ITUPublications

International Telecommunication Union
Standardization Sector

ITU-T Focus Group Deliverable

(09/2022)

Focus Group on Artificial Intelligence for Health
(FG-AI4H)

FG-AI4H DEL2.2

Good practices for health applications of machine learning: Considerations for manufacturers and regulators

ITU-T FG-AI4H Deliverable

DEL2.2 – Good practices for health applications of machine learning: Considerations for manufacturers and regulators

Summary

This Technical Paper recommends a set of good machine learning (ML) practice guidelines to manufacturers and regulators of data driven artificial intelligence (AI) based healthcare solutions on conducting comprehensive requirements analysis and streamlining conformity assessment procedures for continual product improvement in an iterative and adaptive manner. This set of good machine learning practice guidelines gives prime priority to the factor of patient safety and focuses on a streamlined process for risk minimization and quality assurance for AI / ML based health solutions and tries to establish a system of transparency and accountability of all the processes involved in AI / ML based health solutions. The proposed set of good machine learning practices adopts, extends and leverages the best practices and recommendations provided by internationally recognized medical device regulatory agencies such as the international medical device regulators forum (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

AI/ML based medical devices, AI checklist, regulatory framework, software-as-a-medical device.

Note

This is an informative ITU-T publication. Mandatory provisions, such as those found in ITU-T Recommendations, are outside the scope of this publication. This publication should only be referenced bibliographically in ITU-T Recommendations.

Change Log

This document contains latest version of the Deliverable DEL2.2 on " *Good practices for health applications of machine learning: Considerations for manufacturers and regulators* "

Editors:	LuisOala HHI Fraunhofer, Germany	Email: luis.oala@hhi.fraunhofer.de
	Christian Johner Johner Institute for Healthcare IT, Germany	Email: christian.johner@johner-institut.de
	Peter. G. Goldschmidt World Development Group Inc., USA	Email: pgg@has.com
	Pradeep Balachandran Technical Consultant (Digital Health), India	Email: pbn.tvm@gmail.com

Contributors: (in alphabetical order)

Aaron Y. Lee
University of Washington
E-mail:leeay@uw.edu

Alixandro Werneck
LeiteMachine Learning Laboratory of Finance and Organizations,
University of Brasilia, Brazil
E-mail: alixandrowerneck@outlook.com

Andrew Murchison
Company Oxford University Hospitals NHS Foundation Trust, United Kingdom
E-mail: agmurchison@gmail.com

AnleLin, Health Sciences Authority, Singapore
E-mail: LIN_Anle@hsa.gov.sg

Christoph Molnar
Expert for Interpretable Machine Learning
Technische Universität München
E-mail: Christoph.molnar@gmail.com

Johannes Tanne
dentalXrai GmbH, Berlin
E-mail:johannes.tanne@dentalxr.ai

Juliet Rumball-Smith
New Zealand Ministry of Health
E-mail: juliet@rumballsmith.co.nz

Pat Baird
Philips, U S A
E-mail: pat.baird@philips.com

Peter. G. Goldschmidt
World Development Group Inc., USA
E-mail: pgg@has.com

Pierre Quartarolo
Danish Medicines Agency
E-mail: jepq@dkma.dk

Shan Xu
China Academy of Information and Communications Technology (CAICT)
China
E-mail: xushan@caict.ac.cn

Sven Piechottka
Open Regulatory GmbH
Germany
E-mail: sven@openregulatory.com

Zack Hornberger
Medical Imaging & Technology Alliance
E-mail: zhornberger@medicalimaging.org

Table of Contents

ITU-T FG-AI4H Deliverable

Good practices for health applications of machine learning: Considerations for manufacturers and regulators

1 Background

This document covers only data driven AI systems and does not take into account the aspects of non-data driven rule based/expert AI systems. Artificial intelligence (AI)-based technologies find extensive use in medical applications and the proliferation of AI-based technologies holds great potential in improving the accessibility, quality, and value of healthcare 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. Therefore, streamlined and systematic regulatory compliance processes can help to expedite regulatory approval and to reduce the time-to-market for these products. AI-based medical devices are, by definition software 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 diagnoses, prevention, monitoring, prediction, prognoses, treatment or alleviation of disease and other conditions.

Machine learning (ML) technologies, a subset of AI technologies have the unique ability to learn from real-world data feedback and adapt their performance over time. The complexity introduced by these technologies is relevant to the clinical safety and performance of the medical device and may introduce new risks or lead to modification of existing risks, which act as barriers to acceptance. Black box AI / ML algorithms that resist comprehensive explanation–also create 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, safety and effectiveness of AI / ML based 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 / ML based medicals devices, which create barriers such as interoperability for these types of devices to scale up at the global level. Many medical devices companies do not have adequate awareness of machine learning best practices and 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 good practice guidelines applicable to AI / ML based devices. There is also need for regulatory policies and guidelines to be tailored for AI / ML based medical devices. The main aim of these good machine learning practice guidelines is to safeguard patient safety as a 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 food and drug administration (FDA).

2 Target of this guideline

2.1 Aims

– To help manufacturers get familiarized with international laws and regulations that applies to AI / ML based medical devices and to bring them to the market quickly and effectively.

– To help internal and external auditors test the legal conformity of AI / ML based medical devices and the associated life-cycle process.

2.2 Objectives

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

– To understand the expectations of the regulatory bodies.

– To promote step-by-step implementation of safety and effectiveness of AI / ML based software-as-medical device.

– To fill the current gap in international AI / ML based medical device standards to the greatest extent possible.

2.3 Target audience

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

– quality assurance auditors / managers

– developers

– testers

– regulatory specialists

– data scientists

– clinical specialists

– physicians

– product managers

– medical device consultants

– risk assessment specialists

– service and support providers

These roles can be found at different organizations:

– Medical device manufacturers

– Sub-contractors, suppliers, service providers (e.g., engineering services)

– Authorities

– Notified bodies

– Policy makers

– Operators e.g., in hospitals

– Research organizations

3 Scope

– This Technical Report defines a set of guidelines intended to guide the developers and manufacturers of healthcare AI solutions with requirements pertaining to good practices and processes for AI / ML based medical devices (AI / ML-MD) development.

– This scope of the guidelines covers only data driven AI systems and does not take into account the aspects of non-data driven rule based/expert AI systems.

– This set of guidelines promotes a common understanding between the manufacturers, the notified bodies, and other pertinent authorities on the best practices to conduct a comprehensive requirements analysis and to streamline the conformity assessment procedures for continual product improvement in an iterative and adaptive manner in conformance to the appropriate standards and regulations.

– This set of guidelines is not intended to be a primer on artificial intelligence health applications or machine learning but is intended to serve as a resource guide for regulators

when shaping regulations pertaining to artificial intelligence / machine learning based medical devices (AI / ML-MDs).

– The regulatory requirements scope of AI / ML-MD pertains only to technical aspects and functional safety and efficacy of its entire product life cycle; and not to commercial or business aspects, such as strategic positioning, market assessment, profitability, etc.

– This set of guidelines is not intended 1) to be comprehensive and/or 2) to replace any regulation, directive, standard, or similar legally binding regulatory framework or guidance document of any geographic jurisdiction.

3.1 Regulatory scope

The guideline addresses medical device and accessories:

– with enforcement of regulations

– without enforcement of regulations.

However, it is not tailored for software applications for non-medical devices, e.g., for:

– healthcare facility administrative support

– maintaining or encouraging healthy lifestyle, behaviour and wellness.

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 and scope of the proposed guidelines is illustrated in Figure 1.

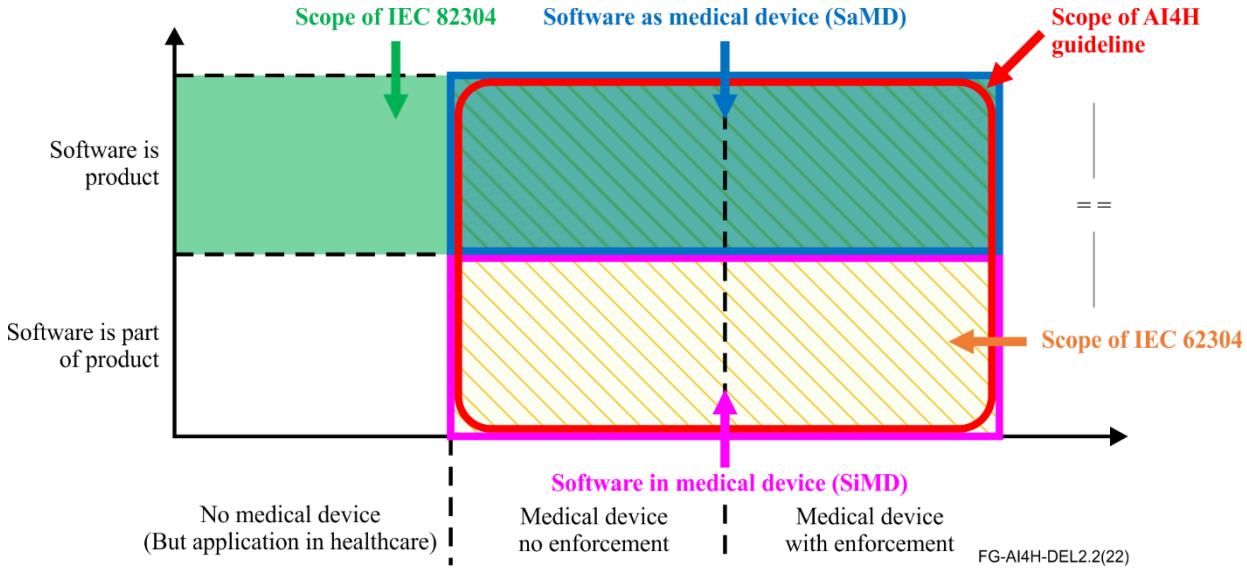

Figure 1 – AI / ML-MD classification criteria and scope

3.2 Product scope

The guideline addresses software that

– is the product (standalone software): Software-as-a-medical device (SaMD)

– is integral (including embedded) in the product: Software-in-a-medical device (SiMD).

It includes both

– static AI model-based systems

– continuous/Incremental learning AI / ML model-based systems.

The scope of the guideline is not limited to a particular software, respectively system architecture. It includes for example:

– Medical devices consisting of components that run on one or on different hardware platforms
– Medical devices that run on mobile and cloud platforms
– Web-based medical devices
– Systems (not necessarily in the sense of the medical device regulation (MDR) article 22) that consist of two (or more) medical devices and/or non-medical devices, and/or medical device accessories.

3.3 Process scope

Within the scope of this guideline are the product related processes, in particular:

– Design and development
– Verification and validation
– Post-market surveillance
– Change (regardless, whether there is significant change or not).

With the exception of the process in maintaining this guideline (as described in clause 14), this guideline does not describe processes related to itself:

– Change of the structure of the underlying data model
– Change of the maintenance process (clause 14.4.1).

Purely quality system related processes as control of documents and records are outside the scope of this guideline as well. Example, this guideline does not provide any guidance on how to change existing documents and records (if the latter is permitted at all).

3.4 Application scope

In healthcare:

– to improve medical outcome (Examples include supporting diagnosis, treatment, prevention, monitoring and prediction of diseases and injuries)
– to improve workflow efficiency (For example, AI recommender systems for 'clinical process' efficiency improvement', NLP pipeline based unstructured clinical data analysis to alert treatment preparations and monitoring of adverse effects, etc.).

4 Future of this guideline

– This is the first edition of the guidelines document. Future editions are planned to update this set of guidelines.
– Clause 13 describes how changes are triggered and processed. Clause 13.3 addresses how to provide feedback.
– Annex D provides a template and further instructions on how to submit feedback.
– Clause 14 proposes a process for assessing and adopting changes and may serve as input for a tool development.

5 References

The following list of reference documents were reviewed as part of a 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. A detailed list of regulatory references considered towards the formulation of the proposed guidelines is included in Annex C: Relationship to other guidelines and standards.

[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*.
[IEC 62304]	IEC 62304:2006/AMD1:2015, *Medical device software – Software life cycle processes – Amendment 1*.
[IEC 62366-1]	IEC 62366-1:2015, *Medical devices – Part 1: Application of usability engineering to medical devices*.
[IEC 82304-1]	IEC 82304-1:2016, *Health software – Part 1: General requirements for product safety*.
[FDA 21 CFR]	FDA 21 CFR – Code of Federal Regulations Title 21, part 820, *Quality System Regulations*.
[FDA SW]	GUIDANCE DOCUMENT (2002), *General Principles of Software Validation; Final Guidance for Industry and FDA Staff*.
[FDA SaMD]	FDA (2021), *Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD)*.
[GDPR]	European Union, *General Data Protection Regulation*.
[GHTF/SG5/N2R8:2007]	Global Harmonization Task Force (2007), *Clinical Evaluation, SG5/N2R8:2007*.
[GHTF/SG1/N071:2012]	Global Harmonization Task Force (2012), *Definition of the Terms 'Medical Device' and 'In Vitro Diagnostic (IVD) Medical Device', GHTF/SG1/N071:2012*.
[IMDRF/GRRP WG/N47]	IMDRF/GRRP WG/N47 Final:2018, *Essential Principles of Safety and Performance of Medical Devices and IVD Medical Devices*.
[EU-MDR (2017/745)]	REGULATION (EU) 2017/745 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL (April 2017), on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC.

6 Definitions

6.1 Terms defined elsewhere

This Technical Report uses the following terms defined elsewhere:

6.1.1 artificial intelligence (AI) [b-ISO/IEC 22989]: Capability of an engineered system to acquire, to process and to apply knowledge and skills (Note 1 to entry: knowledge are facts, information, and skills acquired through experience or education).

6.1.2 AI system [b-ISO/IEC 22989]: Technical system that uses artificial intelligence to solve problems.

6.1.3 clinical evaluation [GHTF/SG5/N2R8:2007]: 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.

6.1.4 in vitro diagnostic (IVD) medical device [GHTF/SG1/N071:2012]: 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.

6.1.5 lifecycle [b-ISO/IEC Guide 51]: All phases in the life of a medical device, from the initial conception to final decommissioning and disposal.

6.1.6 machine learning [b-ISO/IEC 23053]: Process using computational techniques to enable systems to learn from data or experience.

6.1.7 manufacturer [b-ISO 7396-2]: Natural or legal person with responsibility for the design, manufacture, packaging and labelling of a device before it is placed on the market under his own name, regardless of whether these operations are carried out by that person himself or on his behalf by a third party.

6.1.8 medical device [GHTF/SG1/N071:2012]: 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.

6.1.9 process [b-IEC 60050-351]: Complete set of interacting operations in a system by which matter, energy or information is transformed, transported or stored.

6.1.10 product [b-ISO 9000]: Result of a process.

6.1.11 requirement [b-ISO/IEC Guide 2]: Provision that conveys criteria to be fulfilled.

6.1.12 software-as-a-medical device [IMDRF/GRRP WG/N47]: Software intended to be used for one or more medical purposes that perform these purposes without being part of a hardware medical device.

6.1.13 software validation [b-IEEE 610]: The process of evaluating software during or at the end of the development process to determine whether it satisfies specified requirements.

6.1.14 software verification [b-IEEE 610]: The process of evaluating software to determine whether the products.

6.2 Terms defined in this Technical Report

This Technical Report does not define any terms.

7 Abbreviations and acronyms

This Technical Report uses the following abbreviations and acronyms:

AI	Artificial Intelligence
AI4H	Artificial Intelligence for Health
AI / ML-MD	Artificial Intelligence / Machine Learning based Medical Devices
CSV	Computerized Systems Validation
DAISAM	Data and Artificial Intelligence Assessment Methods
EP	Essential Principle
FDA	Food and Drug Administration

GDPR	General Data Protection Regulation
HIPAA	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	Medical Device Regulation
ML	Machine Learning
SaMD	Software-as-a-Medical Device
SiMD	Software-in-a-Medical Device
WG	Working Group
WHO	World Health Organization

8 Conventions

None.

9 Method

This guideline was developed as follows:

– Identification of processes that must be covered (e.g., development and post-market surveillance)
– Collection of all potentially relevant sources (laws, guidelines, standards, best practices) by literature search and by expert interviews (AI, regulatory affairs, quality management)
– Analysis of these sources and extraction of requirements
– Consolidation of these requirements and alignment of degree of abstraction
– Adding specific "sub-requirements" (in the following referred to as "checklist items")
– Adding specific examples
– Consolidation of all the input in a tabular structure
– Adding front matter

10 General requirements

10.1 Process requirements

Table 1 – Process requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s)applicable
PROC-1	The manufacturers should establish a quality management (QM) system that covers all life cycle phases.	– There is at least one SOP[1] covering the design and development process including verification and validation. – There is/are SOP(s) covering the post-market surveillance and vigilance. – There is an SOP covering risk management. – There is an SOP covering computerized systems validation (CSV). – There is an SOP covering the data management (process). – There is/are SOP(s) covering software delivery, service, installation and decommissioning. – There is an SOP covering customer communication including handling of customer complaints.		[EU-MDR (2017/745)] Article 10.9 [ISO 13485] e.g., clause 7.1 [ISO 13485] clause 4.1.6 [FDA 21 CFR] part 820 Good machine learning practices (GMLP) guiding principle (2) (by the Food and drug administration (FDA) et al.)

[1] Standard operating procedure. All SOPs have to be approved and be under version control.

Table 1 – Process requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s)applicable
PROC-2	The manufacturer should compile all product specific plans as required by the 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.		[EU-MDR (2017/745)] Annex I (3) [EU-MDR (2017/745)] Annex III [IEC 62304] clause 5.1 [ISO 14971] clause 4.2 [b-21 CFR 820.30] (b) [FDA SW] validation guidance 5.2.1 GMLP guiding principle (2) (by FDA et al.)

10.2 Competency requirements

Table 2 – Competency requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
COMP-1	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 and 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	[ISO 13485] clause 5.5.1 [ISO 13485] clause 6.2 [EU MDR (2017/745)] Article 10 (9) [21 CFR 820.30] (b) [FDA SW] validation guidance 5.2.1.

Table 2 – Competency requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
			– service technicians e.g., update, upgrade, configuration, installation, capturing audit logs, etc. – support staff.	
COMP-2	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.	[ISO 13485] clause 6.2. [ISO 14971] clause 4.3 [ISO 13485] clause 7.3.2 [IEC 82304-1] clause 6.1 GMLP guiding principle (1) (by FDA et al.) FDA: Culture of quality and organizational excellence: "Continuous development of employees through robust knowledge management, employee development options, coaching, training, and succession planning." (software pre-cert programme.)

11 Pre-market requirements

11.1 Intended use and stakeholder requirements

11.1.1 Intended medical purpose

Table 3 – Intended use requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
USE-1	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 the International Classification of Diseases ICD-10 codes (at least 3 digits). Increasing adherence is an example for improving treatment. The description answers questions such as: – Is it a self-contained device with application or an operational supporting system? – Is it health related or operations support?	[EU-MDR (2017/745)] Annex II (1.1) [ISO 13485] clause 4.2.3 [ISO 14971] clause 5.2 [b-21 CFR 814.20] (b)(3)(i) [b-21 CFR 820.30] (c) GMLP guiding principle (6) (by FDA et al.)
USE-2	The manufacturer should specify other positive impacts on health care.		– Faster patient care e.g., treatment, diagnosis. – Reductions in workload. – Reductions in costs of healthcare.	[b-MEDDEV 2.7/1] revision 4 [EU-MDR (2017/745)] Annex I (23.4) FDA guidance on "Factors to consider when making benefit-risk determinations in medical device premarket approval and de novo classifications".

