art. When risk reduction is required, manufacturers should control risks so that the residual risk associated with each hazard as well as the overall residual risk is judged acceptable. In selecting the most appropriate solutions, manufacturers should, in the following order of priority: | Risk control measures; Safety principles compliance; State of the art; Risk control |
| 5.1.3  | a) eliminate or appropriately reduce risks through safe design and manufacture; | Safe design |
| 5.1.3  | b) where appropriate, take adequate protection measures, including alarms if necessary, in relation to risks that cannot be eliminated; and | Alarms; Risks that cannot be eliminated |
| 5.1.3  | c) provide information for safety (warnings/precautions/contra-indications) and, where appropriate, training to users. | Alarms; User training |
| 5.1.4  | The manufacturer should inform users of any relevant residual risks. | Residual risk information for user |
| 5.1.5  | In eliminating or reducing risks related to use, the manufacturer should: | Risk reduction |
| 5.1.5  | a) appropriately reduce the risks related to the features of the medical device and IVD medical device and the environment in which the medical device and IVD medical device are intended to be used (e.g. ergonomic/usability features, tolerance to dust and humidity) and | Risk reduction; Intended usage environment |
| 5.1.5  | b) give consideration to the technical knowledge, experience, education, training and use environment and, where applicable, the medical and physical conditions of intended users. | Consider user knowledge |
| 5.1.6  | The characteristics and performance of a medical device and IVD medical device should not be adversely affected to such a degree that the health or safety of the patient and the user and, where applicable, of other persons are compromised during the expected life of the device, as specified by the manufacturer, when the medical device and IVD medical device is subjected to the stresses which can occur during normal conditions of use and has been properly maintained and calibrated (if applicable) in accordance with the manufacturer's instructions. | Stress resistance; Intended use; Expected life of device |
| 5.1.7  | Medical devices and IVD medical devices should be designed, manufactured and packaged in such a way that their characteristics and performance, including the integrity and cleanliness of the product and when used in accordance with the intended use, are not adversely affected by transport and storage (for example, through shock, vibrations, and fluctuations of temperature and humidity), taking account of the instructions and information provided by the manufacturer. The performance, safety, and sterility of the medical device and IVD medical device should be sufficiently maintained throughout any shelf-life specified by the manufacturer. | - |
| 5.1.8  | Medical devices and IVD medical devices should have acceptable stability during their shelf-life, during the time of use after being opened (for IVDs, including after being installed in the instrument), and during transportation or dispatch (for IVDs, including samples). | Stability; Shelf life |
| 5.1.9  | All known and foreseeable risks, and any undesirable side-effects, should be minimized and be acceptable when weighed against the evaluated benefits arising from the achieved performance of the device during intended conditions of use taking into account the generally acknowledged state of the art. | Risk; Side-effects |

Table A.2: IMDRFEP 5.2 – Clinical evaluation

| EP # | EP requirements | EP key concepts |
| --- | --- | --- |
| 5.2.1  | Where appropriate and depending on jurisdictional requirements, a clinical evaluation may be required. A clinical evaluation should assess clinical data to establish that a favourable benefit-risk determination exists for the medical device and IVD medical device in the form of one or more of the following:* clinical investigation reports (for IVDs, clinical performance evaluation reports)
* published scientific literature/reviews
* clinical experience
 | Clinical evaluation; Benefit-risk determination; Clinical investigation report; Published scientific literature; Clinical experience |
| 5.2.2  | Clinical investigations should be conducted in accordance with the ethical principles that have their origin in the Declaration of Helsinki. These principles protect the rights, safety and well-being of human subjects, which are the most important considerations and shall prevail over interests of science and society. These principles shall be understood, observed, and applied at every step in the clinical investigation. In addition, some countries may have specific regulatory requirements for pre-study protocol review, informed consent, and for IVD medical devices, use of leftover specimens. | Ethicalprinciples; Declaration of Helsinki Rights; Safety; Well-being; Pre-study protocol review; Informed consent; Leftover specimen |

Table A.3: IMDRFEP 5.8 – Medical devices and IVD medical devices that incorporate software or are software as a medical device

| EP # | EP requirements | EP key concepts |
| --- | --- | --- |
| 5.8.1  | Medical devices and IVD medical devices that incorporate electronic programmable systems, including software, or are software as a medical device, should be designed to ensure accuracy, reliability, precision, safety, and performance in line with their intended use. In the event of a single fault condition, appropriate means should be adopted to eliminate or appropriately reduce consequent risks or impairment of performance. | Electronic programmable systems; Software; Software as a medical device; Accuracy; Reliability; Precision; Safety; Performance; Single fault conditions; Risk reduction |
| 5.8.2  | For medical devices and IVD medical devices that incorporate software or are software as a medical device, the software should be developed, manufactured and maintained in accordance with the state of the art taking into account the principles of development life cycle (e.g., rapid development cycles, frequent changes, the cumulative effect of changes), risk management (e.g., changes to system, environment, and data), including information security (e.g., safely implement updates), verification and validation (e.g., change management process). | State of the art; Principles of development life cycle (e.g., rapid development cycles, frequent changes, the cumulative effect of changes); Risk management (e.g., changes to system, environment, and data); Information security (e.g., safely implement updates); Verification; Validation; Change management process |
| 5.8.3  | Software that is intended to be used in combination with mobile computing platforms should be designed and developed taking into account the platform itself (e.g. size and contrast ratio of the screen, connectivity, memory, etc.) and the external factors related to their use (varying environment as regards level of light or noise). | Mobile computing platforms; Size; Contrast ratio of the screen; Connectivity; Memory; External factors related to their use (varying environment as regards level of light or noise) |
| 5.8.4  | Manufacturers should set out minimum requirements concerning hardware, IT networks characteristics and IT security measures, including protection against unauthorized access, necessary to run the software as intended. | Minimum requirements; Hardware; IT networks characteristics; IT security measures; Protection against unauthorized access |
| 5.8.5  | The medical device and IVD medical device should be designed, manufactured and maintained in such a way as to provide an adequate level of cybersecurity against attempts to gain unauthorized access. | Cybersecurity; Protection against unauthorized access |