Table 3 – Intended use requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
USE-3	The manufacturer should specify the target patients.	There is a documented specification of: – demographics (e.g., age, sex) – Indications – contraindications – co-morbidities.	Comment: The intended use has to be specified in relevant detail for all pertinent aspects. Certain derivative requirements would pertain only to the specified uses. E.g., if the product is intended to support diagnosis in white women, there is no need to know the product's performance in black men. There might be a requirement to warn users to restrict use of the product to white women.	[EU-MDR (2017/745)] Annex I (23.4) [EU-MDR (2017/745)] Annex II (1.1) [IEC 62366-1] clause 5.1 [b-21 CFR 814.20] (b)(3)(i) FDA guidance on "Factors to consider when making benefit-risk determinations in medical device premarket approval and de novo classifications" GMLP guiding principles (3), (5), (6), and (8) (by FDA et al.)
USE-4	The manufacturer should specify the intended part of body or type of tissue the medical device shall interact with.			[IEC 62366-1] clause 5.1.
USE-5	The manufacturer should specify the operating principle.	– There is a description of the task the ML-model may perform. – There is a specification of the type of machine learning. – There is a description whether an intervention of the user before treatment or diagnosis is necessary, possible, not possible. – There is a clarification whether the AI can trigger an autonomous	Typical tasks include: – segmentation – detection – decision support – recommendation – process automation – search (e.g., similarities). Typical dimensions include:	[IEC 62366-1] clause 5.1 [EU-MDR (2017/745)] Annex II (1.1) [b-21 CFR 814.20] [b-XAVIER University] "Building explainability and trust for AI in healthcare" [FDA SaMD] Proposed regulatory framework for modifications to artificial intelligence/machine

Table 3 – Intended use requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
		action / decision or just provides information for human analysis. – There is a description of the level of independence.	– Type of learning (supervised, unsupervised, semi-supervised, reinforcement) – Time and type of learning (before placing on the market→ locked algorithm, during use, globally, per product instance, per hospital) – Technical task (classification, regression, clustering, control). XAVIER differentiates these user interactions: – intervention before treatment or diagnosis is not possible – intervention before treatment or diagnosis is possible by overriding – intervention before treatment or diagnosis is necessary by approval – there is no direct diagnosis or treatment possible with the system.	learning (AI/ML)-based software as a medical device (SaMD).
USE-6	The manufacturer should provide explicit task description by distinguishing it from the particular algorithm used.	– Background information, including a review of the evidence, the purpose of the task, all relevant definitions, and discussion of limitations and special cases.		

Table 3 – Intended use requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
		– A thorough description of the diagnostic task, including criteria for making the clinical assessment, descriptions and definitions of the measurement, or a description of all classification categories. – Detailed image labelling instructions for the task, including specific labelling strategies and relevant pitfalls. – Illustrated prototypical examples and relevant counter - examples, such as an atlas.		

11.1.2 Intended users and context of use

Table 4 – Intended users and intended context of use specification

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
ENV-1	The manufacturer should characterize the intended 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 may include: – education – experience in medical domain – technical skill knowledge – training to be accomplished	[EU-MDR (2017/745)] Annex I (5) [EU-MDR (2017/745)] Annex II (1.1) [IEC 62366-1] clause 5.1 [b-XAVIER University] "Building explainability and trust for AI in healthcare"

Table 4 – Intended users and intended context of use specification

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
			– 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.	[b-FDA HFE] guidance "Applying human factors and usability engineering to medical devices" (clause 5.1) GMLP guiding principle (7) (by FDA et al.)
ENV-2	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:	[EU-MDR (2017/745)] Annex I (5) [IEC 62366-1] clause 5.1 [b-XAVIER University] "Building explainability and trust for AI in healthcare" [b-FDA HFE] guidance "Applying human factors and usability engineering to medical devices" (clause 5.2) [b-ISO 13407] Human-centred design processes for interactive systems.

Table 4 – Intended users and intended context of use specification

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
			– wearing of gloves or other personal protection equipment – usage of tools – physical stress.	
ENV-3	The manufacturer should specify the product lifetime.		The product lifetime may depend on: – technologies applied in the product – technical environment such as operating systems, browsers, networks – development of the state of the art, e.g., progress in medical research – competitive products.	

11.1.3 Stakeholder requirements

Table 5 – Stakeholder requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
STKH-1	The manufacturer should operationalize the goals listed in the intended use with quantitative values for the 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:	[EU-MDR (2017/745)] Annex I (23.4) [EU-MDR (2017/745)] Annex III (1.1)

Table 5 – Stakeholder requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
			– 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. Comment: For different sub-groups there might be different specifications of performance requirements.	[FDA SW] validation guidance 5.2.2 FDA guidance for the content of premarket submissions for software contained in medical devices (s. "Software requirements specification").
STKH-2	The manufacturer should specify the runtime environment of the product regarding hardware and software.	– It is specified whether the software runs inside a medical device, as a mobile application, as a wearable device, as a desktop application, in the cloud or another environment. – The minimum hardware requirements are specified. – The minimum software requirements are specified.	Hardware requirements may include: – CPU – RAM – screen size, resolution and orientation – physical storage – network connectivity e.g., bandwidth, latency, reliability – required peripherals such as printers, scanners, input devices – Sensors – Energy source – Periphery (keyboard, mouse, etc.) – AI acceleration hardware / inference acceleration hardware Software requirements may include: – operating system (including the version) – browser (type, version)	[ISO 13485] clause 7.3.3 [EU-MDR (2017/745)] Annex 1, 17.3 and 17.4 [IEC 62304] clause 5.2 [b-XAVIER University] "Building explainability and trust for AI in healthcare" [FDA SW] validation guidance 5.2.2 FDA guidance for the content of premarket submissions for software contained in medical devices (s. "Software requirements specification").

Table 5 – Stakeholder requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
			– virtualization (e.g., Java Runtime Environment, NET, Docker, virtual machines).	
STKH-3	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 – wireless communication protocols (Bluetooth, 4/5G cellular, Wi-Fi, etc.) – physical hardware connections (e.g., USB) 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 – information exchange e.g., FHIR, DICOM, HL7, etc.	[IEC 62304] clause 5.2 FDA guidance for the content of premarket submissions for software contained in medical devices (s. "Software requirements specification").
STKH-4	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: – ranges – data types – sensor requirements – type of data capturing device	[IEC 62304] clause 5.2 [ISO 14971] clause 5.3 FDA guidance for the content of premarket submissions for software contained in medical

Table 5 – Stakeholder requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
			– precision of data – size/ quantity of data – type and technical parameters of recording procedure (e.g., strength of magnetic field, number of electrodes – frequency of data.	devices (s. "Software requirements specification").
STKH-5	The manufacturer should determine the regulatory requirements.	– There is a list of countries / markets that the product may be placed 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., [EU-MDR (2017/745)], IVDR.	[EU-MDR (2017/745)] Annex IX (2.2) [ISO 13485] (clauses 5.2 and 7.2.1.)

11.1.4 Risk management and clinical evaluation

Table 6 – Inputs to risk management and clinical evaluation requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
RSK_MGNT_1	The manufacturer should evaluate alternatives to the given product (e.g., other products, procedures, technologies) and establish the necessity for using machine learning models.	– There is a clinical evaluation. – The clinical evaluation lists alternative products, technologies and/or procedures. – There is a search protocol that reveals how the manufacturer searches for alternatives.	Alternative technologies might include: – other ML models – non-ML methods e.g., classical algorithms.	[b-MEDDEV 2.7/1] [EU-MDR (2017/745)] Annex I 1. [ISO 14971] clauses 4.2 and 10 FDA guidance on "Factors to consider when making

Table 6 – Inputs to risk management and clinical evaluation requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
		– The clinical evaluation assesses 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.		benefit-risk determinations in medical device premarket approval and de novo classifications" (e.g., Part C).
RSK_MGNT_2	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 the hazards and related harms with related probabilities and severities resulting from the ML models not meeting the requirements. – There is a FMEA that analysis the effects of ML models that do not meet the performance requirements.	Performance requirements might include: – accuracy – specificity – sensitivity – response times – robustness – other Comment: Differential performance by patient demographics.	[ISO 14971] clauses 5.4 and 5.5 [EU-MDR (2017/745)] Annex I (3) DIN SPEC 2 [b-ISO/TR 31004] – Risk management – Guidance for the implementation of ISO 31000.
RSK_MGNT_3	The manufacturer should analyse the reasonably foreseeable risks.	The risk management file analysis the risks associated with: – non-specified users – non-specified use environment – application of the product for patients other than those which are specified – reasonably foreseeable misuse	Non-specified users: – other professions e.g., nurse instead of a physician – missing training. Other patients: – different age, sex, race – other co-morbidities	[EU-MDR (2017/745)] Annex I (14.2) (d) [ISO 14971] clause 5.2 DIN SPEC 2 [b-21 CFR 820.30] (g) FDA guidance on design considerations and premarket submission

Table 6 – Inputs to risk management and clinical evaluation requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
		– hardware failure.	– different severity of disease or injury Comment on 'reasonably foreseeable risks': This is significant. Particularly with respect to perpetuating / maintaining historical biases in treatment / service according to race / ethnicity, or historical issues around systematic misdiagnosis or under investigation in certain groups.	recommendations for interoperable medical devices [b-IEC 31010] – Risk management - Risk assessment techniques.

11.2 Product and software requirements

11.2.1 Functionality and performance

Table 7 – Functionality and performance requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
FNCT-1	The manufacturer should derive traceable quantitative quality criteria and requirements for the software and/or the algorithm from the intended use and from the stakeholder requirements.	– There is a specification of quantitative minimum 'quality criteria'. – There is 'traceability matrix' that links the intended use with the quantitative quality product requirements.	'Quantitative quality criteria' may include the following: – for classification problems: o accuracy (mean or balanced accuracy) o positive and negative predictive value in the intended use population	[IEC 62304] clause 5.2 [ISO 13485] 7.3.3 [b-XAVIER University] "Building explainability and trust for AI in healthcare"

Table 7 – Functionality and performance requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
			○ specificity and sensitivity ○ F1 score area under the ROC curve (AUC) ○ Area under the precision-recall curve – for regression problems: ○ mean absolute error ○ mean square error ○ Bland-Altman plot-difference measure (for bias estimation) NOTE – Further metrics can determine how stable ("non-distractible") and deterministic the model must be. 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.	[FDA SW] guidance on software validation clause 5.2.2.

Table 7 – Functionality and performance requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
			Comment: They should be able to show that the model is equally accurate for different groups of patients. NOTE – "Stability" can be understood from the point of view of: – conditioning analysis (conditioning number via Eigen values) – functional analysis (e.g., through Lipschitz continuity) – epsilon-based robustness as in adversarial research – robust statistics as proposed by Huber, P. J.	
FNCT-2	The manufacturer should derive non-functional requirements from the intended use and stakeholder requirements.	There is a specification of non-functional requirements such as: – repeatability / reproducibility – response times – data volumes to be handled – availability – security e.g., access restrictions.	Self-tests can be a mean to verify the repeatability of a system. The specification of response times might depend on the number of users, number of transactions, frequency and amount of input data, etc. Availability can be expressed as a percentage of time, percentage of usages or as meantime between failure.	[ISO 13485] clauses 7.2.1 and 7.3.3 [b-ISO/IEC 25010] [EU-MDR (2017/745)] Annex I (17.1) [b-XAVIER University] "Building explainability and trust for AI in healthcare" [FDA SW] guidance on software validation clause 5.2.2 Annex C.7 - IT security guidelines.

Table 7 – Functionality and performance requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
FNCT-3	The manufacturer shall derive from the risk analysis product / software requirements to minimize risk.	There is a risk table correlating risks and measures.	The risk table sometimes is referred to as the "FMEA table". Examples for measures are: – Kill switch, overruling through human intervention – Redesign of the user interface (s. user interface requirements) – Locked algorithm instead of continuous learning system – Restriction of intended use – Validation of input data (see next requirement) – Backup, recovery – Redundant design, failover system e.g., without ML functionality.	
FNCT-4	The manufacturer should specify how cybersecurity risk management was incorporated in the device development lifecycle and what risk controls were implemented to ensure that all the interfaces of the product and its communication channels are secured from potential cyber threats. Note that the scope includes protection of the software and protection of the datasets.	There is specification on – list of steps on how to identify and evaluate threats and vulnerabilities, control security risks, and monitor the efficacy of these controls – repeatable, reproducible, testing-oriented criteria to assess a device's cyber vulnerabilities, fight malware, and test the security measures.		[b-AAMI TIR57] Technical information report 57 (TIR57), "Principles for medical device security – Risk management" [b-UL Standard 2900-1]. Standard for software cyber-security for network-connectable products. Annex C.8 – Cybersecurity

Table 7 – Functionality and performance requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
				[b-OECD] Guidelines for the security of information systems and networks: Towards a culture of security.
FNCT-5	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: – adulterated data (integrity problem) – conflicting data – incomplete data sets – missing data, empty data, lack of data sets – wrong data format – excessive data quantities (amount, frequency) – data outside of the specified value ranges – wrong temporal sequence of data, etc.		[b-ISO/IEC 25010] [IEC 62304] clause 5.2 [ISO 14971] clause 5.4 [FDA SW] guidance on software validation clause 5.2.2 [b-FDA Digital health] criteria.
FNCT-6	For continuous learning systems the manufacturer should specify the frequency of the algorithms updates.	It is specified what triggers updates.	Triggers include: – on availability of enough data – periodically – if a minimum change to the algorithm / output is exceeded continuously.	[b-XAVIER University] "Building explainability and trust for AI in healthcare" [FDA SaMD] Proposed regulatory framework for modifications to AI/ML

Table 7 – Functionality and performance requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
				based SaMD, e.g., Appendix B.
FNCT-7	For continuous learning systems the manufacturer should specify how quality control of new data is performed.	There is a specification on how data are cleaned e.g., by: – correction – omission – user interaction.		[b-XAVIER] "Perspectives and good practices for AI and continuously learning systems in healthcare" [FDA SaMD] Proposed regulatory framework for modifications to AI/ML based SaMD, ACP e.g., page 11.
FNCT-8	For continuous learning systems the manufacturer should specify a range within changes to the algorithm and to the system output that are permitted.	There is a description of how: – algorithms are changed over time – the amount of change is quantified – these changes relate to changes to the output.	For example, a change to a neural network can target: – fit parameters such as weights of neurons or cut-off of the activation function – hyper-parameters such as numbers of neurons per layer and number of layers. NOTE – CL systems exhibit "drift" as a learning. They can eventually fall into very different local minima than the original model.	[b-XAVIER] "Perspectives and good practices for AI and continuously learning systems in healthcare" [FDA SaMD] Proposed regulatory framework for modifications to AI/ML based SaMD, SPS e.g., page 10.
FNCT-9	For continuous learning systems the manufacturers should specify how changes to the algorithm are controlled.	There is a specification of: – system self-checks on performance	The decision whether to enforce, prevent, delay or roll-back changes by users or the manufacturer must be taken risk-based.	[b-XAVIER] "Perspectives and good practices for AI and continuously learning systems in healthcare"

Table 7 – Functionality and performance requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
		– functionality to enforce, prevent, delay or roll-back changes to algorithms – change reports, change / audit-logs – control of versions of the algorithms – boundaries of autonomous learning.	The version control must apply to the entire model including fit parameters, hyper-parameters, and model architecture with respective time stamps.	[b-ISO/IEC TR 24028] [FDA SaMD] Proposed regulatory framework for modifications to AI/ML based SaMD, ACP e.g., page 11.

11.2.2 User interface

Table 8 – User interface requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
UI-1	The manufacturer should specify what the user interface must display in case of error if the inputs do not meet the specified requirements.	There is specification of the user interface in case of: – incorrect data inputs (s. "The manufacturer should determine how the system behaves if the inputs do not meet the specified requirements") – internal errors.	See previous checklist item. UI output display modes may include the following: – warning – alert – caution – meantime between failure, etc.	[EU-MDR (2017/745)] Annex I clause 5 [IEC 62304] clause 5.2 [b-FDA HFE] guidance [FDA SW] guidance on software validation e.g., clause 5.2.3.
UI-2	For continuous learning systems the manufacturer should specify how the	There is a specification of user interface parts that provide:	NOTE – For the user to have the option to reject, delay or roll-back an algorithm change, all the previous versions of the	[b-XAVIER] "Perspectives and good practices for AI

Table 8 – User interface requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
	user is informed about significant changes to the algorithms.	– information that an algorithm change was performed or will be performed – the user the option to reject, delay or roll-back an algorithm change.	ML model would need to be maintained	and continuously learning systems in healthcare" [FDA SaMD] Proposed regulatory framework for modifications to AI/ML based SaMD, ACP e.g., page 11.
UI-3	The manufacturer should determine whether there is a need for instructions for the use and training materials.	Either there is an instructions-for-use (IFU) or the user risk analysis reveals no risks that can be further mitigated by an IFU.		[EU-MDR (2017/745)] Annex I (23) Federal Food, Drug, and Cosmetic Act (FD&C Act), [b-21 CFR 801] and [b-21 CFR 820.120] [ISO 13485] clause 4.2.3

11.2.3 Additional software aspects

Table 9 – Additional software requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
SFTW-1	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	[EU-MDR (2017/745)] Annex I (17, 18, 23.4) [IEC 62304] clauses 5.2, 5.3 and 7.1 [ISO 14971] clause 5.4 [FDA SW] guidance on software validation e.g.,

Table 9 – Additional software requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
			– resource overload such as out of memory errors – lack of access to resources such as databases – compromised integrity of data and program code.	clauses 5.2.2, 5.2.3 and 5.2.4. GMLP guiding principle (2) (data integrity) (by FDA et al.)
SFTW-2	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.		Article 22 of the General Data Protection Regulation (GDPR). Exceptions of Article 22 section 2 may apply.
		–		

11.2.4 Risk management

Table 10 – Risk management and clinical evaluation requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
RSK_MGNT_4	The manufacturer should assess the risks arising if the inputs do not meet the specified requirements.	– There is an assessment which inputs and combinations of inputs of the input space that have an (undesirable) impact on the system's output. – 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	[ISO 14971] clause 5.4 [IEC 62304] clause 7.1 DIN SPEC 2 [IEC 82304-1] clause 4.1.c.)

Table 10 – Risk management and clinical evaluation requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
		– There is an assessment of values or ranges for quality metrics that have to be met in order to avoid unacceptable risks. – The risk analysis considers a drift in data. – The risk analysis assesses risk by lack of robustness e.g., for adversarial attacks. – There are adversarial examples defined that lead to unacceptable risk and that have to be evaluated in testing.	– unreasonable combination of data (feature) – wrong meta-data – data drifts can be identified by mean values and distributions. Critical drifts can occur either in single features or combinations of features – use of synonyms in texts – typing errors – malicious attacks e.g., by manipulating a few pixels in images.	
RSK_MGNT_5	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.	[b-XAVIER University] "Building explainability and trust for AI in healthcare" [FDA SaMD] Proposed regulatory framework for modifications to AI/ML based SaMD, "reference standard" GMLP guiding principle (5) (by FDA et al.)
RSK_MGNT_6	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'.		[ISO 14971] clause 5.3 [IEC 62304] clause 7.1 [IEC 82304-1] clause 4.1.c)

Table 10 – Risk management and clinical evaluation requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
				[b-XAVIER University] "Building explainability and trust for AI in healthcare" [FDA SW] guidance on software validation e.g., clause 5.2.2.
RSK_MGNT_7	The manufacturer should assess the consequences if the system provides socially unacceptable / discriminatory outputs.	There are outputs that an assessment reports on the 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 the criticality of the clinical use case and environment.		[b-Ethics AI] Ethics guidelines for trustworthy AI.
RSK_MGNT_8	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 – software using more CPU, GPU, RAM, I/O, bandwidth than specified.		[EU-MDR (2017/745)] Annex I (17.2) [ISO 14971] clause 5.3 [IEC 62304] clause 7.1 [FDA SW] guidance on software validation e.g., clause 5.2.2.