Table A.4: IMDRFEP 5.10 – Labelling

| EP # | EP requirements | EP key concepts |
| --- | --- | --- |
| 5.10.1  | Medical devices and IVD medical devices should be accompanied by the information needed to distinctively identify the medical device or IVD medical device and its manufacturer. Each medical device and IVD medical device should also be accompanied by, or direct the user to any safety and performance information relevant to the user, or any other person, as appropriate. Such information may appear on the medical device or IVD medical device itself, on the packaging or in the instructions for use, or be readily accessible through electronic means (such as a website), and should be easily understood by the intended user. | Information [Manual]; Safety; Performance; Easily understood |

Table A.5: IMDRFEP 5.12 – Protection against the risks posed by medical devices and IVD medical devices intended by the manufacturer for use by lay users

| EP # | EP requirements | EP key concepts |
| --- | --- | --- |
| 5.12.1  | Medical devices and IVD medical devices for use by lay users (such as self-testing or near-patient testing intended for use by lay users) should be designed and manufactured in such a way that they perform appropriately for their intended use/purpose taking into account the skills and the means available to lay users and the influence resulting from variation that can be reasonably anticipated in the lay user's technique and environment. The information and instructions provided by the manufacturer should be easy for the lay user to understand and apply when using the medical device or IVD medical device and interpreting the results. | Lay user; Self-testing; Intended use; Usage variations (user technique, usage environment); Instructions; Easy to understand; Easy to apply |
| 5.12.2  | Medical devices and IVD medical devices for use by lay users (such as self-testing or near-patient testing intended for use by lay users) should be designed and manufactured in such a way as to: | Lay user; Self-testing; Near-patient testing |
| 5.12.2  | a) ensure that the medical device and IVD medical device can be used safely and accurately by the intended user per instructions for use. When the risks associated with the instructions for use cannot be mitigated to appropriate levels, these risks may be mitigated through training. | Safety; Accuracy; Instructions; Risk reduction; Training |
| 5.12.2  | b) appropriately reduce the risk of error by the intended user in the handling of the medical device or IVD medical device and, if applicable, in the interpretation of the results. | Risk reduction; Risk of error; Handling; Interpretation of results |
| 5.12.3  | Medical devices and IVD medical devices for use by lay users (such as self-testing or near-patient testing intended for use by lay users) should, where appropriate, include means by which the lay user: | Lay users; Self-testing; Near-patient testing |
| 5.12.3  | a) can verify that, at the time of use, the medical device or IVD medical device will perform as intended by the manufacturer, and | Verification; Intended use; Performance |
| 5.12.3  | b) is warned if the medical device or IVD medical device has failed to operate as intended or to provide a valid result. | Warning; Failure; Valid result |

Table A.6: IMDRFEP7.2 – Performance characteristics

| EP # | EP requirements | EP key concepts |
| --- | --- | --- |
| 7.2.1 | Performance Characteristics IVD medical devices should achieve the analytical and clinical performances, as stated by the manufacturer that are applicable to the intended use/purpose, taking into account the intended patient population, the intended user, and the setting of intended use. These performance characteristics should be established using suitable, validated, state of the art methods. For example: | Performance characteristics; Analytical performance; Clinical performance; Validation; State of the art |
| 7.2.1 | a) The analytical performance can include, but is not limited to,* + - * 1. Traceability of calibrators and controls
				2. Accuracy of measurement (trueness and precision)
				3. Analytical sensitivity/Limit of detection
				4. Analytical specificity
				5. Measuring interval/range
				6. Specimen stability
 | Traceability of calibrators and controls; Accuracy of measurements (trueness and precision); Analytical sensitivity/Limit of detection; Analytical specificity; Measuring interval/range; Specimen stability |
| 7.2.1 | b)The clinical performance, for example diagnostic/clinical sensitivity, diagnostic/clinical specificity, positive predictive value, negative predictive value, likelihood ratios, and expected values in normal and affected populations.  | Clinical performance; Diagnostic/clinical sensitivity; Diagnostic/clinical specificity; Positive predictive value; Negative predictive value; Likelihood ratios; Expected values in normal and affected populations. |
| 7.2.1 | c) Validated control procedures to assure the user that the IVD medical device is performing as intended, and therefore the results are suitable for the intended use. | Validation; Control procedures; Intended use |
| 7.2.2  | Where the performance of an IVD medical device depends on the use of calibrators or control materials, the traceability of values assigned to such calibrators or control materials should be ensured through available reference measurement procedures or available reference materials of a higher order. | Calibrators; Control materials; Traceability of values; Reference measurement procedures; Reference materials of higher order |
| 7.2.3  | Wherever possible, values expressed numerically should be in commonly accepted, standardized units and understood by the users of the IVD medical device. | Numerical values; Standardized units; User understanding |
| 7.2.4  | The performance characteristics of the IVD medical device should be evaluated according to the intended use statement which may include the following: | Performance evaluation; Intended use |
| 7.2.4  | a) intended user, for example, lay user, laboratory professional; | Intended user |
| 7.2.4  | b) intended use environment, for example, patient home, emergency units, ambulances, healthcare centres, laboratory; | Intended use environment |
| 7.2.4  | c) relevant populations, for example, paediatric, adult, pregnant women, individuals with signs and symptoms of a specific disease, patients undergoing differential diagnosis, blood donors, etc. Populations evaluated should represent, where appropriate, ethnically, gender, and genetically diverse populations so as to be representative of the population(s) where the device is intended to be marketed. For infectious diseases, it is recommended that the populations selected have similar prevalence rates. | Relevant population; Appropriate representation; Ethnicity; Gender; Genetic diversity; Representative population; Prevalence rates |

Annex B:
IMDRFSaMD Risk Categorization Framework

The IMDRFpublication "Software as a Medical Device: Possible Framework for Risk Categorization and Corresponding Considerations"characterizes the medical devices by assigning different risk levels to them based on combination of the significance of the information provided by the SaMD to the healthcare decision and the healthcare situation or condition as shown in Table B.1.

Table B.1: IMDRFSaMD risk categories



The four categories (I, II, III, IV) shown in Table B.1 are based on the levels of impact on the patient or public health where accurate information provided by the SaMD to treat or diagnose, drive or inform clinical management is vital to avoid death, long-term disability or other serious deterioration of health, mitigating public health.

The categories are in relative significance to each other. Category IV has the highest level of impact, Category I the lowest

The criteria for determining (a) SaMD category and (b) Levels of Autonomy are explained as follows.

## B.1 Criteria for determining the SaMDcategory

The criteria for determining whether anSaMD is Category IVare:

* SaMD that provides information to treat or diagnose a disease or conditions in a critical situation or condition is a Category IV and is considered to be of very high impact.

The criteria for determining whether anSaMD is Category IIIare:

* SaMD that provides information to treat or diagnose a disease or conditions in a serious situation or condition is a Category III and is considered to be of high impact.
* SaMD that provides information to drive clinical management of a disease or conditions in a critical situation or condition is a Category III and is considered to be of high impact.

The criteria for determining whether anSaMD is Category IIare:

* SaMD that provides information to treat or diagnose a disease or conditions in a nonserious situation or condition is a Category II and is considered to be of medium impact
* SaMD that provides information to drive clinical management of a disease or conditions in a serious situation or condition is a Category II and is considered to be of medium impact.
* SaMD that provides information to inform clinical management for a disease or conditions in a critical situation or condition is a Category II and is considered to be of medium impact.

The criteria for determining whether anSaMD is Category I are:

* SaMD that provides information to drive clinical management of a disease or conditions in a non-serious situation or condition is a Category I and is considered to be of low impact.
* SaMD that provides information to inform clinical management for a disease or conditions in a serious situation or condition is a Category I and is considered to be of low impact.
* SaMD that provides information to inform clinical management for a disease or conditions in a non-serious situation or condition is a Category I and is considered to be of low impact

## B.2 Levels of autonomy

The IMDRFSaMD Categories table was revised to accounted for various levels of autonomy as shown in the shown in Table B.2 below. Additional levels have been added to the "Treat or diagnose" category:

Table B.2: IMDRFSaMD risk categories(revised)

|  | Significance of information provided by software to healthcare decision |
| --- | --- |
| State of Healthcare situation or condition | Treat or diagnose with no intervention possible | Treat or diagnose with override | Treat or diagnose with approval | Drive clinical management | Inform clinical management |
| Critical | VI | V | IV | III | II |
| Serious | V | IV | III | II | I |
| Non-serious | IV | III | II | I | I |

Three different levels of autonomy proposed are:

1. **Approval:** the software may make suggestions to the user, but either it cannot take action on its own, or it requires operator approval before taking action.
2. **Override**: the software can take action without approval, but the operator has the ability to over-ride (cancel) the software if need be. For example, a human driver in a self-driving car can take control.
3. **No Intervention**: the operator is not involved in the treatment and has no ability to override the software.

Annex C:
Johnerregulatory guidelines for AI- for medical devices

The Johner Guideline for AI-MDs is prepared and released by the JohnerInstitute,Germany.The guideline is published under the Creative Commons License of type BY-NC-SA.This document is managed via the version management system git or the GitHub platform. Only the documents listed in this repository are valid. Full documentation of Johner Guidelines can be found at: <https://github.com/johner-institut/ai-guideline/blob/master/Guideline-AI-Medical-Devices_EN.md>.