Table 10 – Risk management and clinical evaluation requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
RSK_MGNT_9	The manufacturer should analyse risks if (run-time) environment does not meet the specifications.	The risk analysis assesses risk from: – insufficient or faulty hardware – software environment not meeting the specifications – network environment not meeting the specifications – interfaces not meeting the specifications.	Hardware related risks: – CPU, RAM, I/O, hard disk space not as specified – memory, CPU, GPU flaws – hard disk full – RAM, CPU, I/O overutilization by other applications. Software related risks: – other type or version of operating system, browser, virtualization layer (.NET, JRE, VM), libraries – software patches not installed – software bugs. Network related risks: – bandwidth, latency not as specified – endpoints, protocols not supported or blocked. Interface related requirements: – S. wrong input data – unspecified data volumes and frequencies.	[EU-MDR (2017/745)] Annex I e.g., mobile platforms, network characteristics, e.g., 14.2.(d) DIN SPEC 2
RSK_MGNT_10	The manufacturer should identify use related risks.	The risk analysis assesses risks caused by users:	– User does not update the system.	[IEC 62366-1], clause 5.3 *f.* [b-FDA HFE] guidance

Table 10 – Risk management and clinical evaluation requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
		– not following the instructions for use – not understanding warnings and explanations – using the system in a technical social environment that does not meet the specifications.	– User installs software on wrong. – User connects product to other systems not meeting requirements.	[FDA SW] guidance on software validation e.g., clause 5.2.3.
RSK_MGNT_11	The manufacturer should analyse risks from malicious / adversarial attacks	– There is an analysis of potential attackers and motivation – There is a list of attack vectors – There is a vulnerability analysis.	– Potential attacks include manipulating input data such as images or even of (public) training data ("poisoning attack") – The vulnerability increases if the attacker has access to the model internals (e.g., architecture) or even to the model itself. Also, the chance of accessing the model via an API and thereby evaluating different attacks increases the vulnerability	
RSK_MGNT_12	With continuous learning systems, the manufacturer should mitigate risks that are specific to continuously learning systems.	– The risk analysis assesses risks that are specific to continuous learning systems. – The risk management file specifies the respective risk mitigation.	Examples of risk mitigation: – option to reset the systems – self-tests.	[EU-MDR (2017/745)] Annex I (17) [FDA SaMD] Proposed regulatory framework for modifications to AI/ML based SaMD (entire document).

Table 10 – Risk management and clinical evaluation requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
RSK_MGNT_13	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.		[ISO 14971] clause 6 DIN SPEC 2 FDA guidance on determining benefit risk.
RSK_MGNT_14	The manufacturer should mitigate risks.	There is a risk mitigation for risks caused by: – input data not meeting the requirements – inability of the system to meet the non-functional requirements – ML algorithms not meeting the quality metrics – Adversarial attacks – software bugs. The measures implemented are: – The measures are specified as product or component requirements – There are tests verifying / validating the implementation and effectiveness of these measures.	Means for risk mitigation might include: – System shutdown – Warnings to users, alarm systems – Validation of input data – Self-tests – Robustness: Adversarial training (*arXiv:1706.06083), generative methods (See section "uncertainty" in DAISAM paper)* – Adversarial attacks: Training with adversarial data sets or operating with different classifiers or learning invariant transformation of feature	[ISO 14971] clause 7 DIN SPEC 2 Data and artificial intelligence assessment methods (DAISAM) [FDA SW] guidance on software validation e.g., clause 6.1 [b-FDA HFE] guidance e.g., clauses 8.1.3 and 8.1.4 [b-ISO/IEC TR 24028] clauses 10.4, 10.5, 10.7 *ff*.
RSK_MGNT_15	The manufacturer should repeat these risk management activities after training of the model as well as prior to product release.			

11.3 Data management requirements

11.3.1 Data collection

Table 11 – Data collection requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standards / Regulations applicable
DAT_CL-1	The manufacturer should specify the number of required data sets.	– There is a specification of number of data sets. – There is a rationale for this number.	The division into training, test and validation data sets is scope of chapter 12.4.1.	[ISO 13485] clause 7.3.7 [FDA SaMD] Proposed regulatory framework for modifications to AI/ML based SaMD: "statistical analysis plan".
DAT_CL-2	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. – There is – if applicable – a specification for the timeframe within data must be collected.	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 – character encoding – sampling rates – image parameters such as compression, image sizes, resolution, colour coding, zoom – language	[b-ISO/IEC TR 24028] [b-XAVIER University] "Building explainability and trust for AI in healthcare."

Table 11 – Data collection requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standards / Regulations applicable
			Inclusion / exclusion criteria of patient data may include the following attributes: – demographic data (age, gender) – physical parameters (height, weight) – diseases – vital parameters – laboratory parameters – presence of additional tests – case history – special conditions (e.g., patients having heart pacemaker or lung surgery).	
DAT_CL-3	The manufacturer should specify quality control of data.	– There is a list of allowed / expected data sources. – There is a specification of data source requirements. – There is a description on how invalid input data are identified and excluded. – There is a validation of surveys (justify the selection of the surveys, the time of survey and possibly the method of assessment, in particular if no standardized survey exists).	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. Invalid data may be caused by:	[b-ISO/IEC TR 24028] [b-PSO NAVIGATOR] [b-OECD PF] Privacy framework GMLP guiding principle (2) (data management) (by FDA et al.)

Table 11 – Data collection requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standards / Regulations applicable
			– violation of inclusion and exclusion criteria – mistyping e.g., in EMRs, confusion of patients or form fields, misunderstanding of information that has to be entered – different coding of data (It is not uncommon that hospitals apply coding rules differently e.g., for reimbursement reasons) – different units (e.g., kg for babies and pounds for adults). Survey methods may include the type of questions, the types of answers, the decision to have open or closed questions, etc.	
DAT_CL-4	The manufacturer shall analyse the factors that might cause a bias.	– There is a list of potential biases.	Analysis can be performed (and visualized) by: – Directed acyclic graphs – QUADAS-2 ([b-Whiting, 2011]), and PROBAST ([b-Moons, 2019]) – DeLong test – FairML (python toolbox) – AI Fairness 360 toolkit – Conditional generative adversarial networks (GANs) Factors causing biases include:	[b-ISO/IEC TR 24028] e.g., clause 10.5 DAISAM GMLP guiding principle (3) (bias) (by FDA et al.)

Table 11 – Data collection requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standards / Regulations applicable
			– non representative patient population e.g., volunteers, sex, race, age, size, weight, diseases, treatments, social and geographic environment – data collection e.g., types of questionnaires or using channels (e.g., social media) predominantly by certain groups – attributes that are irrelevant for the expected output – confusion of correlation and causation – preparation of source data e.g., histopathological slides – specific data sources e.g., different types, accuracy – location of data collection e.g., size and type of hospital, rural versus urban – Aggregation that combines data that are not representative for the single population – "Over-curation" e.g., excluding data from poor quality MRI scans that, however, are common. "Over-curation" also might exclude certain patient profiles	

Table 11 – Data collection requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standards / Regulations applicable
DAT_CL-5	The manufacturer should specify a distribution of input data that is representative for the target system / population.	– There is a specification of the distribution of relevant patient characteristics.	Characteristics can include: – demographics: age, sex, race – health status, comorbidities – social status, education – motivation to participate in studies. NOTE – Even if all individual data sets meet the specification, still the distribution of data might not be representative and/or cause a bias.	[b-ISO/IEC TR 24028] e.g., clause 9.8.1 DAISAM GMLP guiding principle (3) (by FDA et al.)
DAT_CL-6	The manufacturer should validate that the test and training data meet the specified criteria.	– There is a description on how ensured are 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).	[b-ISO/IEC TR 24028].

Table 11 – Data collection requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standards / Regulations applicable
DAT_CL-7	The manufacturer should ensure data protection.	– There is a documented patient data protection policy. – This policy describes the roles (persons, systems) – There is a documented description which roles have which type of access (e.g., via user interface, APIs etc.) to which data with which rights (create, delete, change, read) – There should be a documented procedure for data anonymization / pseudonymization. – The policy describes how to decommission data – Data scientists do not have access to protected data. – There is a data protection officer. – There is an ethical approval e.g., for genetic data if legally required.	Data could be derived from machine-to-machine (M2M) communication as well.	[GDPR] Health Insurance Portability and Accountability Act ([HIPAA]) [b-ISO/IEC TR 24028] clause 10.6 [b-ISO/IEC 20889] (data de-identification.)

11.3.2 Data annotation

Table 12 – Data annotation requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
DAT_AN-1	The manufacturer using "supervised learning" should derive the labels from	There is specification for "label" selection criteria in case of		

Table 12 – Data annotation requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
	the intended use and justify this selection.	"supervised learning" based machine learning task.		
DAT_AN-2	The manufacturer using "supervised learning" should have 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. – The procedure specifies the data format and/or syntactic and or standards (e.g., coding system) for annotations. – There is a detailed instruction for the task including background information and prototypical examples.	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. DAISAM addresses "label bias".	[ISO 13485] clause 4.1
DAT_AN-3	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'.	The results of the monitoring of the labelling can be used to continuously verify the fitness of persons responsible for labelling.	[ISO 13485] clauses 6.2 and 7.3.2 [FDA 21 CFR] part 820.25 (Personnel.)

Table 12 – Data annotation requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
		– 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.		

11.3.3 Data pre-processing

Table 13 – Data pre-processing requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
DAT_PR-1	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	Data pre-processing steps may include the following: – conversion – transformation – aggregation – normalization – format conversion – calculation of feature – conversion of numerical data into categories, etc. – statistical analysis e.g., descriptive statistics	[ISO 13485] clauses 4.1.6, 7.3.6 [b-XAVIER University] "Building explainability and trust for AI in healthcare" [FDA 21 CFR] part 820.70 (i) Automated processes.

Table 13 – Data pre-processing requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
		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. – This procedure describes how data for training, testing and validation are kept separately. – This procedure describes how new data can be added after initial processing already has been performed (if applicable). – The procedure describes how uniqueness of data is ensured.	"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 (minimum/maximum), 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. – Uniqueness of data is for example data that is not imported twice accidentally.	
DAT_PR-2	The manufacturer shall analyse and mitigate all risks caused by data processing	– There is a list of factors that can cause distortion and perturbation of data.	Examples for factors causing distortion and perturbation of data are:	Artificial intelligence for health (AI4H)-DAISAM.

Table 13 – Data pre-processing requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
			– Rounding errors – Compression, decompression – Noise reduction, filtering – Normalization, transformation, – Resampling – Dealing with outliers, missing values, handling of artefacts	

11.3.4 Documentation and version control

Table 14 – Documentation and version control requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
DOC_VC-1	The manufacturer should describe and control all data processing steps.	– There is a list of data sources. – The document describes all the data processing steps mentioned in the previous clause. – 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 document describes how all data can be traced back to its source.	The description of data sources might include: – location (e.g., clinic) – capture device. The procedure might specify conventions for: – file formats and types – file names – character encoding Means to understand and reproduce the data processing and to prove compliance are: – Audit logs – Intermediary data sets	[b-ISO/IEC TR 24028] clause 9.8.2.2.

Table 14 – Documentation and version control requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
		– The document describes how the validity of personnel operation is ensured. – The document describes how compliance with the data collection, annotation and pre-processing requirements in clauses 11.3.1 to 11.3.3 is verified.	– File name conventions – Application of version control – Regression testing with sample data.	
DOC_VC-2	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.	[ISO 13485] clauses 4.1.6, 4.2.4 and 7.5.6 [FDA 21 CFR] part 820.70 (i) Automated processes.
DOC_VC-3	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 is actually under version control. – The software libraries and frameworks are identified and under version control.		[IEC 62304] clause 8 [FDA SW] guidance on software validation e.g., clause 5.2.1.
DOC_VC-4	The manufacturer should put all training, test and validation data under version control.	– The version of data is aligned with the corresponding software versions (software for processing and product).		[ISO 13485] clause 4.2.5.

Table 14 – Documentation and version control requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
DOC_VC-5	The manufacturer shall protect all data and code from loss and unwanted changes.	– There is a documented procedure for backups and restoring – There are backup records.		[ISO 13485] clause 4.2.5.

11.4 Model development requirements

11.4.1 Model preparation

Table 15 – Model preparation requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
MD_PRP-1	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 an analysis of feature dependencies.	Dependencies can be visualized e.g., with a directed acyclic graph (DAG).	[b-ISO/IEC TR 24028] clause 9.8.2.2.
MD_PRP-2	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 into 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).	Example, for data with rare features or labels, it may be necessary to distribute the data not just at random. An example for an object can be a CT scan. The images of one series should not be distributed into the three different "buckets". The splitting strategy for time series data must ensure correct chronological order.	[b-ISO/IEC TR 24028] clause 9.8.2.1 DAISAM GMLP guiding principle (4) (by FDA et al.)

Table 15 – Model preparation requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
		– There is a justification if data are not distributed at random.		
MD_PRP-3	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 the test data.		GMLP guiding principle (4) (by FDA et al.)

11.4.2 Model training

Table 16 – Model training requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
MD_TR-1	The manufacturer should document model specific data processing.	– The document describes which feature has been recorded specifically for a model or technology.	Examples of this are normalization, selection of class labels (e.g., 0 or 1), selection of column names and distribution of categorical values over multiple columns.	[ISO 13485] clause 4.1 [FDA 21 CFR] part 820.70 (i) Automated processes.
MD_TR-2	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 justify 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.		
MD_TR-3	The manufacturer should avoid over-fitting.	– There is a policy forbidding the use of test data to optimize the	Visualization (e.g., learning curves) might be helpful for	[b-ISO/IEC TR 24028] clause 9.8.2.23.

Table 16 – Model training requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
		model (only training and validation data may be used).	justification and to illustrate the impact of hyperparameter and epochs on quality metrics.	
MD_TR-4	The manufacturer should verify that the training actually trains the model	– There is a documentation revealing that the training process improves the model's performance.	There is a graph that shows how the loss gets smaller with increasing iterations / epochs.	

11.4.3 Model evaluation

Table 17 – Model evaluation requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standards / Regulations applicable
MD_EV-1	The manufacturer should plan the model evaluation.	– There is an evaluation plan. – The plan specifies the evaluation activities, the roles involved and the milestones at which these activities have to be performed. – The plan foresees the evaluation with clinically relevant data sets independent from training data sets.	– The evaluation plan can include activities prior and after product release. The latter activities can be part of the post-market surveillance plan. – The evaluation can include activities in a controlled environment, in closely monitored real-world settings and every local site. – Clinically relevant data sets are representative for the conditions as specified in the intended use.	[ISO 13485] clauses 7.3.2, 7.3.6 and 7.3.7 [ISO 14971] clause 10. GMLP guiding principle (8) (by FDA et al.)
MD_EV-2	The manufacturer should gain an understanding on how the machine makes a decision to evaluate the	– There is a validation specification and validation results for the	– A residual analysis in which the errors are listed via the feature values.	[EU-MDR (2017/745)] Annex I (17), Annex II (6.1).

Table 17 – Model evaluation requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standards / Regulations applicable
	correctness and robustness of the model.	evaluation of the model with validation data set. – 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 may be an analysis of datasets that have exhibited good model performance versus datasets that have performed badly. – 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.	– 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 (including 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: (https://yosinski.com/deepvis). This, however, depends on the ML method and cannot be demanded as a general best practice.	[IEC 62304] clauses 5.5 *ff*. [ISO 13485] clause 7.3.4 *ff*. [b-XAVIER] "Perspectives and good practices for AI and continuously learning systems in healthcare" [b-XAVIER University] "Building explainability and trust for AI in healthcare" DIN SPEC 2 [b-ISO/IEC TR 24028] clauses 10.2 and 10.3 GMLP guiding principles (6) (e.g., overfitting) and (8) (confounding factors) (by FDA et al.)

Table 17 – Model evaluation requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standards / Regulations applicable
			– A manufacturer using synthesized data may argue why this data mimic the actual data respectively why they are suitable to assess the robustness of the model. – 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. – Cross-validation helps to estimate the over fitting.	
MD_EV-3	The manufacturer should justify the selection of the model based on its intended use and performance on a representative dataset.	– There is a documentation of various models that have been compared. – There is a comparison of these models (architectures). – The comparison includes quality metrics. – There are clearly designed, representative datasets and model performance on those datasets is shown to be adequate following an assessment criterion e.g., an acceptable risk-benefit-ratio. – There is a risk-benefit assessment that discusses interpretability, performance (e.g., quality metrics, efficiency) and robustness.	Example for quality metrics: see above.	[ISO 14971] [b-XAVIER University] "Building explainability and trust for AI in healthcare" DIN SPECT 2 GMLP guiding principle (6) (by FDA et al.)

Table 17 – Model evaluation requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standards / Regulations applicable
MD_EV-4	The manufacturer should evaluate the model according to the evaluation plan.			GMLP guiding principle (8) (by FDA et al.)

11.4.4 Model documentation

Table 18 – Model documentation requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standards / Regulations applicable
MD_DC-1	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. – There is a documentation of data the model has been trained on. – There is a documentation of potential problems (e.g., biases) and limitations.	Ways 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	[ISO 13485] clauses 4.2.3, 4.2.5, and 7.3.6. [b-XAVIER University] "Building explainability and trust for AI in healthcare" [IEC 62304] on software on unknown provenance (SOUP) e.g., clause 8.1.2 [b-FDA OTS] guidance [ISO 14971].

Table 18 – Model documentation requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standards / Regulations applicable
			– 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 autoencoders – Deep Boltzmann machine (DBM) – Deep belief networks (DBNs), etc.	
MD_DC-2	The manufacturer should apply version and configuration control to development artefacts.	– There is an SOP for the 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)	E.g., trained models can be serialized.	[ISO 13485] clauses 7.3.10, 7.5.9.1 [FDA SW] guidance on software validation e.g., clause 5.2.1 [IEC 62304.]

Table 18 – Model documentation requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standards / Regulations applicable
		○ software libraries and frameworks.		

11.5 Product development requirements

11.5.1 Software development

Table 19 – Software development requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
SFTW-3	The manufacturer should perform and document the required activities pursuant to [IEC 62304].	– There is a 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).	– 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.	[IEC 62304] [IEC 82304-1] [b-XAVIER University] "Building explainability and trust for AI in healthcare" [FDA SW] guidance on software validation [b-FDA OTS] guidance [b-ISO/IEC TR 24028] e.g., clause 10.10 GMLP guiding principle (2) (by FDA et al.)

Table 19 – Software development requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
		– The SRS specifies user interface related requirements. – There is a documented software architecture.		
SFTW-4	The manufacturer should perform software unit, integration and system tests.	– There are unit-, integration and system tests results. – There are coverage reports. – There is a documented strategy for black box testing. – The tests cover all software / product requirements (including non-functional requirements). – The tests verify whether risk mitigation measures are effective. – Tests verify that the system safely manages unseen attacks. – There is a description of tested software version, test data, test environment (e.g., hardware), tester and evaluation of test results. – After changes to the software, the tests are repeated unless the	There are specific testing strategies for testing AI-based systems as described in the syllabus of the Korean software testing and qualifications board. (KSTQB & CSTQB Certified tester AI testing (CTFL-AIT) (http://www.kstqb.org/eng/sw/sw3_6.asp) To simulate unseen attacks a test data generator respectively Fuzzing/Fuzz tests might be used. The software / product requirements typically include: – performance – functionality e.g., meeting the quality metrics, dealing with invalid data (including warnings) – portability (see testing on target hardware) – interoperability – IT security.	[EU-MDR (2017/745)] Annex I e.g., 17.1) [IEC 62304] clauses 5.5-5.7 [b-XAVIER University] "Building explainability and trust for AI in healthcare" [b-ISO 2911] (testing techniques) DIN SPEC 2 [FDA SW] guidance on software validation [b-ISO/IEC TR 24028] e.g., clause 10.10 Annex C.7 – IT security guidelines GMLP guiding principle (2) (data integrity, cybersecurity) (by FDA et al.)