Johner Guidelines - Objectives

The objective of JohnerGuidelines is to provide medical device manufacturers and notified bodies instructions and to provide them with a concrete checklist

* to understand what the expectations of the notified bodies are,
* to promote step-by-step implementation of safety of medical devices, that implement artificial intelligence methods, in particular machine learning,
* to compensate for the lack of a harmonized standard (in the interim) to the greatest extent possible.

Johner Guidelines - Scope

Johner guidelines do not set forth specific requirements for the products, but for the processes. It contains the following chapters:

1. General requirements
2. Requirements for product development
3. Intended use
4. Software requirement specification
5. Data management
6. Model development
7. Product development
8. Product release
9. Requirements for phases following development

Annex D:
FG-AI4H data and AI solution quality assessment criteria

Data and AI solution quality assessment criteria were formulated by the ITU-T Focus Group on AI for Health's DAISAM Working Group,following the data and FGAI4H-F-032-A01: Data and AI solution assessment methods, governed by FGAI4H-F-103: Updated FG-AI4H data acceptance and handling policy.

Based on these criteria, a quality assessment questionnaire was prepared to serve as a preliminary checklist intended to guide the various AI4 Health Topic Groups in following a uniform procedure for preparing the data and AI solution technical requirements specifications and submitting them in a common reporting format.

This DAISAM quality assessment questionnaire includes a glossary that contains definitions for technical terms specific to data and AI solution quality criteria.This is provided to guide the FG-AI4 Health Topic Groups in interpreting the quality assessment checklist in a clear and concise manner and in mapping the respective technical requirement specifications.

The data and AI solution quality assessment criteria are listed in Table D.1

Table D.1:FG-AI4H data and AI solution quality assessment criteria

| AI Model Development Workflow | Assessment Criteria | Description  | Examples |
| --- | --- | --- | --- |
| Problem Definition | Underlying Task | Underlying Task refers to the broad taxonomy followed in organizing Machine Learning (ML) Tasks based on how the solution will be applied to solve or address the specific business problem of the respective practice domain use cases. Please refer to sections- Level-1A and Level-1B of FGAI4H-C-104 for domain use-case thematic classifications) | * Classification
* Regression/Prediction
* Clustering
* Association rule learning
* Decision Support / Virtual Assistance / Recommendation systems
* Matching
* Labelling
* Detection
* Segmentation
* Sequential data models
* Anomaly detection and Fraud Prevention
* Compliance Monitoring / Quality Assurance
* Process optimization / Automation
* Other
 |
| Data Preparation | Input Data Sources, Types & Formats | * Input Data refers to the subset of the dataset that is used to train the AI model
* Data Type refers to the type of the different data attributes involved
* Data Format refers to the standard representation formats of the different data attributes involved
 | Input data sources include:* Electronic Health Records (Anonymised)
* Medical Images
* Vital signs signals
* Lab test results
* Photographs
* Non-medical data-Socioeconomic, Environmental, etc)
* Questionnaire responses
* Free Text (Discharge / Summary, Medical History / Notes, etc.)
* Other

Input Data Types include:* Real valued
* Integer-valued
* Categorical value
* Ordinal value
* Strings
* Dates
* Times
* Complex data type
* Other

Standard Input Data Formats include:* DICOM PS3.0 (latest versions)- for Diagnostic Image (X-Ray, CT,MRI, PET, other pathological slides, etc)
* JPEG / PNG – for Static Image
* MP3 / OGG – for Audio:
* MP4 / MOV- for Video
* SNOMED – for clinical observations/terminology
* LOINC- for laboratory observations
* WHO ICD-10 for disease classifications
* RxNORM for Medication Code
* Other
 |
| Data Preparation | Output Data Types | Output Data refers to type of data generated by the AI Model, when a particular ML algorithm is applied on the Input Data  | * Binary/Class output (0 or 1) as in case of classification problems
* Probability output (0-1) as in case of classification problems
* Continuous valued output as in case of regression problems
 |
| Data Preparation | Target Data Types | Target Data refers to the output data in the training dataset that is defined as the reference (ground truth) for AI Model validation/testing | * Binary/Class output (0 or 1) as in case of classification problems
* Probability output (0-1) as in case of classification problems
* Continuous valued output as in case of regression problems
 |
| AI Model Selection | Model Type | Model Type refers to the specific machine learning algorithm and its configuration that is applied on the training dataset in order to learn the Model | Broad Classification of ML Algorithms include:* Supervised Learning based algos
* Linear Regression
* Logistic Regression
* k-nearest neighbours
* Decision Trees
* Random Forest
* Gradient Boosting Machines
* XGBoost
* Support Vector Machines (SVM)
* Neural Network
* other
* Unsupervised Learning based algos
* k means clustering
* Hierarchical clustering
* Neural Network
* other
* Reinforcement Learning based algos
* Association rule learning based algos
* Apriori algorithm
* Eclat algorithm
* Deep learning based algos
* Convolutional Neural Network (CNN)
* Recurrent Neural Networks (RNNs)
* Long Short-Term Memory Networks (LSTMs)
* Stacked Auto-Encoders
* Deep Boltzmann Machine (DBM)
* Deep Belief Networks (DBN)
* other
 |
| AI ModelEvaluation | Evaluation Metrics | Metrics used to quantify the errorsand to evaluate the performance quality of the trained model on the test datasetSelection of metrics depends on the type of the problem & the type of the model under consideration | * Model Accuracy (%)
* Model Accuracy -Mean & Standard Deviation
* Model Accuracy –Box Plot Summarization
* Root Mean Squared Error(RMSE)
* Sensitivity (True Positive Rate)
* Specificity (True Negative Rate)
* F1-Score (class wise performance determination)
* Confusion matrix
* K-fold Cross-validation
* Gain and Lift Charts
* Kolmogorov Smirnov Chart
* Gini Coefficient
* Log [Loss](https://developers.google.com/machine-learning/crash-course/descending-into-ml/training-and-loss)
* [Area under the ROC curve (AUC)](https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc)
* Concordant – Discordant Ratio
* Other user defined performance measures
* Other
 |
| AI Model Optimization | Optimization Objective(s) | This deals with the iterative process (feedback principle) of reconfiguring or tweaking the Model Parameters to their optimal values in order to achieve the desired level of accuracy or performance score in comparison with the baseline definition.Model performance can besystematically tracked by maintaining progressive versions of Code, Model, and Data. | Optimization techniques include:* Adding or deleting Features /Attributes of the input data
* Aggregating or Decomposing Features /Attributes of the input data
* Tuning Model Hyper-parameters
* Normalization & Standardization of input data
* Changing the learning rate of the algorithm
* Examining the Statistical Significance of results
* Recruiting Ensemble Methods for combining / augmenting the prediction scores of multiple models
* Monitoring and tracking API response times and Computational Memory requirements of the serving infrastructure
* Etc.
 |
| Safety Standards Compliance | Safety tool(s)training | This deals with the user training/orientation given on how to identify potential human safety risks occurring due to accidental or malicious misuse of the technology involved in AI Model deployment | Safety Risk Mitigation and Management Plan & Procedure |
| Safety tool(s) deployment | This deals with the incorporation of necessary preventative system measures/tools as per the defined Risk Mitigation Plan to ensure that no damage or harm is caused to human safety out of potential physical or cyber-attacks on the AI Model being applied. | * Adopting governance procedures to assert alternative system fault tolerance plans
* Adopting security mechanisms like
* Authentication
* Role based Access Control
* Encryption
* Transport Level Security
* Informed Consent
* Anonymisation
* etc
* Maintaining Data Audit Logs for secure content verification, based on
* Blockchain Technology
* Merkle Trees
* etc
* Implementing Security Standards based onDigital Certificate, SSL, SHA-256, etc
 |
| AI Model Testing | Test Data Quality Tests | Test Data refers to the subset of the dataset and not part of the training dataset that is used to evaluate the ML Model accuracy after its primary vetting by the validation datasetQuality tests are performed to minimize the noise and variance of the test data in order to maximize the performance accuracy of ML algorithm applied on it | Standard Test Options include:* Training and testing on the same dataset
* Split tests
* Multiple split tests
* Cross validation
* Multiple cross validation
* Statistical significance
 |

Annex E:
ITU ML5G high-level requirements mapping to AI for health requirements

Requirements analysis was performed on the ITU-T FG-ML Technical Specification "Unified architecture for machine learning in 5G and future networks" to identify high-level requirements that could be translated and applied for regulatory assessment of AI-MDs. The list of high-level requirements is given in the following tables.

|  |  |
| --- | --- |
| ITU ML5G Req. Code | ML-unify-001 |
| Requirement | Multiple sources of data are recommended to be used to take advantage of correlations in data. |
| Description | In future networks, sources of data may be heterogeneous, integrated with different NFs, and may report different formats of data. These varied "perspectives" can provide rich insights upon correlated analysis.Example: Analysis of data from UE, RAN, CN and AF is needed to predict potential issues related to quality of service (QoS) in end-to-end user flows.Thus, an architecture construct to enable the ML pipeline to collect and correlate data from these varied sources is needed.  |
| Relevance for Healthcare / Assessment Sven (MXH) | Despite explicit parameters set for representativeness, a small number of data sources increases the risk of a data bias. Since always application dependent, this should not be required but recommended. |
| Required / Recommended? | Recommended |

|  |  |
| --- | --- |
| ITU ML5G Req. Code | ML-unify-005 |
| Requirement | Logical entities of the ML pipeline are required to be capable of splitting their functionalities or be hosted on separate technology-specific nodes. Similarly, multiple logical entities are required to be capable of being implemented on single node.  |
| Description | In future networks, HAS for NFs will optimize the location and the performance accordingly. The network function virtualization orchestrator (NFVO) plays an important role in this. To carry forward such benefits to the ML use case, similar optimizations should also be applied to ML pipeline nodes. Moreover, the constraints applicable to an ML pipeline [e.g., training may need a graphic processor unit (GPU) and may need to be done in a sandbox domain] may be unique.  |
| Relevance for Healthcare / Assessment Sven (MXH) | This roughly falls into the category of distributed training / inference / federated learning. At MXH, we are not too convinced by the latter yet, since far away from concrete application and too high-level to make a reasonable judgement on. For pragmatic reasons, we would assess this as 'recommended'. |
| Required / Recommended? | Recommended |