Table 19 – Software development requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
		manufacturer can provide a rationale for skipping test activities. – The tests are reproducible.		
SFTW-5	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.	[EU-MDR (2017/745)] Annex II, 6.1 [FDA SW] guidance on software validation e.g., clauses 5.2.5 and 5.2.6.
SFTW-6	The manufacturer should identify and verify all SOUP / 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 the 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 tables. Examples for prerequisites are: – hardware (e.g., processor architecture, RAM) – software (e.g., operating system, run-time environments e.g., .NET, browser) – AI acceleration hardware/inference acceleration hardware Comment: For comparative definitions, similarities and differences of SOUPs, COTS, OTS terms, please refer to (https://www.johner-	[IEC 62304] clauses 5.3 and 8.1.2 [b-FDA OTS] guidance.

Table 19 – Software development requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
			institute.com/articles/software-iec-62304/soup-and-ots/)	
SFTW-7	The manufacturer shall validate the software tools	– There is a validation plan for the training functionality of ML library – There are respective validation results		[IEC 13485] clause 4.1.6.

11.5.2 Risk management

Table 20 – Risk management

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
RSK_MGNT-16	The manufacturer shall assess and mitigate the risks that occur specifically to the chosen ML libraries	– There is a specification of functionalities of the chosen ML libraries that are used for training – There is a specification of functionalities of the chosen ML libraries that are used for prediction – There is an analysis of risks of a training function not meeting the specifications – There is an analysis of risks of predict function not meeting the specifications.	– Input for risk-based tool validation – Input for risk-based SOUP validation.	[EU-MDR (2017/745)], IVDR Annex I e.g., section 3. [ISO 13485] clause 3.1.6 [IEC 62304] clause 5.3.3 [ISO 14971] GMLP guiding principle (3) (by FDA et al.)

Table 20 – Risk management

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
RSK_MGNT-17	The manufacturer shall assess and mitigate the risks that occur specifically to the chosen software architecture.	– The risk analysis analyses risk for at least the most important components – The risk analysis analyses risks that are specifically to chosen technologies.	– Risks related to client server architecture – Risks related to (de)serialization of data – Risks related to format and protocol conversions – Risks related to multiple API versions and API gateways – Risk related specifically for programming language – Risks related to compiler and compiler settings.	[ISO 14971] GMLP guiding principle (3) (by FDA et al.)
RSK_MGNT-18	The manufacturer shall assess and mitigate risks related to data processing (e.g., during training).	– There is a list of all steps of data processing and annotation – There is an analysis of errors that can occur for each processing step – There is an analysis of risks arising from these errors.	– Error in format conversion – Errors in detecting and dealing with missing values – Errors in detecting and handling outliers – Errors in unit conversions – Errors in converting numeric in categorical values – Errors due to loss of data – Errors due to confusing data sources – Errors in feature extraction.	[ISO 14971] GMLP guiding principle (1) (by FDA et al.)
RSK_MGNT-19	The manufacturer shall assess the risks related to design transfer.	The risks analysis analyses consequences of porting the software and data to the target system.	The target system includes for example: – Hardware	

Table 20 – Risk management

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
			– Operating system – Other software	
RSK_MGNT-20	The manufacturer shall assess the risks caused by the specific selection of data.	– The risk analysis analyses the consequences of model bias, overfitting, variance of model performance dependent on input data – The risk analysis analyses the consequences of wrong reference data (e.g., wrong gold standard, wrong comparison).	See Table 11 DAT_CL-4	GMLP guiding principles (5) and (6) (by FDA et al.)
RSK_MGT-21	The manufacturer shall assess the risks by (unforeseen) operation conditions.	The risk analysis considers – data from patients that are not foreseen in the intended use – invalid data – data from systems and devices that are not foreseen in the intended use – operation of the device by users that are not foreseen in the intended use – operation of the device in clinical use conditions that are not foreseen in the intended use	– examples for invalid data: out of range and missing values, wrong data formats and units	GMLP guiding principle (6) (by FDA et al.)

11.5.3 Accompanying materials

Table 21 – Accompanying materials requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
MTR-1	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 document on instructions for use is under version control.	The identification of the product should be achieved by the product's UDI-DI.	[EU-MDR (2017/745)] Annex I (23.4) FD&C Act, FDA 21 CFR parts [b-21 CFR 801] and [b-21 CFR 820.120] [b-ISO/IEC TR 24028] e.g., clause 10.11.3.
MTR-2	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 reveal limitations. – The instructions for use specify the requirements of the input data. – The instructions for use specify the intended primary and secondary users pursuant to the intended use. – The instructions for use describe the other conditions applicable to	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 the accompanying diseases Examples for input data requirements are: – formats – resolutions – value ranges, etc.	[EU-MDR (2017/745)] Annex I (23.4) [b-21 CFR 801] [b-21 CFR 814.20] [b-XAVIER] "Perspectives and good practices for AI and continuously learning systems in healthcare" [b-ISO/IEC TR 24028] [b-XAVIER University] "Building explainability and trust for AI in healthcare" [b-ISO/IEC TR 24028] e.g., clause 10.11.3 GMLP guiding principle (9) (by FDA et al.)

Table 21 – Accompanying materials requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
		the product (e.g., runtime environment, use environment). – The instructions for use describe how to update the product. – The instructions for use of continuous learning systems describe what triggers algorithm updates and how to identify the version of this algorithm. – The instructions for use of continuous learning systems describe how to permit, delay and roll-back algorithm updates.		
MTR-3	The instructions for use should specify the performance of the product.	– The instructions for use specify the quality metrics.	– Examples of quality metrics are specificity, sensitivity, precision.	[EU-MDR (2017/745)] Annex I (23.4)
MTR-4	The instructions for use should explain the product and its working principle / underlying principle.	– 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. – The instructions for use provide information whether, and if yes how the system learns over time.		[EU-MDR (2017/745)] Annex I (23.4) [b-XAVIER] "Perspectives and good practices for AI and continuously learning systems in healthcare" TODO: Reference to AI/ML standards.
MTR-5	The instructions for use should reveal residual risks.	– The instructions for use list the factors that could have a negative	Examples of negative factors are:	[EU-MDR (2017/745)] Annex I (23.4)

Table 21 – Accompanying materials requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
		effect on the product's performance. – The instructions explain risks arising from a product not meeting the performance requirements. – The instructions for use list the possible ethical problems.	– patient population deviating from the specified population – data not meeting the specified criteria (e.g., formats, value ranges).	[ISO 14971] clause 8 [b-ISO/IEC TR 24028] e.g., clause 10.11.3.
	The instructions for use should further information that is legally required.	– The instructions for use identify the manufacturer. – The instructions for use list the 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 I (23.4) [b-EU-Regulation 207/2012].

11.6 Product validation requirements

11.6.1 Usability validation

Table 22 – Usability validation requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
U_VLD-1	The manufacturer should identify risk arising from a lack of usability.	– The risk management file lists risks that arise from misunderstanding, overlooking or	The product's visual output includes:	[ISO 14971] clause 5.2 [IEC 62366-1] clause 4.1

Table 22 – Usability validation requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
		ignoring the product's visual output. – The risk management file lists risks that arise from users blindly trusting or mistrusting the product.	– results e.g., treatment recommendations, diagnosis – limitations of the system – warnings e.g., whether preconditions are met – trustworthiness of results – reports, printouts – the manufacturer could evaluate how obvious the systems output is before users become suspicious.	[b-XAVIER University] "Building explainability and trust for AI in healthcare" b-FDA HFE] guidance [b-ISO/IEC TR 24028] clause 9.7 [FDA SaMD] Proposed regulatory framework for modifications to AI/ML based SaMD, clause 4 GMLP guiding principle (7) (by FDA et al.)
U_VLD-2	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 are adequate to mitigate risks.		[IEC 62366-1] (instructions for use are considered to be part of the accompanying documentation that is considered to be part of the user interface) [b-FDA HFE] guidance.

Table 22 – Usability validation requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
U_VLD-3	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 over all the risk mitigation measures.		[IEC 62366-1] clause 5.4 *ff*.
U_VLD-4	The manufacturer should define and specify the usability metrics for (a) understandability, (b) learnability and (c) operability of AIMD.	Specifications for following metrics: – Product description completeness – Function understandability – Input and outputs understandability – Ease of learning product functions – User documentation effectiveness – Operational error recoverability – Customizability – Physical accessibility – Other.	– What proportion of functions are understood by reading the product description / manual? – What proportion of interface functions are understandable? – How long does the user take to learn to use a function? – How easily the user can understand the messages from the software system? – How easily the user can recover from their worse situation? – How easily the user can customize operation procedures for their convenience?	[b-ISO/IEC TR 9126-2] (Part 2: External metrics).

Table 22 – Usability validation requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
			– What proportion of functions are accessible to users with physical impairments.	
U_VLD-5	The manufacturer should define and specify the 'quality in use' metrics to measure the extent to which AIMD meets the needs of the target users to achieve specified goals of effectiveness, productivity, and satisfaction in a specified context of use.	Specifications for following metrics: – Task effectiveness – Task completion – Error frequency – Task time – User wait time – Frequency of use of system help features – User satisfaction scale – Other.	– What proportion of the task is completed correctly by the user? – What is the frequency of errors encountered by the user? – How long does the user take to complete a task? – What proportion of the time do users spend waiting for the system to respond?	[b-ISO/IEC TR 9126-4] (Part 4: Quality in use metrics.)

11.6.2 Clinical evaluation

Table 23 – Clinical evaluation requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
CLC_EV-1	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 which support and contradict the hypothesis, that the benefits have been achieved. – If the data has been collected from other products, then 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 data. – The technical equivalence has to consider the software algorithms.	[EU-MDR (2017/745)] Article 61 and Annex XIV and XV [b-MEDDEV 2.7/1] revision 4 [b-XAVIER University] "Building explainability and trust for AI in healthcare" [b-21 CFR 820.30] (g) GMLP guiding principle (6) (clinical benefits and risks are understood) (by FDA et al.)
CLC_EV-2	The manufacturer should assess whether the promised medical benefit is achieved and 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: – a non- continuously learning model in comparison with a continuously learning model – classic algorithm in comparison with a machine learning model.	[EU-MDR (2017/745)] Article 61 and Annex XIV and XV [b-MEDDEV 2.7/1] revision 4 [ISO 14971] clause 4.2

11.7 Product release requirements

Table 24 – Product release requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
PDT_RLS-1	The manufacturer should verify the completeness of the 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 mentioned in the chapter "instructions for use".	[EU-MDR (2017/745)] Annexes I and II [ISO 13485] e.g., clause 7.3.5 [b-21 CFR 820.30] (e).
PDT_RLS-2	If the manufacturer of a continuous learning system plans to market its product in the United States of America (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 how these changes for systems will be performed. NOTE – Manufacturer may further clarify with authorities if SPS / ACP is desired for submission.		[FDA SaMD] Proposed regulatory framework for modifications to artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD.)

12 Post-market requirements

12.1 Production, distribution and installation requirements

Table 25 – Production, distribution and installation requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
PDT_INST-1	The manufacturer should apply version- and configuration control.	– There is an 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 control apply to the software as well as to the 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 the European Union (EU) and in the US, there is typically the need for a UID-DI and UDI-PI. Comment: For comparative definitions, similarities and differences of SOUPs, COTS, OTS terms, please refer to (https://www.johner-institute.com/articles/software-iec-62304/soup-and-ots/)	[IEC 62304] clause 8 [b-FDA Cybersecurity] guidance [ISO 13485] clause 7.5.8 [FDA SW] Guidance on software validation Annex C.8 - Cybersecurity.
PDT_INST-2	The manufacturer should ensure the design transfer.	– There is an 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 are instructions for installation, update and decommissioning. – These instructions specify the runtime environment.	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 [b-21 CFR 820.30] (h) [b-21 CFR 820.170].

Table 25 – Production, distribution and installation requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
		– There instructions specify how the correct installation can be verified.		
PDT_INST-3	The manufacturer should ensure effective and efficient communication with operators and users, including any security-related requirements that are the responsibility of the operators / users. The manufacturer also needs a mechanism to communicate security issues with their customers (e.g., a new vulnerability is discovered and the customer should take adequate steps to minimize potential harm until the manufacturer can fix the problem.)	– There is a SOP covering customer communication including handling of customer complaints. – There is a website that contains information about latest product releases and news related to security vulnerabilities. – 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.		[ISO 13485] clauses 5.2 and 7.2 [EU-MDR (2017/745)] Article 10 (9)-(j) [EU-MDR (2017/745)] Annex I (23.1) Annex C.7 - IT security guidelines.

12.2 Post-market surveillance requirements

Table 26 – Post-market surveillance requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
PST_MRK-1	When determining threshold values the manufacturer should analyse how	– There is an analysis, whether feedback loops can influence input values.	– Example for feedback loop: An algorithm provides prognoses. Therefore, the physician will	[EU-MDR (2017/745)] Annex III (1.1).

Table 26 – Post-market surveillance requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
	the application of the product might impact the features (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.	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.	
PST_MRK-2	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 the relevant data sources to be monitored. – These sources include information from SOUP manufacturers (also of ML libraries) and also includes security disclosures by those vendors. – The plan describes for each data source how, how often and by whom data is collected. – The plan specifies how data has to be analysed.	– "By whom" not only persons / roles, but also systems can be listed Examples for data sources are: – results from leading ML conferences – scientific literature – customer communication (e.g., complaints) – IT security databases – bug reports and release notes for SOUP / OTS – databases of authorities (e.g., FDA) – actual input values (features) for continuous training and or usage of the product – audit-logs. Examples for additional quality metrics see above. Also, the	[EU-MDR (2017/745)] Article 10 (9)-(i) [EU-MDR (2017/745)] Article 83 [EU-MDR (2017/745)] Annex III (1.1) [b-FD&C] Act 522 [b-21 CFR 822] [IEC 62304] clause 7.1.3 [b-XAVIER] "Perspectives and good practices for AI and continuously learning systems in healthcare" [b-ISO/IEC TR 24028] DIN SPEC 2 [FDA SaMD] Proposed regulatory framework for modifications to AI/ML based SaMD, clause IV

Table 26 – Post-market surveillance requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
		– 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 analyse whether the data in the field is consistent with the expected data or training data. – The plan requires to collect and analyse data to assess how the use of the system changes over time. – The plan for continuous learning systems specifies whether and if how often which data sets have to be retested after algorithm updates. – The plan for continuous learning systems specifies how, and how frequently changes in the algorithm updates are assessed. – 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.	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). The post-market plan should consider shifts such as: – Concept drift – Distribution shifts (labels) – Distribution shifts (feature) Actions include update of risk analysis and re-evaluation of risk-benefit analysis, re-training of algorithm, product recall, implementation of better risk mitigation measures.	Annex C.7 - IT security guidelines GMLP guiding principle (10) (by FDA et al.)

Table 26 – Post-market surveillance requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
		– The plan is approved.		
PST_MRK-3	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 the product type. – The post-market surveillance reports clearly identify the respective products via its UDI. – The post-market surveillance reports identify the post-market data and concludes whether activities are required.		[EU-MDR (2017/745)] Article 85. [FDA SaMD] Proposed regulatory framework for modifications to AI/ML based SaMD, clause IV GMLP guiding principle (10) (by FDA et al.)
PST_MRK-4	The manufacturer should establish a post-market risk management system.	– There is a specification on 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 on how, how often and by whom post-market data are evaluated for new or changed hazards, hazardous situations and risks.	– It is possible to combine post-market risk management and post-market surveillance. – The interpretability includes transparency and explainability. – The foreseeable misuse may include radiologists that rely on the software and do not look at the images anymore, so they overlook the findings. – The foreseeable misuse can include users or operators not updating the software or using the product after the communicated end of life.	[ISO 14971] clause 10 GMLP guiding principle (10) (by FDA et al.)

Table 26 – Post-market surveillance requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
		– 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.		
PST_MRK-5	The manufacturer, must 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 an "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) – programming language – user interface including warning – data interfaces.	[EU-MDR (2017/745)] Article 87. [ISO 13485] clause 7.3.9 [b-21 CFR 820.30] (i) [FDA SaMD] Proposed regulatory framework for modifications to artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD.)

12.3 Decommissioning requirements

Table 27 – Decommissioning requirements

REQ. ID	Requirement(s)	Checklist item(s)	Checklist examples and comments	Standard(s) / Regulation(s) applicable
DE_CMSN-1	The manufacturer should establish a plan before decommissioning the medical device.	The plan describes: – information of users and operators – disposal of product – archiving of product and data (e.g., training, test, validation data), software, documentation, considering security and privacy concerns.	Disposal can include: – putting product into trash – de-installation – sending back to manufacturer – logging-off – "de-registration" – confirmation of disposal.	[b-ISO/IEC TR 24028]
DE_CMSN-2	The manufacturer should analyse risks of decommissioning.	The risk analysis assesses: – risks for patients due to a product that is no longer available. – risks due to negative impact on other systems.	– Disturbance of workflows. – Interoperability problems.	[EU-MDR (2017/745)] Annex I (3) [ISO 14971] clause 10 in combination with terms and definitions 3.8 and 3.12 [b-ISO/IEC TR 24028]

13 Feedback

13.1 Publishing future versions

This is the first edition of the guidelines document. Future editions are planned to update this set of guidelines based on

1) scanning pertinent sources for additions and/or changes to the applicable international standards, jurisdictional laws and regulations, good practices, and processes (as well as additional expert augmentation of established guidelines to fill perceived gaps), and
2) incorporating feedback from manufacturers, regulators, and other users and readers.

This chapter described the types of feedback being sought.

Annex D provides a template for submitting feedback.

13.2 Seeking feedback

13.2.1 Types of feedback

The publisher of this set of guidelines is seeking feedback from document users and readers

1) to update contents (guidelines), and
2) to improve usability of the document.

13.2.2 Contents

Types of feedback about contents include the following:

– Change to an included guideline because it was revised by the source, e.g., regulatory authority.
– Applicable guideline added by a regulatory authority, etc. in a revision to an existing source, additional regulation, or otherwise.
– Needed but missing guideline (not in any source; expert suggestion to fill the identified gap).
– Update wording to a guideline because it was misstated in this document, source was misidentified, or similar editorial error.
– Request to change characterization, classification, etc. of the guideline in this document, including change to its priority score.
– Guideline should be deleted, e.g., because it does not apply within the stated purpose/scope of this document.
– Other contents consideration.

13.2.3 Usability

Usability pertains to the organization and presentation of guidelines and other information in this document, including the following:

– Organization of guidelines, including explanation of organization.
– Presentation of guidelines, e.g., in tables.
– Clarity and completeness of the descriptive material.
– Annex, including existence, organization, contents, etc.
– Type fonts, faces, style, size, etc.
– Needed but missing descriptive information, explanations, etc.
– Superfluous material (that can be deleted without affecting or thereby improving usability).

– Other usability and/or editorial consideration.