|  |  |
| --- | --- |
| ITU ML5G Req. Code | ML-unify-011 |
| Requirement | Intention is required to specify the sources of data, repositories of models, targets/sinks for policy output from models, constraints on resources / use case. |
| Description | The separation between technology agnostic part of the use case and technology-specific deployment (e.g., 3GPP) is captured in the design time of future network services. Intent specification for the ML use cases achieves this separation for the ML overlay. See clauses 3.2.5 and 3.2.6 for definitions.  |
| Relevance for Healthcare / Assessment Sven (MXH) | Specification of data sources is required to provide transparency on robustness, e.g. to exclude misfit situations with unclear model outcome. |
| Required / Recommended? | Required |

|  |  |
| --- | --- |
| ITU ML5G Req. Code | ML-unify-017 |
| Requirement | Model training is required to be done in the sandbox using training data.A sandbox domain is recommended to optimize the ML pipeline. Simulatorfunctions hosted in the sandbox domain may be used to derive data foroptimizations. |
| Description | Model training is a complicated function, it has several considerations: use of specific hardware for speed, availability of data (e.g., data lakes), parameter optimizations, avoiding bias, distribution of training (e.g., multi-agent reinforcement learning), the choice of loss function for training. The training approach used exploration of hyper parameters, for example.Moreover, in future networks, operators will want to avoid service disruptions while model training and updates are performed.These considerations point to the use of a simulator for producing the data for training the models, as well as its use in a sandbox domain.  |
| Relevance for Healthcare / Assessment Sven (MXH) | Separation of development and production setting is required because uncontrolled, continuous learning imposes the risk of unexpected model biases. |
| Required / Recommended? | Required |

|  |  |
| --- | --- |
| ITU ML5G Req. Code | ML-unify-018 |
| Requirement | The capabilities to enable a closed loop monitoring and update, based on theeffects of the ML policies on the network, are required. |
| Description | Closed loop is needed to monitor the effect of ML on network operations. Various KPIs are measured constantly and the impact of the ML algorithm on them as well as on the ML pipeline itself (due to operations of the MLFO) are monitored and corrected constantly. These form inputs to the simulator that generate data. These data can cover new or modified scenarios accordingly in future (e.g., a new type of anomaly is detected in the network, the simulator is modified to include such data. which can also train the model to detect that data type).  |
| Relevance for Healthcare / Assessment Sven (MXH) | Sounds like monitoring of ML algorithm performance in the production setting. Reasonable thing to do in order to be able to intervene if outcomes don't hold up to expectations and might cause risks to patient safety. |
| Required / Recommended? | Required |

|  |  |
| --- | --- |
| ITU ML5G Req. Code | ML-unify-019 |
| Requirement | A logical orchestrator (MLFO: ML function orchestrator) is required to be usedfor monitoring and managing the ML pipeline nodes in the system.MLFO monitors the model performance, and model reselection is recommendedwhen the performance falls below a predefined threshold. |
| Description | The varied levels and sources of data (core, edge), including the simulator and the sandbox domain, imply that there could be various training techniques including distributed training. Complex models that are chained (or derived) may in fact be trained using varied data. The performance of such models can be determined and compared in the sandbox domain using a simulator.Based on comparisons, operators can then select the model for specific use cases. This can be used in conjunction with the MLFO to reselect the model. Note: evaluation may involve network performance evaluation along with model performance.  |
| Relevance for Healthcare / Assessment Sven (MXH) | Sounds like previous point, monitoring of model outcomes |

Annex F:
DIN SPEC 92001-AI devices life cycle processes requirements

**DIN SPEC 92001-1**

**ICS 35.080; 35.240.01**

**Artificial Intelligence – Life Cycle Processes and Quality Requirements – Part 1: Quality Meta Model;**

1. **Introduction**

Challenge: For these reasons, quality assessment of an AI module still poses a major challenge. It becomes more difficult to confirm, verify, and validate an AI module during conception, development, deployment, operation, and retirement which are wide-ranging tasks.

Abstract: This document introduces an AI quality meta model to outline key aspects of AI quality including the previously mentioned AI quality pillars. For AI quality analysis, an approach for risk evaluation and a suitable software life cycle are provided. The given AI life cycle is consistent with the international standard for systems and software engineering. The second part of this specification, DIN SPEC 92001-2, provides specific AI quality requirements.

|  |
| --- |
| **Scope** |
| Purpose | Establish a quality-assuring and transparent life cycle of AI modules. Critical quality criteria are identified and AI-specific problems are addressed. To achieve this, this document presents a set of quality requirements that are structured in an AI specific quality metamodel. It is important to note that not all AI modules impose the same quality requirements. document proposes the differentiation between AI modules with regard to their safety, secureThe document outlines and defines the three central quality pillars functionality & performance, robustness, and comprehensibility.  |
| Field of Application | This document applies to all life cycle stages of an AI module — concept, development, deployment, operation, and retirement — and addresses a variety of different life cycle processes.  |

1. **Terms and definitions**

For the purposes of this document, the following terms and definitions apply.

DIN and DKE maintain terminological databases for use in standardization at the following addresses:

— DIN-TERMinologieportal: available at <https://www.din.de/g>

1. **Quality Meta model**

The key quality characteristics, the so-called quality pillars, that need to be taken into account throughout the whole life cycle of an AI module, are functionality & performance, robustness and comprehensibility. These three quality pillars are not fully disjoint. For instance, robustness may be conceived as part of functionality & performance, since the adaptation to unknown environments can be a functionality requirement in a given application. In this way, AI modules are divided into two risk classes. In the following, AI modules with safety, security, privacy, or ethical relevance are summarized in components with (potentially) high risk and the latter in components with low risk. For high risk AI modules, a deviation from the quality requirements is either not permitted or is to be justified, while for low risk AI modules this is less strict.