13.3 Providing feedback

Annex D provides a template for submitting feedback. Practical considerations preclude:

1) communicating individually with entities submitting feedback beyond acknowledging its receipt, and
2) publishing item-by-item disposition of comments and suggestions.

Nevertheless, readers can see results of their submissions in subsequent editions of the guidelines.

13.4 Processing feedback

The publisher assesses periodically the applicability to guidelines' purposes of all feedback provided. If accepted, a suggestion will be incorporated in the next edition of the guidelines, possibly, in a modified form. The process for acceptance is as follows: an expert work group (EWG) assembled for the purpose.

1) decides if the comment or suggestion should be accepted, possibly in a modified form (including those submitted by the EWG members themselves, e.g., arising from scanning existing and for additional sources),
2) if accepted, decides how the document should be modified,
3) after all the changes for the next edition have been settled, reviews the edited document for coherence, completeness, clarity, etc. (and may also solicit feedback on the draft from other experts or pertinent organizations), and),
4) produces the final draft document.

The publisher 1) edits and formats the document for publication and 2) publicizes the availability of the new edition.

14 Maintenance of AI checklist

14.1 Background and objectives

Medical device manufacturers must proof compliance with regulatory requirements and thereby that the devices are state-of-the-art with respect to safety, effectiveness and clinical benefits. This state of the art is rapidly evolving, in particular in the domain of artificial intelligence.

Therefore, best practice guides such as this checklist must be kept up to date with this state of the art. This requires a systematic process to update and to use this checklist on a regular basis.

14.2 Limitations and scope of this section

The objective of this clause is to describe this process of updating und using the checklist. It is neither the scope to describe the process of changing this "change process" itself nor the change of the underlying data structure.

14.3 Process requirements

This process must fulfil the following requirements:

– The process must be lightweight to minimize the burden for all stakeholders involved.
– The process must be explicitly documented to fulfil quality system requirements.
– The checklist must comply with regulatory requirements related to the control of documents. This includes requirements such as:
    • The checklist must be verified (i.e., reviewed) before approval.

- The checklist must be approved before release.
- All changes to the checklist must be traceable (who changed when and what).
- Any version of the checklist is clearly identifiable.
- The status of any version of the checklist is clearly identifiable (e.g., draft, verified / reviewed, approved, released, rejected).
- There are defined roles for authoring, verification and approval.
- There are criteria to verify and approve new versions of the checklist.

– The process must involve all relevant stakeholders:
  - Subject matter experts (i.e., AI experts)
  - Authorities and notified bodies including auditors, inspectors and tech file reviewers
  - Manufacturers.

– The checklist must be tailorable to specific use cases that differ e.g., in
  - duration of an audit, inspection, review
  - focus of an audit, inspection, review (e.g., focus on specific chapters of [ISO 13485])
  - type of product (e.g., risk, role of AI, technologies applied).

14.4 Process description

There are two processes:

1. Process to update, review and approve the checklist
2. Process to tailor the checklist to a specific use case (audit, inspection, review).

14.4.1 Process no. 1: Requesting changes and maintaining checklist

The maintenance process involves the following roles:

– **AI expert**

  These experts have practical and scientific knowledge in artificial intelligence and oversee both, state of the art and state of science. The expertise relates to:
  - AI architectures and models
  - AI technologies and libraries
  - Interpretable AI

– **Regulatory expert**

  These experts have a deep understanding of regulatory requirements on an international scope. They can map general requirements to AI specific requirements and best practices and vice versa. Therefore, these experts are able to estimate how completely the checklist items cover the regulatory requirements.

– **Authority and notified body**

  The members of authorities and notified bodies have the authority, oversight and competency to decide when a new checklist becomes effective and when an older version is withdrawn.

– **"Anybody"**

  Any person (no particular competency requirements) can request a change / modification to the checklist. Especially users (auditors, inspectors) are expected to submit these change requests.

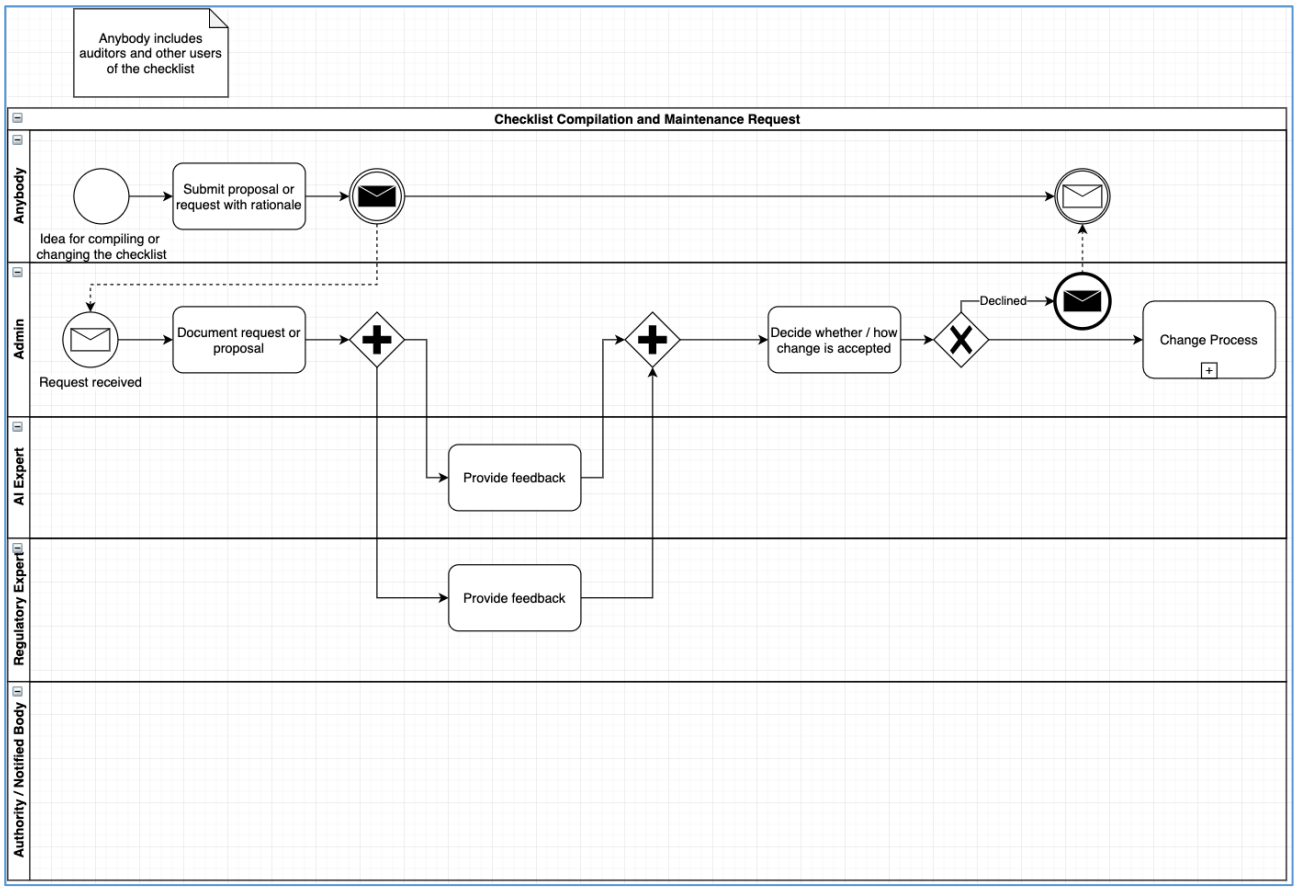

Figure 2 – AI checklist-requesting changes

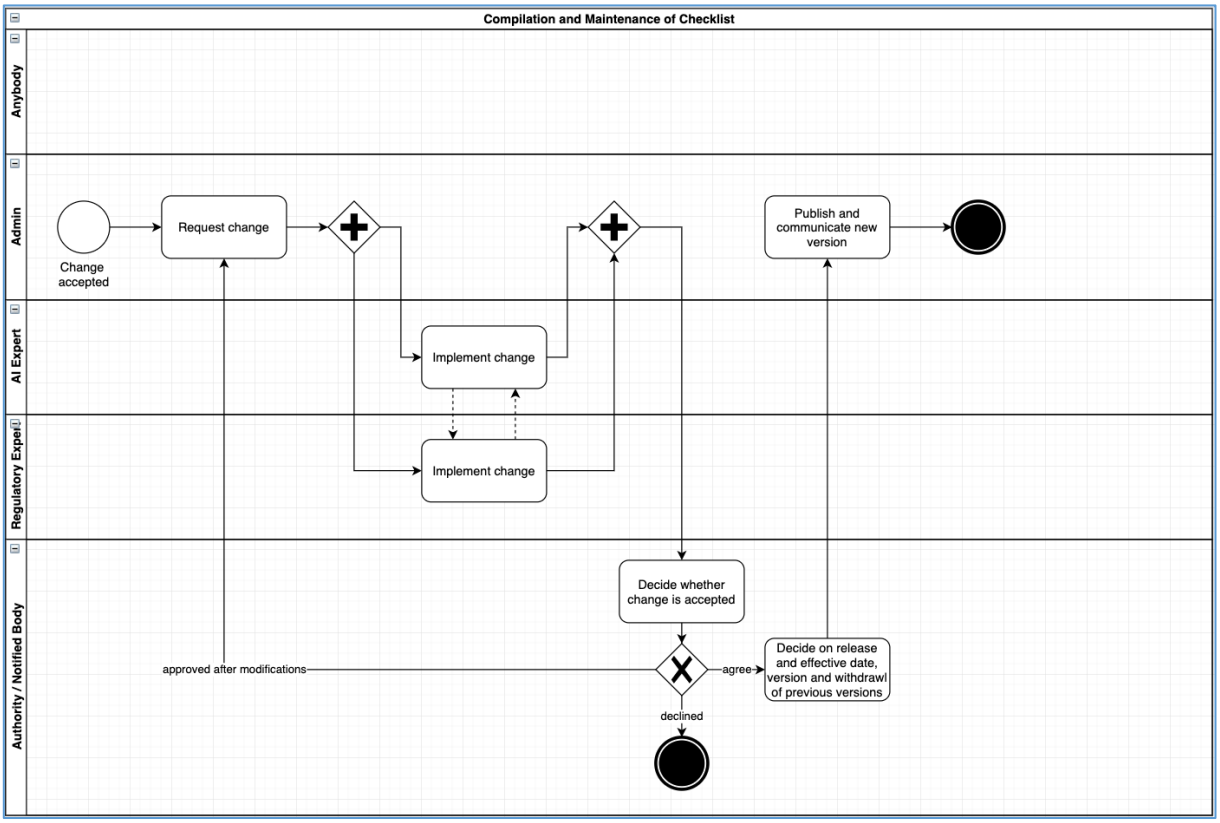

Figure 3 – AI checklist-performing, evaluating and approving changes

14.4.2 Process no. 2: Tailoring and using the checklist

The tailoring process mainly affects auditors, inspectors, and reviewers of technical files. They prepare an inspection / review checklist for a given situation:

– Scope of audit / inspection / review

– Available time

– Previous history with manufacturer

– Medical device (risk, technologies, role of AI)

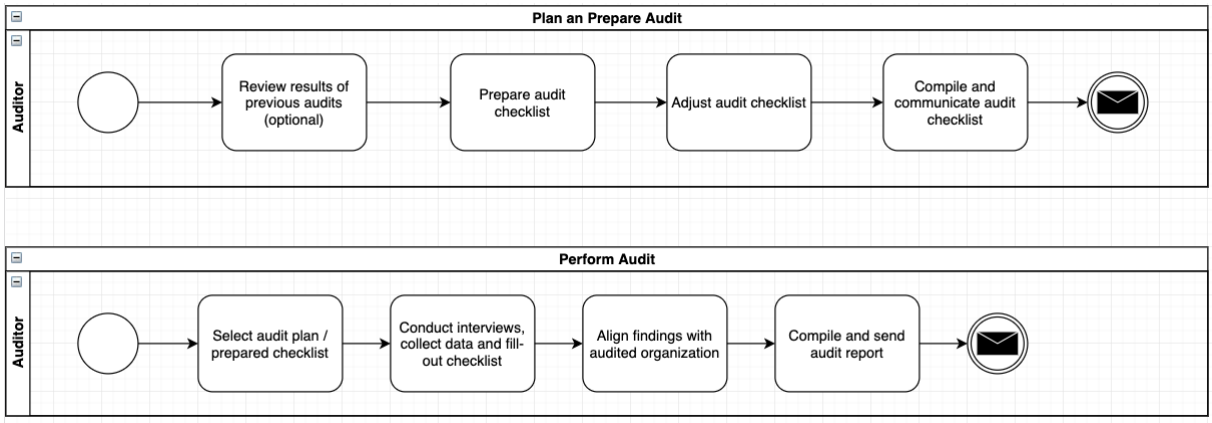

Figure 4 – AI checklist-processes for tailoring and using the checklist

Annex A

AI / ML related activities in the product life cycle

To facilitate continual product improvement in an iterative and adaptive manner with conformance to appropriate standards and regulations, it becomes a good practice for any regulatory framework to establish a system that can ensure transparency and accountability of all the life cycle processes involved in AI4MD development shown in Figure A.1. A brief rationale is provided in this annex on the need for a product development lifecycle process-oriented approach that forms the basis of the proposed regulatory requirements guidelines.

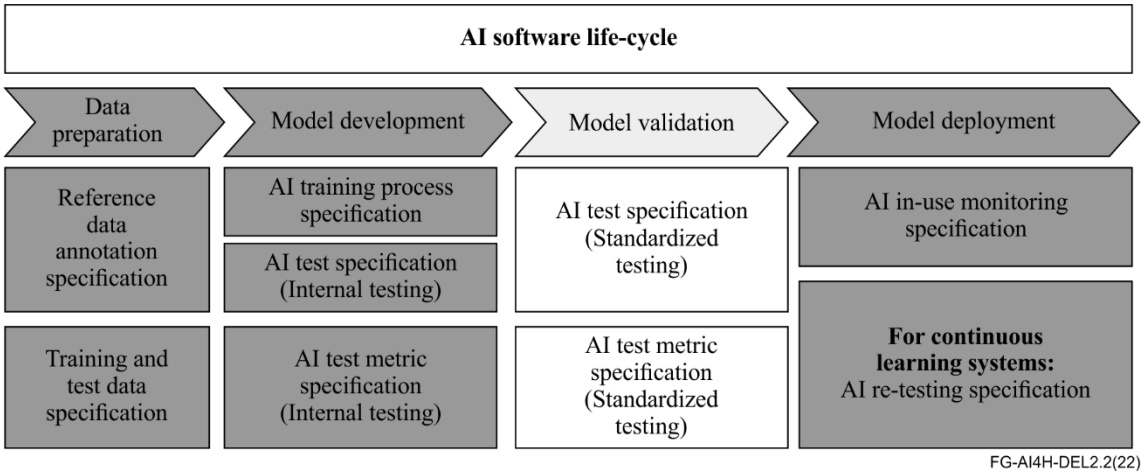

Figure A.1 – AI software life-cycle diagram

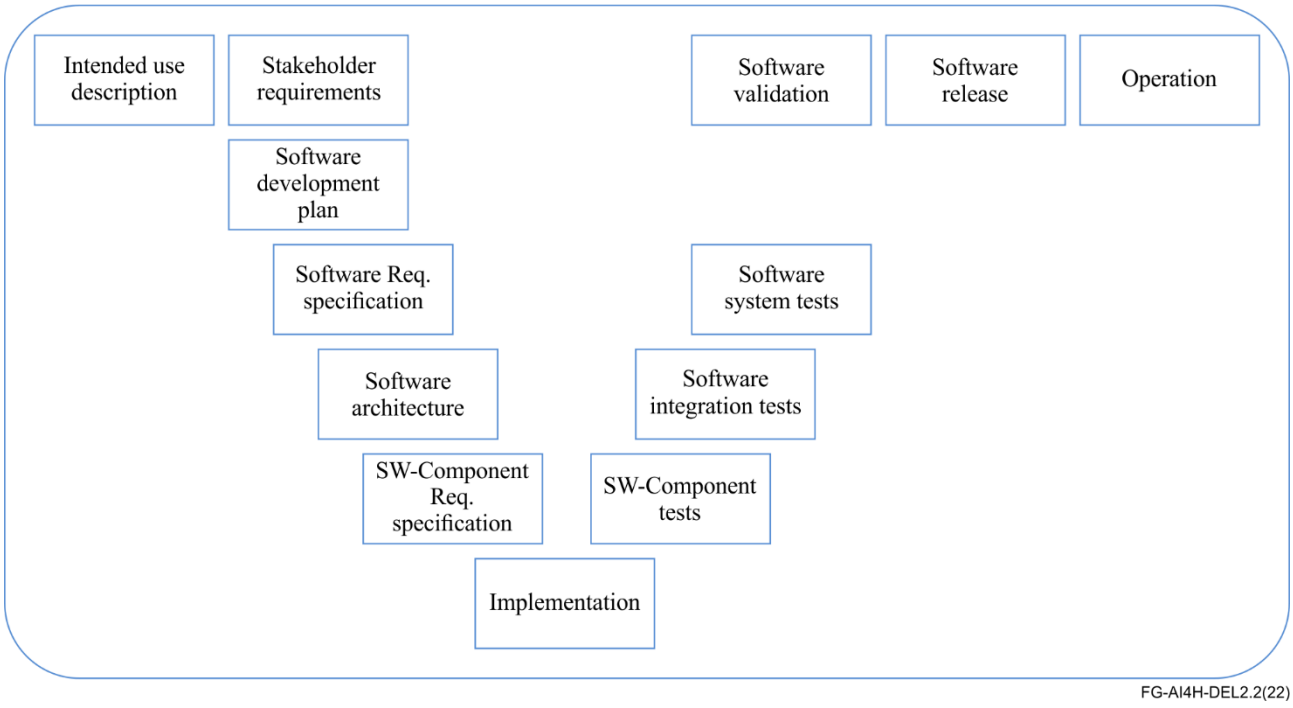

Figure A.2 – Product development life-cycle process (V-model)

Figure A.2 shows the V-model, which is widely accepted as a good practice 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 the medical device.

Annex B

Priority assessment scheme

B.1 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 the 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.

NOTE – The checklist contained in this document is an aide-mémoire. It cannot substitute for adequate training of people who will use the checklist and who therefore must understand the expressed requirements in context. Regulators, manufacturers, and other organizations adopting the checklist should ensure that the personnel designated to use it have the requisite educational background, professional experience, and specific training required to use it properly.

B.2 Requirements checklist: Priority assessment scheme

B.2.1 About priority score

All of the requirements listed in this document are necessary, if not sufficient in all circumstances for manufacturing every type of AI health application. That said, it may not be possible 1) to meet all requirements when building a QMS or, 2) to apply them all when auditing manufacturing processes or evaluating products. In such circumstances, one could decide that some requirements are more important than others, e.g., based on the potential for harm relative to the expected benefit to patients, and thus should be addressed first. Again, priority may depend on the type of AI health application and/or other considerations. The failure to meet the most important requirements may be considered to be major deficiencies; other requirements may be considered lesser deficiencies.

To develop a priority score, it is necessary to

1. establish a priority scale (and to define each scale point),
2. to develop an operating detail criteria for differentiating one scale point from another, and
3. to decide the appropriate scale point for each requirement.

While each regulator could decide these matters for itself, to assist regulators, this document provides the following guidance with respect to these matters. Additionally, users of priority scores need to ensure that they are implemented reliably, both by a given person and among people. Ensuring such inter-reliability may require specific educational background or professional experience, development of an implementation tool, and training in its use, among other methods. Further, through appropriate data collection and analysis, regulators could assess the validity of the assigned priority scores based on examining them in relation to the patient outcomes, and, if necessary, could accordingly adjust the scoring criteria, implementation methods, etc. Such application considerations are beyond the scope of this document.