This document, each AI module is considered to be either of high or low risk or it is assumed that a mapping of internal risk classes to high risk and low risk, respectively, is carried out. For safety, security, privacy, or ethically relevant AI modules this document requires the consideration of all listed quality requirements. Potential deviations of such AI modules need a profound justification.

* 1. **AI parts Module of the and AI Software quality metamodel System are Relation**

Software systems are composed of interacting system elements, where each has its own purpose and requirements, respectively. The AI module is one of these elements that consist of AI methods and algorithms, respectively. As an element of the software system, it relates to and interacts with other elements such as hardware, software or data and with the surrounding environment such as humans. Henceforth, this document focuses on the quality assurance of AI artifacts within the software system. These artifacts can be hybrid systems. It is required to keep in mind that further standards, requirements, and regulations can apply to the overall software system and consequently to the AI module. In order to give a framework for DevOps of trustworthy AI modules, a quality metamodel is proposed and described in this document.

* 1. **Risk Evaluation**

|  |  |
| --- | --- |
| **Risk-grade** | **Description** |
| High-risk | AI modules (so called “critical” AI modules) have safety, security, privacy, or ethical relevance. Domains with such relevance can be autonomous driving, medical diagnostics, and credit ratings. |
| Low-risk | For low risk AI modules, deviations from recommended requirements are permitted without further justification. A deviation from highly recommended requirements for low risk AI modules is only permitted in exceptional cases and with appropriate justification, whereas deviations from mandatory requirements such as the establishment of a risk identification and assessment process are not accepted. Deviations from recommended and highly recommended requirements are only permitted in exceptional cases and with appropriate justification, whereas deviations from mandatory requirements are not allowed.Low risk is called “comfort” AI modules. |

* 1. **Environment, Platform, Dthis is our ata, Model**

|  |  |
| --- | --- |
| **Model type** | **Description** |
| Model Space | The model space includes all sets of potential approaches to solve the problem task at hand. Algorithms, mathematical models, architectures, and parameter configurations that can lead to suitable solutions for the prescribed task are included within this set. |
| Inference Model | The inference model is one specific element of the model space. Thus, it is composed of particular model architecture with a fixed parameter configuration. This configuration is derived from the model space via a selection method, such as a training algorithm on some data set. The inference model can be used to solve the intended task to a certain degree. |

1. **Life Cycle**
	1. **General**

|  |  |  |
| --- | --- | --- |
| **Stage** | **Definition** | **Context of AI** |
| Concept | Creation of all process and defining of the problem definition, analysis, and finding a suitable model space. Based on the specific problem suitable models should be identified and analyzed concerning properties like convergence and input assumptions. In this stage, no model hyper parameters are chosen and no final model evaluation is done. | Additionally, acceptance criteria should be defined for further quality assurance steps. It is, for instance, recommended to operationalize the problem such that its formulation contains possible actions for a solution. |
| Development | Means a number of activities, including the system design and specification, prototyping and implementation, integration, bug tracking and bug fixing, verification and validation including testing on various levels (functional, integration, testing, performance & robustness), packaging, documentation, versioning, etc.  | Data driven development approaches are used to construct an interference model in connection with classical software engineering approaches: Such activities contain data acquisition, data analysis, and the actual programming or training efforts. In the case of ML models, the data set should be analyzed, understood, and variables that are relevant for the goal or problem should be identified. In this stage, model hyper parameters are compared concerning the quality of the specific model. Different measures and metrics for the evaluation of the model quality can be considered. The aim is to find one model with specific hyper parameters that adequately solves the problem. The representation of the data set is possibly adapted to the chosen model since some ML models need a specific input shape. |
| Deployment | Transition from development to operation. | 2 levels: a) High degree of database learning, deployment includes the training of the model on the host system and the export to the target system. b) Low degree of data based learning: the transition from host to target system is also relevant. For instance, the acceptance of the AI module by the stakeholder is part of the target system and has to be obtained. Note that deployment starts the operation stage. Therefore, it is impossible to delineate clearly between deployment and operation. |
| Operation | Maintenance and evaluation aspects in the environment where the AI module is used. | Since ML algorithms can continue to learn from data through online learning and thus continue to change after training in the experimental environment. |
| Retirement | Disintegration and discontinuation of the AI module as well as the transition to a new AI module | This stage can be deleted from the software system or significantly changed such that a new AI module is created. This starts a new life cycle. Thus, this can be interpreted as a retirement of the original AI module as well. |

One important in these stages is that everything is part of development stage.