B.2.2 Priority score: purpose

The priority score was developed to suggest to regulators, auditors, and QMS personnel which checklist items should be given highest priority when resources are not sufficient to address them all simultaneously or when only limited time is available in which to complete an audit. If a manufacture fails to meet or comply with the requirements, it should take appropriate action to resolve the underlying problems. If there are many such failures, the priority score may help to decide the order in which they should be addressed.

B.2.3 Priority score: decision anchor

The importance of a checklist item is risk-based, i.e., it depends upon specific consequences of not meeting or not complying with the requirement. These consequences are

1. ultimately, patient safety, i.e., the risk of harm to patients exposed to the product (and, when applicable, the risk of harm to users of the product, bystanders, and other involved persons), and
2. proximally, failure to meet established product specifications (which in turn has the potential to impact patient safety adversely).

In other words, failure to meet an important requirement can be expected to result in

1. products that do not meet one or more product specifications and
2. if such products were to be used it would be expected to cause serious harm to patients.

NOTE – It is possible that a product meets all product specifications and can still cause serious harm to patients. This is a different problem; one that requires appropriate clinical testing or monitoring, and sound decision-making based on appropriate weighing or risks and benefits. Such decisions may include changing the product specification and/or indications for use of the product. It is possible that a manufacturer meets all process requirements but still produces products that do not reliably meet product specifications. This is a manufacturing quality management system problem; one to be resolved through appropriate investigation and subsequent action.

B.2.4 Priority scale

We established the following 3-point importance scale. This scale is independent of whether or not a requirement is currently mandated by a regulator in at least one jurisdiction. In other words, the scale may differentiate the importance of different regulatory requirements. The checklist of requirements included in this document indicates in which regulations, a requirement may be found, if any, i.e., the source of each listed requirement.

B.2.5 Priority criteria

We established the following operational criteria to assign each listed requirement to a priority scale point.

1. **High importance:** Not fulfilling this requirement can be expected to cause serious patient harm or a major non-compliance in audits / inspections.
2. **Intermediate importance:** Not fulfilling this requirement can be expected to cause severe patient harm or a minor non-compliance.
3. **Less important:** It is not expected that not fulfilling this requirement will lead to patient harm nor to a non-compliance over the product's entire lifespan.

B.2.6 Priority scores for checklist requirements

Using the above-described criteria, manufacturers can assign a priority score to each listed requirement, arriving at such score by consensus. The resultant priority scores can be included in the table of requirements before submitting it to the regulators, auditors, and QMS personnel.

Annex C

Relationship to other guidelines and standards

The proposed guidelines were formulated based on a critical review of the 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 proposed guidelines are included here.

C.1 International medical device regulators forum (IMDRF) essential principles

IMDRF – Essential principles provide broad, high-level, criteria for design, production, and postproduction throughout the lifecycle of all medical devices and in vitro diagnostics (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 working group (WG)/N47 FINAL, 31 October 2018. (https://www.imdrf.org/sites/default/files/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 control artificial intelligence and machine learning algorithms utilized in medical technology.

The scope of Eps applicable to AI-MDs cover the following:

A. Safety and performance of medical devices – General essential principles

B. 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

C. 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 nos. refers to the 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 C.1 – IMDRF EP 5.1 – General

EP No.	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

Table C.1 – IMDRF EP 5.1 – General

EP No.	EP requirements	EP key concepts
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 C.2 – IMDRF EP 5.2 – Clinical evaluation

EP No.	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.	Ethical principles; Declaration of Helsinki Rights; Safety; Well-being; Pre-study protocol review; Informed consent; Leftover specimen

Table C.3 – IMDRF EP 5.8 – Medical devices and IVD medical devices that incorporate software or are software as a medical device

EP No.	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

Table C.3 – IMDRF EP 5.8 – Medical devices and IVD medical devices that incorporate software or are software as a medical device

EP No.	EP requirements	EP key concepts
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 C.4 – IMDRF EP 5.10 – Labelling

EP No.	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 C.5 – IMDRF EP 5.12 – Protection against the risks posed by medical devices and IVD medical devices intended by the manufacturer for use by lay users

EP No.	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 C.6 – IMDRF EP 7.2 – Performance characteristics

EP No.	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

Table C.6 – IMDRF EP 7.2 – Performance characteristics

EP No.	EP requirements	EP key concepts
7.2.1	a) The analytical performance can include, but is not limited to, a. Traceability of calibrators and controls b. Accuracy of measurement (trueness and precision) c. Analytical sensitivity/Limit of detection d. Analytical specificity e. Measuring interval/range f. 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

Table C.6 – IMDRF EP 7.2 – Performance characteristics

EP No.	EP requirements	EP key concepts
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

C.2 IMDRF SaMD risk categorization framework

The IMDRF publication "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 the 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 C.7.

Table C.7 – IMDRF SaMD risk categories

State of healthcare situation or condition	Significance of information provided by SaMD to the healthcare decision		
	Treat or diagnose	Drive clinical management	Inform clinical management
Critical	IV	III	II
Serious	III	II	I
Non-serious	II	I	I

The four categories (I, II, III, IV) shown in Table C.7 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.

C.2.1 Criteria for determining the SaMD category

The criteria for determining whether an SaMD is Category IV are:

– 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 an SaMD is Category III are:

– 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 an SaMD is Category II are:

– SaMD that provides information to treat or diagnose a disease or conditions in a non-serious 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 an SaMD 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.

C.2.2 Levels of autonomy

The IMDRF SaMD categories table was revised to account for various levels of autonomy as shown in the shown in Table C.8 below. Additional levels have been added to the "Treat or diagnose" category:

Table C.8 – IMDRF SaMD risk categories (revised)

	Significance of information provided by software to the healthcare decision				
State of healthcare situation or condition	**Treat or diagnose with no possible intervention**	**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 it either 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.

C.3 Johner regulatory guidelines for AI- for medical devices

The Johner guideline for AI-MDs is prepared and released by the Johner Institute, 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 the Johner guidelines can be found at: (https://github.com/johner-institut/ai-guideline/blob/master/Guideline-AI-Medical-Devices_EN.md).

C.3.1 Johner guidelines – objectives

The objective of Johner guidelines 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 the 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.

C.3.2 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
    a) Intended use
    b) Software requirement specification
    c) Data management
    d) Model development
    e) Product development
    f) Product release
3. Requirements for phases following development

C.4 FG-AI4H data and AI solution quality assessment criteria

Data and AI solution quality assessment criteria were formulated by the International Telecommunication Union (ITU)-T Focus Group on AI for Health's DAISAM working group (WG), 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 the 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 C.9.

Table C.9 – 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 and 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.)

Table C.9 – FG-AI4H data and AI solution quality assessment criteria

AI model development workflow	Assessment criteria	Description	Examples
			– 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 – World Health Organization (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 the case of classification problems

Table C.9 – FG-AI4H data and AI solution quality assessment criteria

AI model development workflow	Assessment criteria	Description	Examples
			– Probability output (0-1) as in case of the classification problems – Continuous valued output as in case of the 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 the case of classification problems – Probability output (0-1) as in the case of classification problems – Continuous valued output as in the 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 – Unsupervised learning based algos – k-means clustering – Hierarchical clustering – Reinforcement learning based algos – Association rule learning based algos

Table C.9 – FG-AI4H data and AI solution quality assessment criteria

AI model development workflow	Assessment criteria	Description	Examples
			– *A priori* 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 model evaluation	Evaluation metrics	Metrics used to quantify the errors and to evaluate the performance quality of the trained model on the test dataset. Selection of metrics depends on the type of the problem and the type of the model under consideration.	– Model accuracy (%) – Model accuracy - Mean and 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 – Area under the ROC curve (AUC) – Concordant – Discordant ratio – Other user defined performance measures

Table C.9 – FG-AI4H data and AI solution quality assessment criteria

AI model development workflow	Assessment criteria	Description	Examples
			– 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 be systematically tracked by maintaining progressive versions of the 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 and 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 and 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 o Authentication o Role based access control o Encryption

Table C.9 – FG-AI4H data and AI solution quality assessment criteria

AI model development workflow	Assessment criteria	Description	Examples
			○ 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 on digital certificates, 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 dataset. Quality tests are performed to minimize the noise and variance of the test data in order to maximize the performance accuracy of the 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.

C.5 ITU ML5G high-level requirements mapping to AI for health requirements

Requirements analysis was performed on the ITU-T FG-ML5G 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.

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	This roughly falls into the category of distributed training / inference / federated learning.
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 of ITU-T FG-ML5G for definitions.
Relevance for healthcare / assessment	Specification of data sources is required to provide transparency on robustness, e.g., to exclude misfit situations with unclear model outcomes.
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. Simulator functions hosted in the sandbox domain may be used to derive data for optimizations.
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 of 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	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 the effects 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	Similar to the monitoring of ML algorithm performance in the production setting. Reasonable thing to do in order to be able to intervene if outcomes do not 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 used for monitoring and managing the ML pipeline nodes in the system. MLFO monitors the model performance, and model reselection is recommended when 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.

ITU ML5G Req. code	ML-unify-019
	NOTE: evaluation may involve network performance evaluation along with model performance.
Relevance for healthcare / assessment	Similar to the previous point, monitoring of model outcomes.

C.6 DIN SPEC 92001 - AI devices life cycle processes requirements

DIN SPEC 92001-1:2019, Artificial Intelligence – Life Cycle Processes and Quality Requirements – Part 1: Quality Metamodel; ICS 35.080; 35.240.01.

C.6.1 Introduction

Challenge: For the following 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 Technical Report 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. The Technical Report proposes the differentiation between AI modules with regard to their safety and security. The Technical Report also outlines and defines the three central quality pillars i.e., functionality and performance, robustness, and comprehensibility.
Field of application	This Technical Report 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.

C.6.2 Terms and definitions

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

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

– DIN-TERMinology portal: available at (https://www.din.de/en/services/din-term).

C.6.3 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 and performance, robustness and comprehensibility. These three quality pillars are not fully disjoint. For instance, robustness may be conceived as part of functionality and 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.

In this Technical Report, 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 report requires the consideration of all listed quality requirements. Potential deviations of such AI modules need a profound justification.

C.6.3.1 AI module of the AI software quality metamodel system

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 report 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 Technical Report.

C.6.3.2 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.

C.6.3.3 Environment, platform, data, 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.

C.6.4 Life cycle

C.6.4.1 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 analysed concerning the properties like	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.

Stage	Definition	Context of AI
	convergence and input assumptions. In this stage, no model hyper parameters are chosen, and no final model evaluation is done.	
Development	This 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 and 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 analysed, 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.	Two 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 point to note in these stages is that everything is part of the development stage.

C.6.4.2 Life cycle processes

Processes are defined by title, purpose and outcome.

a) Organizational project-enabling processes: This part is important for 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 report 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 and 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 fulfil 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 and 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." 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.

C.6.4.3 AI quality pillars

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

The report 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 C.10 – Three key qualities

Key quality	Definition	AI meaning
Functionality and 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 analyse it with respect to the 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

Table C.10 – Three key qualities

Key quality	Definition	AI meaning
		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 defence 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 there is no clear explain to the stakeholder (white-box), some difficulties can be created to the project (grey-box or black-box.)

C.6.5 Conclusion of quality assurance

Three 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 qualities. All requirements for quality assurance are collected in these quality characteristics. Thus, the AI quality meta model covers all the aspects of the AI quality assurance.

C.6.6 Bibliography

[ISO/IEC/IEEE 12207] ISO/IEC/IEEE 12207:2017, Systems and software engineering – Software life cycle processes. <https://www.iso.org/standard/63712.html>

C.7 IT security guidelines

C.7.1 Meta information

C.7.1.1 Guideline objectives

The objective of these guidelines is to provide medical device manufacturers and notified bodies with instructions and a specific checklist in order to:

– Explain what notified bodies' expectations are

– Encourage the step-by-step implementation of IT security for medical devices

– Compensate for the absence of a harmonized standard (until there is one) to the very best.

Unlike a lot of other guidelines on IT security, these guidelines only relate to medical devices and focus on patient safety.

These guidelines are **not** intended to act as a textbook or guidelines for implementing IT security. Instead, they are intended as a guide for reviewing IT security.

The annex details the considerations that led to the creation of these guidelines.

C.7.1.2 Scope of application

These guidelines are intended for manufacturers of medical devices, especially networkable medical devices, and their service providers, as well as for people and organizations who have to evaluate the IT security of these devices.

It focuses on the IT security of the medical devices, not the organization's IT security.

The guidelines are also suitable for assessing the technical measures required for data protection. Nevertheless, the focus is on patient safety, not the confidentiality of the data.

C.7.1.3 Notes on use

C.7.1.3.1 Structure of the guidelines

These guidelines are based on the idea that IT security is based on three fundamental pillars:

1. Process requirements
2. Product requirements
3. Documented evidence that these process and product requirements have been met

The structure of these guidelines is based on these ideas: In clause C.7.2 it starts off with the general requirements, in clause C.7.3 it establishes the process requirements (including documentation), and in clause C.7.4 it establishes the product requirements (including documentation). Within these "main chapters", the requirements are structured along software life cycle process lines:

1. Process requirements

a) Requirements for the development process

1. Intended purpose and stakeholder requirements
2. System and software requirements
3. System and software architecture
4. Implementation and development of the software

5. Evaluation of software units
6. System and software tests
7. Product release

b) Requirements for the post-development phase

8. Production, distribution, installation
9. Market surveillance
10. Incident response plan

2. Product requirements

a) Preliminary remarks and general requirements
b) System requirements
c) System and software architecture
d) Support materials

The risk management requirements are woven into the requirements throughout the product life cycle.

C.7.1.3.2 Applicable chapters and requirements

Manufacturers should first use the guidelines to check the completeness of the specification documents (procedural and work instructions, checklists, etc.). For this, they should look at clauses C.7.2 to C.7.4.

Subsequently, manufacturers and the people who evaluate IT security on a product-specific basis (including internal and external auditors and technical documentation reviewers) should use the guidelines to evaluate IT security for the product. In this case, they can use clauses C.7.3 and C.7.4 of these guidelines as a checklist.

These guidelines contain requirements that do not apply to all products. Manufacturers must justify the exceptions that are not obvious.

C.7.1.3.3 Prioritization

If the manufacturers are not able to meet all the requirements of these guidelines from the outset, the requirements should be met in the order of their priority (from level 0 to level 3) as far as possible and where reasonable. These levels are described in this annex.

Acceptance of the security level achieved must be evaluated.

C.7.1.3.4 Comments

These guidelines contain "comments" on most of the requirements. These comments include justifications, references, comments and, above all, tips for auditors and reviewers.

Since the German term "Sicherheit" does not distinguish exactly between the important protection aims of freedom from danger and IT security, the term security is also used to emphasize IT security. Accordingly, the term "risk" means the technical possibility of reducing freedom from danger, while the term "threat" means potential attacks on IT security.

With regard to the further development of the guidelines, there is a trend towards the implementation of the ISO 2700x series of standards. This is due to detected attempts by professional attackers, who in the future will introduce malware into medical devices via the manufacturing organization's IT infrastructure, via means of communication, configuration tools, software tools and libraries. Additional security measures will therefore have to be initiated "earlier" in the development process, which will bring IT security issues in the company to the fore.

C.7.1.3.5 Liability

These guidelines are neither a legal requirement nor a harmonized standard. Accordingly, they do not differentiate between normative and informative elements.

Instead, the guidelines bring together best practices to describe the legally mandated "state-of-the-art" to the very best.

C.7.1.4 Authors and rights of use

These guidelines were prepared by the following authors:

– Andreas Purde (TÜV SÜD)
– Olaf Teichert (TÜV SÜD)
– Christian Johner (Johner Institute)

Georg Heidenreich (Siemens Healthcare GmbH) has made a significant contribution as a reviewer.

C.7.1.5 Document control and document identification

This document is managed via the version control system Git or the platform GitHub. Only the documents named in this repository are valid.

The version history including the respective authors can be found in the document history.

The released versions are identified as such in the repository using a tag. Versions without a tag are documents in the draft stage.

C.7.2 General requirements

C.7.2.1 Process

Manufacturers should cover all the aspects mentioned below either in the procedural instructions or in the corresponding plans in order to ensure that IT security is systematically ensured. Usually, the following procedural instructions and plans are affected:

– Development
– Risk management
– Verification and validation (if not part of development)
– Post-market surveillance and vigilance
– Service, installation, decommissioning
– Customer communication
– Management evaluation ([ISO 13485] requires "applicable new or revised regulatory requirements" to be taken into account).

If the manufacturer uses outsourced processes, the requirements apply accordingly. For example, a (software) development service provider would have to observe the sections of these guidelines that are relevant for it.

C.7.2.2 Expertise

Manufacturers must ensure and demonstrate that they have sufficient expertise to ensure IT security in line with the state of the art. This evidence is often most easily obtained through internal or external training.

In this way, manufacturers can also access the expertise of external resources.

No.	Requirement	Level	Comments
1	The manufacturer has created a list of all roles that are directly or indirectly involved with IT security	1	Examples are developers, testers, software architects, regulatory affairs, quality and product managers, as well as service and support staff.
2	The manufacturer has provided evidence of the IT-security expertise for each role.	1	Expertise refers here to the competency level "can do" (application) and not just "know" or "understand" (comprehension).
3	The manufacturer has records (e.g., training documents) that lead to the conclusion that the people in question actually have this expertise.	1	
4	The (software) development plans define the (additional or deviating) expertise on a product-specific basis.	2	Requirement since [ISO 13485].

C.7.2.3 Documentation

The manufacturer should be able to provide evidence that it has complied with the relevant requirements of these guidelines. There are no specific requirements for the documentation and "objective evidence".

In Europe (unlike in the USA), there is also no obligation to create a specific document on IT security. Instead, manufacturers can integrate these aspects into existing documents, such as the QM system specification documents and the technical documentation (e.g., software files, risk management files).

C.7.3 Process requirements

C.7.3.1 Product development requirements

C.7.3.1.1 Intended purpose and stakeholder requirements

No.	Requirement	Level	Comments
1	The manufacturer has identified all the neighbouring systems (medical devices, IT systems) that may be connected to the product.	0	
2	The manufacturer has created a list of roles (people, neighbouring systems) that may interact with the product.	0	Ask for the list of roles to be shown.
3	The manufacturer has identified all the markets and all the regulatory requirements that are relevant in these markets.	0	Ask for the list of IT security regulatory requirements to shown.

No.	Requirement	Level	Comments
4	The manufacturer has identified the intended primary and secondary users with their IT expertise.	1	Primary users use the product to achieve its intended purpose (e.g., diagnose a patient). Secondary users use the product according to the other intended use (e.g., cleaning, servicing, upgrading, configuring, transporting the product).
5	The manufacturer has defined the intended user environment.[4]	1	Aspects of this user environment are covered in the section with the accompanying materials.
6	The manufacturer has analysed the risks (hazards) that result if the system is used in the specified user environment by someone who is not a specified user.	1	Examples are users that: • are not trained according to specification, • use the system in a different (technical) environment than the specified (e.g., without malware scanner or unpatched iOS version)
7	The manufacturer has described in the risk management documentation what the IT security threats are and what the consequences would be for patients, users and third parties.	1	
8	The manufacturer has traceably generated the risk acceptance criteria based on the product's use and the state-of-the-art.	1	
9	The manufacturer has developed a system it can use to evaluate IT security-related risks.	2	Examples of systems to classify risks are DREAD and CVSS. However, these systems do not classify impact on safety.