* 1. **Life Cycle Processes**

Processes are defined by title, purpose, and outcome.

a) Organizational project-enabling processes: This part is important to concept and provides each asset to make the project work and obtain all the expectations of company stakeholders. Most processes within this group are only slightly affected by new challenges introduced by AI. Nevertheless, the user of this document needs to evaluate whether changes to existing processes are required. For instance, ways in which these processes need to be refined include establishing quality evaluation criteria that are applicable to functionality & performance, robustness, and comprehensibility of AI modules.

b) Technical management processes: “are concerned with managing the resources and assets allocated by organization management and with applying them to fulfill the agreements into which the organization or organizations enter [...]. In particular they relate to planning in terms of cost, timescales and achievements, to the checking of actions to help ensure that they comply with plans and performance criteria and to the identification and selection of corrective actions [...]”. Additionally, specific measures with respective quality criteria need to be defined that allow evaluating if the AI module satisfies functionality & performance, robustness, and comprehensibility criteria.

c) Technical processes: “transform the needs of stakeholders into a product or service by means of technical actions throughout the life cycle”. They ensure that sustainable performance and overall quality is reached when the AI module is applied. This is the group of processes that is mostly affected by AI-specific challenges. An important aspect that needs to be considered within the system analysis process is, for instance, to ensure the needed extent of interpretability of the AI module.

Agreement processes is a part of process group but IN THIS DOCUMENT, authors did not use.

Agreement processes “are organizational processes that apply outside of the span of a project’s life, as well as for a project’s lifespan. Agreements allow [...] to realize value and support business strategies for […] organizations.” [3]. While agreement processes apply to the overall software system, they bear no reference to one software component and AI-specific challenges. Thus, this DIN SPEC does not include agreement processes.

1. **AI Quality Pillars**

AI quality characteristics in the form of requirements need to be considered.

The document introduces an approach to cover a sufficiently wide spectrum of AI-related software quality aspects and to emphasize the importance of AI-specific requirements. It enables the development and implementation of performance, robust, safe, and trustworthy AI modules.

**Table 2. Three key qualities**

|  |  |  |
| --- | --- | --- |
| **Key Quality** | **Definition** | **AI Meaning** |
| Functionality & performance | The degree to which an AI module is capable of fulfilling its intended task under stated conditions. | Performance evaluation and model selection are further topics that are addressed in this quality pillar. It is required to precisely define the problem or goal before development and analyze it with respect to constraints and assumptions concerning environment, platform, data, and model. After problem analysis, potential solutions need to be formalized and evaluated. To find suitable solutions, adequate performance measures and metrics shall be chosen for the given task and data. |
| Robustness | The ability of an AI module to cope with erroneous, noisy, unknown, and adversarial input data. Due to the complexity of the AI module’s environment that can result from its non-stationary and high-dimensional, robustness is a key AI quality issue. | Therefore, the AI module’s robustness needs to be adequately quantified and meet requirements that are defined in the risk analysis. The dependence of the model on environment, platform, and data has to be considered. Distributional shifts occur when the AI module is exposed to data points outside the training or testing data set. The possibility of an adversarial attack must be specifically addressed, since this poses a major risk to the operation of AI modules in safety and security relevant settings. For this, the adversary’s knowledge of the AI module and the perturbation scope, respectively, are to be assessed and defense strategies are required to be chosen accordingly and continuously monitored during development and deployment. |
| Comprehensibility | The degree to which a stakeholder with defined needs can understand the causes of an AI module’s output. The causes include the reason for a specific output, i.e. the input leading on to it, and the whole process of decision-making.  | This means that the AI component is transparent and explainable. Furthermore, a qualitative understanding between the input variables and the response is provided with respect to the stakeholder’s level of expertise and need for comprehension. For instance, the developer of an AI module needs to understand not only the data and inference model but also the model space and the mathematical framework. This quality pillar focuses on the transparency and interpretability of the chosen model. If you doesn’t explain to the stakeholder clearly (white-box), you can create some difficulties to the project (grey-box or black-box). |

**Conclusion of quality assurance**

3 parts of quality assurance is the life cycle, influencing factors, and three quality pillars. The project manager needs to join different points like the influencing factors environment, platform, data, and model. It raises awareness of possible quality issues that can arise during the different life cycle stages and processes of the AI module. The points to consider when the project manager in the life cycle is guided by the three quality. All requirements for quality assurance are collected in these quality characteristics. Thus, the AI quality meta model covers all aspects of AI quality assurance.

Bibliography

Systems and software engineering— Softwarelifecycleprocesses.ISO/IEC/IEEE 12207:2017.Techrep.ISO, IEC and IEEE, 2017.

Annex G:
AI4H Project Deliverable Reference

Table G: AI4H Project Deliverable Reference ID

|  |  |
| --- | --- |
| **AI4H Project Deliverable**  | **ITU Document Reference ID**  |
| AI Software Life Cycle Specification  | FG-AI4H-G-204  |
| AI4H regulatory [best practices | considerations]  | FG-AI4H-G-202 |
| Mapping of IMDRF Essential Principles to AI for Health Software  | FG-AI4H-G-038-A01 |
| Data Annotation Specification  | FG-AI4H-G-205-A03 |
| AI4H Training Best Practices Specification  | FG-AI4H-G-206 |
| AI4H Evaluation Specification  | FG-AI4H-G-207 |
| AI Technical Test Specification  | FG-AI4H-G-207-A02 |
| AI4H Ethics Considerations  | FG-AI4H-G-201 |
| AI4H Applications and Platform  | FG-AI4H-G-200 |
| AI4H Scale-up and Adoption  | FG-AI4H-G- |

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Standard Operating Procedure. All SOPs have to be approved and be under version control. [↑](#footnote-ref-1)