C.7.3.1.2 System and software requirements

C.7.3.1.2.1 Authentication and authorization

No.	Requirement	Level	Comments
1	The manufacturer has identified all data interfaces.	0	Ask for the list of data interfaces (wired, WLAN, USB, etc.) to be shown.
2	The manufacturer has specified the protocols and standards used for each data interface.	1	These standards address all interoperability (IO) layers: • Structural IO: protocols such as TCP/IP, HTTPS, RS232 • Syntactical IO: formats such as JSON, XML, HL7 V2 • Semantic IO: taxonomies and classification systems such as LOINC, ATC, ICD10 • Organizational IO: standards such as the IHE workflows
3	For each data interface, the manufacturer has specified the functions offered via the interface.	0	Ask for the list of functions to be shown.
4	The manufacturer has analysed each function's security relevance (in terms of hazards).	0	

No.	Requirement	Level	Comments
5	The manufacturer has documented the effects of the safety-relevant (in terms of hazards) functions in the risk management documentation.	0	
6	The manufacturer has tested all the usage scenarios in which risks are generated due to a display of information that has not been specified (e.g., no display, incorrect display or display is too late).	1	Ask for this to be shown in the risk management or usability file.
7	For each role and neighbouring system, the manufacturer has defined the product functions that they may have access to via the corresponding interface.	1	Ask for the "mapping" of roles to functions to be shown, e.g., as a table.
8	The manufacturer has justified its choice of authentication procedure (username/password, biometric procedure, token, e.g., card) for all the roles and all the neighbouring systems.	1	The justification should be risk-based.
9	Where necessary, the manufacturer has requested additional mechanisms to minimize the probability of unauthorized access.	2	e.g., limitation of MAC or IP address ranges, limitation of physical access e.g., by locks and doors.
10	The manufacturer has analysed in the risk management process, the effects on patient safety if a person cannot access patient or device data (e.g., no authorization, they forget their password), and defined the appropriate measures.	1	These risks relate to products that behave according to specification. This is about balancing the protection goals of "confidentiality" and "safety".

C.7.3.1.2.2 Data communication

No.	Requirement	Level	Comments
1	The manufacturer has created a list of all data managed by the system.	1	Examples are patient data, treatment data, configuration data, certificates etc.
2	The manufacturer has assessed how worthy of protection these data are, in relation to confidentiality and their impact on patient safety.	1	
3	The manufacturer has evaluated, in the context of risk management process, the effect if particularly sensitive data is no longer protected.	1	
4	The manufacturer has investigated, in the context of risk management, the consequences of overloading the system with too many requests, for example, disk operating systems (DOS) or requests with volumes that are too large and has defined actions if necessary.	2	
5	The manufacturer has, in the context of risk management, analysed the consequences of the network no longer being available or no longer being available in the expected quality.	2	
6	The manufacturer has, in the context of risk management, analysed the consequences of the loss of data and establishes actions, such as making a backup, if necessary.	2	

No.	Requirement	Level	Comments
7	The manufacturer has established, in general or for specific products, the criteria for the checking of external data before they are processed further. (e.g., by white-listing of data and data ranges).	2	

C.7.3.1.2.3 Patches

No.	Requirement	Level	Comments
1	The manufacturer has a documented plan of how patches are applied and removed again. This plan includes the development, distribution, installation and review of patches.	1	This plan can be part of the incident response plan (see below).
2	The manufacturer has a list of all SOUP / OTS components.	1	This requirement belongs more to the "System and software architecture" section.
3	The manufacturer has assessed how often patches are required and how they should be installed.	2	

C.7.3.1.2.4 Other

No.	Requirement	Level	Comments
1	The manufacturer has established how the medical device informs the users in the event where cybersecurity is compromised.	2	
2	The manufacturer has assessed what functionality the medical device must guarantee in the event where cybersecurity is compromised.		

C.7.3.1.3 System and software architecture

No.	Requirement	Level	Comments
1	The manufacturer has documented all SOUP / OTS components (including version, manufacturer, reference to information on updates, release notes).	1	Ask for the list / table to be shown. The FDA requires a "Cybersecurity bill of materials (CBOM)".
2	The manufacturer has analysed the specific risks resulting from the choice of technologies (in particular programming language, SOUP / OTS components).	2	
3	The manufacturer has taken measures to ensure that the tools used (e.g., development environment, compiler) as well as the platforms and SOUP / OTS components are free of malicious code.	2	E.g., these tools only may be obtained from reliable sources.
4	The manufacturer has created a list of all services that the product offers or uses "externally" (e.g., through its operating system).	1	Ask for this list to be shown.

No.	Requirement	Level	Comments
5	For each service, the manufacturer has justified why it has to be visible externally (no time limitation).	2	Have the manufacturer explain how / where it is required and tested and that services that are not required (no time limit) are not offered (no time limit). The aim of this is "attack surface reduction".
6	If the product provides an interface, the manufacturer has described how attacks via this interface are controlled in the context of risk management.	1	Complete control of these risks is generally not really possible with USB interfaces, but also not necessary in all cases.
7	The manufacturer has identified the process offering / running this service for each externally visible service.	2	
8	For each process, the manufacturer has identified the user (at the operating system level) and, if this user does not run with minimal rights ("worst case" as root), it justifies this.	2	
9	The manufacturer has systematically identified the risks that would be caused by deficient IT security using threat modelling.	2	Have the model show that at least the external actors and/or threats and the threatened objects have to be identified.
10	The manufacturer has analysed the risks that result from the (auto-)update of anti-malware software.	1	
11	The manufacturer has established how the product detects compromised IT security, documents (log) this and reacts to it quickly.		
12	With regard to the audit log, the manufacturer has determined where its data is stored, how it is protected and updated and how this can be automatically analysed.		
13	For all software components, (at least top-level components), services and processes, and data and software components, the manufacturer has analysed which risks arise if they do not behave in accordance with the specifications due to a problem with IT security.	1	Corresponds to the failure modes and effects analysis (FMEA) approach.
14	The manufacturer has taken the software requirements into account in the software architecture.	1	For example, for the above software requirements, ask for the component(s) or technologies in the architecture that implement the requirements to be shown.

C.7.3.1.4 Implementation and development of the software

No.	Requirement	Level	Comments
1	The manufacturer has created coding guidelines that establish specific requirements for IT security.	1	Examples are code metrics, documentation requirements, formatting best-practices.

No.	Requirement	Level	Comments
			Ask the manufacturer to show the coding guidelines and corresponding requirements.
2	The manufacturer only plays code where reverse engineering and RAM readout cannot lead to unacceptable risks.	3	E.g., by restriction of physical access, code obfuscation.
3	The manufacturer either tests the software (source code and binaries) for malicious code before delivery and/or has protected all computers involved in the development and "production" of the software against malware.	0	
4	The manufacturer has defined measures that can find and eliminate buffer overflows.	2	

C.7.3.1.5 Evaluation of software units

No.	Requirement	Level	Comments
1	The manufacturer has defined at least one method that is used to check compliance with the coding guidelines.	1	The manufacturer will achieve this if it uses tools for static code analysis and/or establishes specifications for the code reviews.
2	The manufacturer requires code reviews for all components that map (IT) security-relevant functions.	2	
3	The manufacturer has concrete test criteria in its specification documents for the code reviews.	1	E.g., no use of unsafe functions, sanitization of all external data inputs.
4	The code reviews are carried out according to the four-eye principle and only by people who have the necessary expertise. The manufacturer has documented this expertise.	2	Documentation refers to documentation of requirements and documented evidence of competences.
5	The manufacturer has established which tests (e.g., unit tests) are necessary with which test cases and which degrees of coverage are necessary.	1	
6	The manufacturer has described how all SOUP and OTS components have to be verified.	1	

C.7.3.1.6 System and software tests

No.	Requirement	Level	Comments
1	The manufacture includes port scans at all relevant network interfaces in the test plan and also performs them.	1	
2	The manufacturer includes penetration tests at all relevant data interfaces and/or for all known vulnerabilities of the OTS components used in the test plan and also performs them.	2	For a known OTS component in the NIST Common / National vulnerability database, investigate a vulnerability and have the manufacturer explain how it ensures that it cannot be exploited or why it is not relevant.
3	The manufacturer includes the use of "vulnerability scanners" in the test plan.		

No.	Requirement	Level	Comments
4	The manufacturer includes fuzz tests at all relevant data interfaces with at least one tool in the test plan and also performs them.	2	Focus of fuzz testing should the code, not libraries.
5	The manufacturer includes a security check against the usual attack vectors in the test plan.	2	According to OWSAP top 10orCWE/SANS top 25.
6	The manufacturer includes the testing of robustness and performance in the test plan.		
7	The manufacturer includes the testing of all system / software requirements (see above) in the test plan.	1	
8	The manufacturer also has its software checked by IT security experts with regard to the above measures.	3	To reach level 3, this testing must include fuzz and penetration testing as well as analysis of the system / software architecture and the source code.
9	The manufacturer includes third-party test reports (e.g., from SOUP manufacturers) in the system test (if available).		

C.7.3.1.7 Product release

No.	Requirement	Level	Comments
1	The manufacturer has addressed the most common errors and the resulting hazards in the risk analysis or can at least explain how these risks are controlled.	1	Select an example from one of the linked lists of the most common errors and ask the manufacturer for a justification.
2	The manufacturer discusses the risks posed by all relevant attack vectors (see above) in the risk analysis and shows how these risks are controlled.	1	
3	The manufacturer has checked the effectiveness of all risk-control measures.	1	E.g., ask for references to corresponding tests to be shown.
4	The manufacturer has created a traceability matrix it uses to document that there are measures that control all risks related to IT security.	2	
5	The manufacturer has prepared the risk management report and the IT security report.	2	In Europe but not in the USA, the IT security report can be part of the risk management report.
6	The manufacturer has drawn up the necessary plans for the post-development phase (e.g., post-market surveillance and incident response plan).	1	Details below clause C.7.3.2.
7	The manufacturer has tested the completeness of the tests using a traceability matrix that links the tests to the requirements.	2	

C.7.3.2 Requirements for the post-development phases

C.7.3.2.1 Production, distribution, installation

No.	Requirement	Level	Comments
1	The manufacturer has described how it ensures that only the exact intended artifacts (files) in exactly the intended version are delivered in the product or as a product.	1	This is about configuration management. Also relevant for downloads or app stores.
2	The manufacturer has described how the people responsible for the installation know which is the latest version and how confusion during installation can be prevented.	2	This is only relevant for stand-alone software. A procedural or work instruction would be expected here.
3	The manufacturer has described how it ensures during the installation that the requirements specified in the support materials (see above) are actually met.	1	A procedural or work instruction would be expected here.
4	The manufacturer has established procedures that ensure that it can communicate quickly with operators and users of its products.	1	Level 2 is acceptable for non-critical products.

C.7.3.2.2 Market surveillance

No.	Requirement	Level	Comments
1	The manufacturer has created a post-market surveillance plan.	0	
2	The manufacturer has described which information is collected from the downstream phase.	1	
3	The manufacturer has described how and through which channels information is collected from the downstream phase.	1	
4	The manufacturer has described what information is analysed and evaluated from the downstream phase.	2	Ask the manufacturer to explain how it recognizes and defines a trend reversal and the threshold values it has set.
5	The manufacturer has described the resulting measures. as required EU-MDR Annex III.	2	Ask for the connection to the corrective and preventive actions in the process descriptions to be shown.
6	For each OTS component, the manufacturer has defined at least one source through which it is informed of the IT security problems and how often it is monitored and described the role this analysis performs with which tools.	2	These sources should include the websites of the OTS manufacturer and the NIST vulnerability database.
7	The manufacturer has described how it monitors the technologies and procedures used (e.g., cryptology) that are still secure.	2	

C.7.3.2.3 Incident response plan

(Includes recalls, patches, customer communication)

No.	Requirement	Level	Comments
1	The manufacturer has created an incident response plan.	2	This plan can be part of other plans e.g., vigilance or post-market surveillance plan
2	The incident response plan governs the criteria the manufacturer uses to evaluate information from the market and when it implements the emergency plan.	2	
3	Who develops and releases the patches and how and within what deadlines.	2	
4	How the customer obtains the patches.	2	
5	How the manufacturer ensures that the patches are also installed.	2	
6	Who informs the customers, how and within what deadlines.	2	
7	In which cases decommissioning or other product recalls is ordered and how.	2	

C.7.4 Product requirements

C.7.4.1 Preliminary remarks and general requirements

This clause describes the product's technical functions that support IT security. They must be introduced via the requirement specification (system / software requirements) and implemented as requirements.

The following technical product measures for IT security ("security controls") must, in principle, be appropriate for ensuring the intended purpose, taking into account the intended operating environment: In order to maintain the basic requirements for safety and function, the manufacturer may waive the implementation of individual product measures in justified and documented individual cases. Therefore, for each of the following requirements, instead of implementation, the manufacturer may also include a note in the documentation (e.g., performance specifications) explaining why the requirement has not been implemented with regard to the intended purpose and taking into account the operational environment and explaining the residual risk.

Manufacturers must check each of the measures described below to see whether they introduce new risks which themselves need to be controlled.

C.7.4.2 System / software requirements

C.7.4.2.1 Authentication

No.	Requirement	Level	Comments
1	The product only allows users to use it if they have authenticated themselves to the product.	0	Ask for the associated test cases to be shown.
2	The product allows the neighbouring systems (e.g., other medical devices, IT systems) connected at each data interface to exchange data only if they have been authenticated by the product.	0	Ditto. The requirement that data may only be transmitted in encrypted form is set out below.
3	The product allows password authentication only if this has a defined minimum length of which at least one is a non-alphanumeric character, and it contains at least one uppercase and one lowercase character.	1	The choice of the authentication mechanism has been justified by the manufacturer (see above).

No.	Requirement	Level	Comments
4	The product does not have a default password or requires that a password be changed during the first use.	0	
5	The product blocks users and neighbouring systems for *m* minutes after *n* attempts, with the manufacturer able to define the *n* and *m* values or their lower limits. The manufacturer has analysed the "safety-related" risks resulting from such a blocking and, if necessary, has implemented measures to minimize these risks.	1	
6	In the event of an unsuccessful login, the product only displays information that does not allow the user to identify the exact cause of the blocking, e.g., incorrect username or password.	2	
7	The product terminates user and neighbouring system sessions after *n* minutes of inactivity, with the manufacturer setting the value for *n* or its upper limit.	2	
8	The product assigns a role to each user and each neighbouring system for authentication.	1	Ask for an explanation of which software component(s) this functionality will be implemented in and how this is tested. The FDA even requires a hierarchical role strategy.
9	The product allows each role to access only the functions it is authorized for. This applies in particular for product updates and upgrades.	1	Ditto.
10	The product allows authorized users to block other users and neighbouring systems.	1	
11	The product allows authorized users to reset the authentication of any required elements (passwords, cryptographic keys, certificates) of other users and neighbouring systems.	1	
12	The product allows authorized users to delete other users and neighbouring systems.	1	
13	The product does not allow users to change their own permissions.	2	
14	The product allows permissions to be cancelled ("breaking the glass") and identifies / documents the person and the reasons.	2	
15	In a client-server architecture, all cybersecurity measures are determined and checked on the server side.	2	
16	In a client-server architecture, all client inputs are checked on the server side.	2	

C.7.4.2.2 Communication and storage

No.	Requirement	Level	Comments
1	The product allows users to permanently delete all patient-specific data. The product	2	

No.	Requirement	Level	Comments
	allows you to restrict permissions to do this (e.g., to roles).		
2	The product protects data from accidental deletion.	2	Manufacturers must check that there is no higher value security objective that prevents this, e.g., the above requirement.
3	The product only transmits data (or at least security-related data) via its data interfaces in an encrypted form. This also applies to storage on an external data carrier.	1	Ask which encryption is used and how the initial key exchange is done.
4	The product protects the integrity of the data against unwanted modification, e.g., through cryptographic procedures.	2	This applies in particular to security-relevant data, e.g., patient-specific data
5	By default, the product rejects all incoming connections (e.g., USB, TCP, Bluetooth).	1	FDA requirement.
6	The product checks all user inputs and all incoming data on the basis of verification criteria defined by the manufacturer (see above) before further processing.	1	Select an example of a data input at the user interface and the data interface and ask for the check to be shown in the code.
7	The product does not use wireless transmission for the transmission of time-critical data relevant to patient safety.	2	
8	The product stores passwords as "salted hash" only.	2	E.g., ask about the hash procedure and, if necessary, ask for it to be shown.
9	The product stores characteristics that could be used to identify a person in encrypted form only.	2	Ask for an explanation as to what the manufacturer defines as characteristics that could be used to identify a person and which encryption mechanism it uses.
10	The product protects critical data against accidental change and loss.	2	
11	Every time the program is restarted, it checks whether the mechanisms used to protect the data against loss and modification are in sync.		
12	The product allows users to deactivate data interfaces (e.g., USB, remote access).	2	
13	The product checks the integrity of the program code every time it is restarted.	2	
14	In the event of that security being compromised, the product provides an emergency mode for functions that have an effect on patient safety.	2	

C.7.4.2.3 Patches

No.	Requirement	Level	Comments
1	The product allows patches (own code, SOUP / OTS components) to be applied.	1	Manufacturers should be able to justify exceptions and to explain whether patching may or must be done remotely.

No.	Requirement	Level	Comments
2	The product allows you to remove defective patches again ("roll-back").	2	
3	The product limits the ability to apply or remove patches to users with the corresponding permissions (authenticated and authorized).	2	
4	The product checks changed program code (patches) for integrity before first use and when it is restarted.	2	These checks are usually carried out using signatures, which themselves must be protected against forgery.

C.7.4.2.4 Other

No.	Requirement	Level	Comments
1	The product logs all essential actions on/in the system in an audit log, including day and time and actor (user, system).	2	
2	The product ensures that it has the correct system time.	3	Have the mechanism explained. And how it is ensured that the user cannot unintentionally change the time without noticing.
3	The product protects the audit log against change.	2	Have the manufacturer explain how the protection is ensured and how a change to the audit log is identified by the system. If necessary, even ask for the responsible software components to be shown.
4	The product implements mechanisms that can detect penetration or an attack and react to them.	3	
5	The product allows the exchange of certificates.	2	

C.7.4.3 System / software architecture

No.	Requirement	Level	Comments
1	The software only uses tried and tested libraries / components (no self-implementation) for all cryptographic functions (e.g., encryption, signing).	1	The library must be included in the list of SOUP / OTS components. Ask the manufacturer to explain the selection (criteria) to you.
2	The software uses different technologies or keys for different functions (e.g., encryption of communication, encryption of data).	3	
3	The software is protected against malware (viruses, worms etc.) as far as it is technically possible.	1	Ask for an explanation of how the system is protected against malware and how this protection is maintained.
4	The software is based on versions of the SOUP / OTS components that do not contain any security vulnerabilities. Exceptions are justified.	1	Pick an example from the SOUP list and research which version the manufacturer has and check which vulnerabilities have been patched in the subsequent versions.

C.7.4.4 Support materials

The support materials refer primarily to the instructions for use and installation. If necessary, the manufacturers must also provide training materials.

No.	Requirement	Level	Comments
1	The instructions for use establish the intended IT environment for operation.	1	
2	The instructions for use specify which activities the operator must perform, as well as how and how often they should be performed.	1	
3	The installation and service instructions establish which other roles (operator, service technician) are responsible for which activities and how often they have to be performed.	1	
4	The support materials describe how to deal with lost or stolen authentication elements (e.g., cards, certificates, cryptographic keys) and forgotten passwords.	1	
5	The support materials describe how users can recognize an IT security problem with the product and what to do in this case.	2	This means that the product implements this detection.
6	The support materials describe which anti-malware software has been approved for the product and where (e.g., link) it can be obtained and who is responsible for updating it.	2	Only to the extent applicable.
7	The support materials contain the manufacturer's contact details, which can be used to contact the manufacturer, for example, in the event of problems with the IT security.	1	
8	The support materials also give a technical description of the product.	2	This is an FDA requirement in particular.

C.7.5 Prioritization

C.7.5.1 Prioritization

When prioritizing requirements, the guidelines take the following dimensions into account:

– Risk for an individual patient (combination of severity and probability of harm)

– Scope (only one patient, whole hospital, etc.)

– Feasibility (financial and time expenditure, requirements in terms of tools).

Prioritization leads to the following maturity levels:

– **Level 0 ("Layperson level")**: Even most laypeople would comply with this requirement. Anyone who does not even meet the requirements of this level should not be developing medical devices. An auditor may and must expect these requirements to be met in the very first audit.

– **Level 1 ("Advanced beginner" level)**: The manufacturer has already addressed the issue of IT security. This level can be accepted for less critical products and the initial audits. However, an improvement is expected in each subsequent year until level 2 is reached.

– **Level 2 ("State-of-the-art")**: This is the level that manufacturers generally have to reach in the long run. However, it does not yet reflect the state of scientific knowledge.

– **Level 3 ("Expert level")**: This level is reached by professional IT security experts. It goes beyond what an auditor can normally expect from medical devices. Energy suppliers, intelligence services and the military would have to operate at this level.

Depending on the risk posed by a product, an auditor or test may require a certain level from the outset.

C.7.5.2 Further reading

a) Laws
   - MDR
   - IVDR
   - GDPR
   - 21 CFR Part 11

b) Standards and best practice guides
   - AAMI/TIR57
   - EN IEC 60601-1
   - IEC 62443-2-1
   - IEC 62443-4-1
   - IEC 62443-4-2
   - IEC 82304-1
   - IEC 80001-1
   - IEC/TR 80001-2-2
   - IEC/TR 80001-2-8
   - UL 2900-1
   - UL 2900-2-1
   - BSI-CS 132
   - ISO/IEC 29147:2018 Information technology – Security techniques – Vulnerability disclosure
   - FDA guidance documents
     - "Content of premarket submissions for management of cybersecurity in medical devices"
     - "Postmarket management of cybersecurity in medical devices"

     "Design considerations and premarket submissions recommendations for interoperable medical devices"
     - "Wireless medical telemetry risks and recommendations"
   - BSI Cyber-Sicherheitsanforderungen an netzwerkfähigeMedizinprodukte [Cyber security requirements for network-compatible medical devices]

c) Specialist literature, textbooks
   - Eckert: IT-Sicherheit: Konzepte - Verfahren - Protokolle (De Gruyter Studium)
   - Johner Institute: Video trainings on the IT security of medical devices
   - Current trends in Bruce Schneier's blog.

C.7.5.3 Considerations

1) Manufacturers are developing more and more networked medical devices. As a result, the risks resulting from inadequate IT security (e.g., against cyberattacks) have increased. Customers are not informed about the state of the art during the procurement process and are responsible for the costs of security – before or after IT incidents. The number of IT incidents

is increasing as the professionalism of attackers is rapidly increasing. Many manufacturers do not take sufficient account of this.

2) The EU regulations (MDR, IVDR) explicitly demand IT security. The EU directives demand it indirectly. These requirements can be found in the respective Annex I with the basic (safety and performance) requirements. The IT security risk analysis goes beyond the analysis of intended usage scenarios. IT security should cover scenarios outside the intended use. Therefore, the concept of foreseeable misuse must be analysed more precisely because the manufacturer now has to take all the technical possibilities of invasion into the networked medical device into account.

3) In contrast to most other basic requirements, there are no harmonized standards on IT security. Therefore, there is no canonical catalogue of requirements that is recognized as reflecting the required state of the art.

4) The FDA has published several guidance documents as well as standards such as [b-UL Standard 2900-2-1]. These specifications are inconsistent in terms of granularity, completeness and conceptual integrity. They only meet the requirements that are usually placed on the quality of a standard to a limited extent.

5) A lot of standards are subject to charges (despite some questionable quality). In the authors' opinion, manufacturers should have free access to regulatory requirements.

6) Since most medical device manufacturers do not deal with IT security at all or only deal with it inadequately, they only meet the basic requirements. There is no consensus in Europe with regard to which technical and procedural obligations concern the manufacturer.

7) For most manufacturers, it would not be feasible in terms of time or in terms of finance to reach an IT security level in one fell swoop, as required by UL 2900. Therefore, manufacturers should gradually strive for and reach the state-of-the-art level with regard to IT security. The aim of these guidelines is to have the initial improvements implemented quickly rather than to do nothing due to excessive demands.

8) IT security has to be taken into account in all phases of the product life cycle process. Limiting it to testing is not enough. Together with technical product measures ("controls") and documentation, the guideline aims to refer to the three pillars of IT security: Requirements, process and documentation. The structure of these guidelines reflects these pillars and will continue to apply even after the foreseeable technological adjustments.

9) It must be expected that standards will be developed and harmonized for medical device IT security, but this may still take years. Therefore, a guideline is needed in this intermediate phase (only).

10) These guidelines are made available (by November 2018) so that they can provide guidance to manufacturers in the short term and allow them to act immediately. The speed of its development makes compromises in terms of cooperation with as many parties as possible unavoidable.

11) As the guidelines are based on a step-by-step convergence with the state of the art and have also been produced in a very short time, it cannot claim to be exhaustive.

12) However, the guidelines should represent an extensive and generally accepted level of requirements. The selection and priority of its requirements must therefore be as transparent as possible.

13) Such guidelines must take into account the specifics of the medical devices, including the principles of patient safety and a risk-based approach. In this particular case, selected IT security measures ("controls") may conflict with the basic requirements. For this reason, there cannot be a fixed list of controls for medical devices. The medical device's intended purpose as defined by the manufacturer is vital in each case.

14) For guidelines to have the intended positive effect on IT security, it is vital that they are easy to understand and implement. Therefore, these guidelines do not set any abstract or "high level" requirements but give binary test criteria.

15) In order to make them easier to implement, the authors should also avoid bringing together as many requirements as possible. Instead, they must limit themselves to requirements that they consider to be particularly relevant and feasible.

16) These guidelines should also be and remain available free of charge in order to encourage their distribution and increase awareness of them.

17) These guidelines deliberately do not require any specific technologies or processes. On the one hand, such technologies and processes are subject to too much change, and on the other hand, the authors of the guide do not presume to decide for manufacturers which technologies and processes are best for the specific application.

18) These guidelines should be available in German and English.

19) The focus is on the IT security of medical devices, not on IT security for organizations such as hospitals and medical device manufacturers. The authors of these guidelines are aware that attacks are increasingly affecting medical device manufacturers' supply chains. Future versions of these guidelines will have to take this into account by establishing requirements for organizations.

C.8 Cybersecurity

The purpose of this annex is to provide a brief introduction to cybersecurity. There have been many published reports, guidance, and standards relating to cybersecurity and health systems. This annex does not intend to replace those documents, but rather to provide an explanation about why cybersecurity is important and to give references for further reading. It is also important to note that cybersecurity is distinct from data privacy. This annex also does not provide information on data privacy.

Cybersecurity can be thought of as the measures taken to protect a computer system against unauthorized access or attack. Since AI solutions depend upon computer systems to function, cybersecurity is a concern for health system that utilize AI algorithms. Attackers, known as hackers, are the primary source of cybersecurity risk.

For the device developer, risk of such an attack is often difficult to quantify. Predicting the likelihood of a mechanical or electrical part failing is usually straightforward – i.e., how often a part is used, under what environmental conditions, the stress that it will be under, and one can then design the part accordingly. However, for cybersecurity, the likelihood of something being compromised is a function of many external and qualitative criteria: How attractive of a target is your data? How secure is the network where the device is installed? How often software vulnerabilities are identified and addressed? What is the cybersecurity expertise of the user?

The risk management process described in the [ISO 14971], "Medical devices – Application of risk management to medical devices" standard includes process steps to identify potential risks, evaluate those risks, take action to minimize those identified risks, evaluate any residual risks, and continue to monitor product performance and potential new risks.

A security management process is very similar – identify threat sources, identify vulnerabilities, evaluate those risks, take action to minimize those risks, evaluate residual risks, and continue to monitor the product and the cybersecurity environment for potential new risks. The NIST cybersecurity framework is an internationally recognized document that explores these concepts in more detail.

Many regulatory jurisdictions enforce certain cybersecurity requirements or publish guidance for medical device manufacturers to consider (reference).

References

[AAMI TIR57] AAMI TIR57:2016/(R)2019 Principles for medical device information security risk management – Risk management. <https://store.aami.org/s/store#/store/browse/detail/a152E000006j60WQAQ>

[Morgan] Morgan, C. (2021), *Medical Device Cybersecurity Regulatory Publications*. Apraciti. <https://apraciti.com/2021/01/22/medical-device-cybersecurity-regulatory-publications/>

[OneTrust] OneTrust DataGuidance (2019), *China: CAICT publishes white paper on cybersecurity*. <https://www.dataguidance.com/news/china-caict-publishes-white-paper-cybersecurity>

[CyberReg] Pretty good list of medical cyber regulations: <https://www.apraciti.com/blog/2019/11/25/global-regulatory-authority-publications-on-medical-device-cybersecurity>

[CAICT Cyber] CAICT white paper on cybersecurity: https://www.dataguidance.com/news/china-caict-publishes-white-paper-cybersecurity

[ML RiskMng] Collections of methods and examples of machine learning risk management in health applications:

– Oala, Luis, Jana Fehr, Luca Gilli, Pradeep Balachandran, Alixandro Werneck Leite, Saul Calderon-Ramirez, Danny Xie Li et al. "Ml4h auditing: From paper to practice." In Machine learning for health, pp. 280-317. PMLR, 2020.

– Parziale, Antonio, Monica Agrawal, Shengpu Tang, Kristen Severson, Luis Oala, Adarsh Subbaswamy, Sayantan Kumar et al. "Machine Learning for Health (ML4H) 2022." In Machine Learning for Health, pp. 1-11. PMLR, 2022.

– Roy, Subhrajit, Stephen Pfohl, Girmaw Abebe Tadesse, Luis Oala, Fabian Falck, Yuyin Zhou, Liyue Shen et al. "Machine learning for health (ml4h) 2021." In Machine Learning for Health, pp. 1-12. PMLR, 2021.

– Oala, Luis, Andrew G. Murchison, Pradeep Balachandran, Shruti Choudhary, Jana Fehr, Alixandro Werneck Leite, Peter G. Goldschmidt et al. "Machine learning for health: algorithm auditing & quality control." Journal of medical systems 45 (2021): 1-8.

– Parziale, Antonio, Monica Agrawal, Shalmali Joshi, Irene Y. Chen, Shengpu Tang, Luis Oala, and Adarsh Subbaswamy. "Machine Learning for Health symposium 2022--Extended Abstract track." arXiv preprint arXiv:2211.15564 (2022).

– Falck, Fabian, Yuyin Zhou, Emma Rocheteau, Liyue Shen, Luis Oala, Girmaw Abebe, Subhrajit Roy, Stephen Pfohl, Emily Alsentzer, and Matthew McDermott. "A collection of the accepted abstracts for the Machine Learning for Health (ML4H) symposium 2021." arXiv e-prints (2021): arXiv-2112.

– Oala, Luis, Marco Aversa, Gabriel Nobis, Kurt Willis, Yoan Neuenschwander, Michèle Buck, Christian Matek et al. "Data Models for Dataset Drift Controls in Machine Learning With Images." arXiv preprint arXiv:2211.02578 (2022).

– Fehr, Jana, Giovanna Jaramillo-Gutierrez, Luis Oala, Matthias I. Gröschel, Manuel Bierwirth, Pradeep Balachandran, Alixandro Werneck-Leite, and Christoph Lippert. "Piloting a Survey-Based Assessment of Transparency and Trustworthiness with Three Medical AI Tools." In Healthcare, vol. 10, no. 10, p. 1923. MDPI, 2022.

– Calderon-Ramirez, Saul, Shengxiang Yang, Armaghan Moemeni, Simon Colreavy-Donnelly, David A. Elizondo, Luis Oala, Jorge Rodríguez-Capitán, Manuel Jiménez-Navarro, Ezequiel López-Rubio, and Miguel A. Molina-Cabello. "Improving uncertainty estimation with semi-supervised deep learning for covid-19 detection using chest x-ray images." Ieee Access 9 (2021): 85442-85454.

– Willis, Kurt, and Luis Oala. "Post-hoc domain adaptation via guided data homogenization." arXiv preprint arXiv:2104.03624 (2021).

– Ramirez, Saul Calderon, Luis Oala, Jordina Torrentes-Barrena, Shengxiang Yang, David Elizondo, Armaghan Moemeni, Simon Colreavy-Donnelly, Wojciech Samek, Miguel Molina-Cabello, and Ezequiel Lopez-Rubio. "Dataset similarity to assess semi-supervised learning under distribution mismatch between the labelled and unlabelled datasets." IEEE Transactions on Artificial Intelligence (2022).

– Oala, Luis, Cosmas Heiß, Jan Macdonald, Maximilian März, Gitta Kutyniok, and Wojciech Samek. "Detecting failure modes in image reconstructions with interval neural network uncertainty." International Journal of Computer Assisted Radiology and Surgery 16 (2021): 2089-2097.

Annex D

Template for submitting feedback

Feedback on this Technical Report may be submitted

1) by completing the online template (TBD), or

2) emailing a scan of the completed printed template (TBD).

= Complete the online version of the template at [(TBD)]

= Email feedback to [email address (TBD)].

– *If questions, please email them to the above-referenced email address.*

– *Receipt of feedback will be acknowledged. Further communications about feedback or its eventual disposition are not possible.*

Template

Please submit feedback one item at a time, to enable its proper processing.

Submit as many completed templates as needed to provide all feedback.

– *Do not include trade-secret, propriety, confidential, or any other similar information.*

1. Feedback is about (check one of the following):

☐ Contents – go to section 2

☐ Usability – go to section 3

☐ Something else (or not sure) – go to section 4.

2. Feedback is about the following guidelines' contents (check one item):

☐ Change to an included guideline because it was revised by the source, e.g., regulatory authority.

☐ Applicable guideline added by a regulatory authority, etc. in revisions to an existing source, additional regulation, or otherwise.

☐ Needed but missing guideline (not in any source; expert suggestion to fill identified gap).

☐ Update wording to a guideline because it was misstated in this document, source was misidentified, or similar editorial error.

☐ Request to change characterization, classification, etc. of the guideline in this document, including change to its priority score.

☐ Guideline should be deleted, e.g., because it does not apply within the stated purpose / scope of this document.

☐ Other contents consideration.

Go to section 4.

3. Feedback is about the following aspect of usability, i.e., the organization and/or presentation of guidelines and other information in the document (check one item):

☐ Organization of guidelines, including explanation of organization.

☐ Presentation of guidelines, e.g., in the tables.

☐ Clarity and completeness of the descriptive material.

☐ Annex, including existence, organization, contents, etc.

☐ Type fonts, faces, style, size, etc.

☐ Needed but missing descriptive information, explanations, etc.

☐ Superfluous material (that can be deleted without affecting or thereby improving usability).

☐ Other usability and/or editorial consideration.

Go to section 4

4. Please elaborate on the contents or usability item checked above. Provide such information as 1) purpose and nature of the proposed change (or of improvement); 2) if applicable, reference to the pertinent regulation, standard, or other source document; 3) if no change is proposed, identify the problem (and, if possible, ways in which it might be resolved); 4) any other pertinent information.

_____

_____

_____

Go to section 5.

5. Attestation

In order for comments and suggestions to be considered, the submitter must sign the following attestation:

– The individual or entity submitting feedback is identified below (and referred to here as the "reader"); anonymous feedback cannot be accepted.

– The reader affirms and warrants that none of the feedback provided is confidential, propriety, trade-secret, or otherwise restricted.

– If feedback is being submitted by an organization, the reader is authorized to submit it on the organization's behalf.

– The reader gives the publisher of any and all future editions of the guidelines permission to use the feedback as the publisher sees fit; There is no guarantee that any future edition of the guidelines will incorporate the reader's comments or suggestions.

Name of individual or entity submitting feedback:

Email address (for acknowledgment):

Signature / date:

Annex E

AI4H project deliverables reference

Figure E.1 shows the generic as well as the AI specific aspects that need to be considered under the regulatory roadmap of medical devices. From Figure E.1, it can be inferred that AI-MD, as continuous learning or adaptive systems, are subject to modifications throughout its lifecycle and this results 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 E.2 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 E.1 – AI4H project deliverables reference ID for further reference.

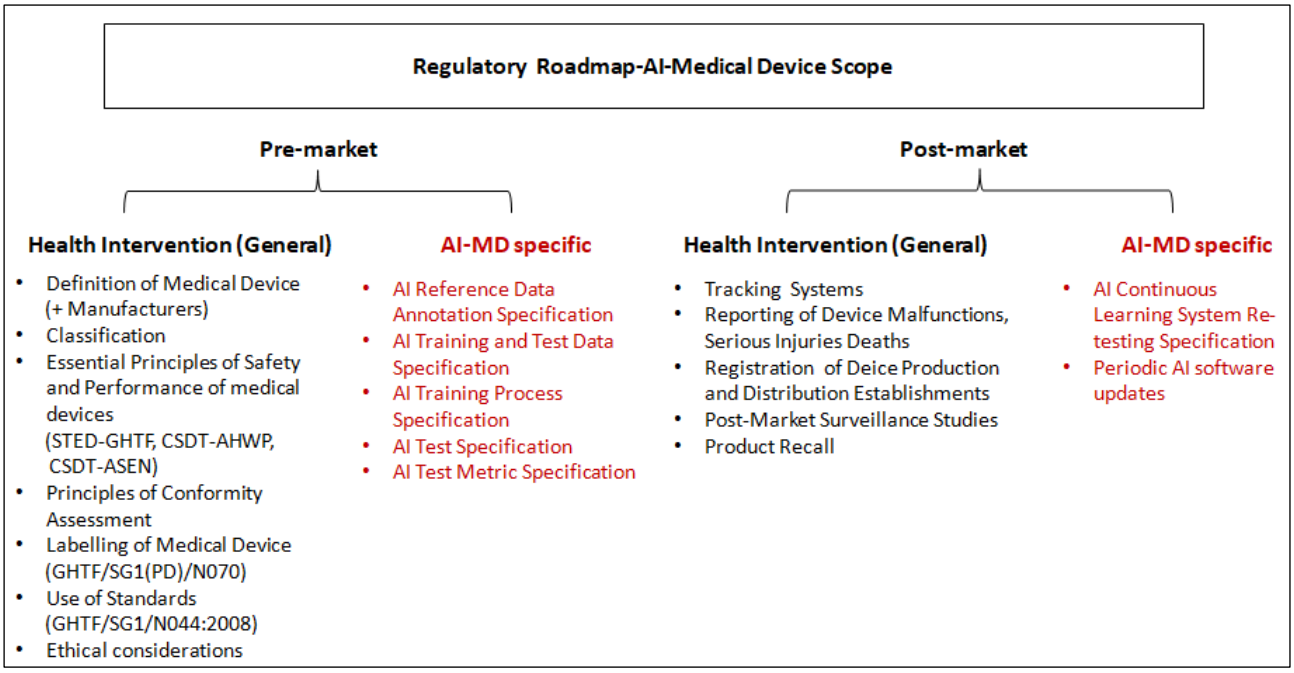

Figure E.1 – Regulatory roadmap-AI-medical device scope

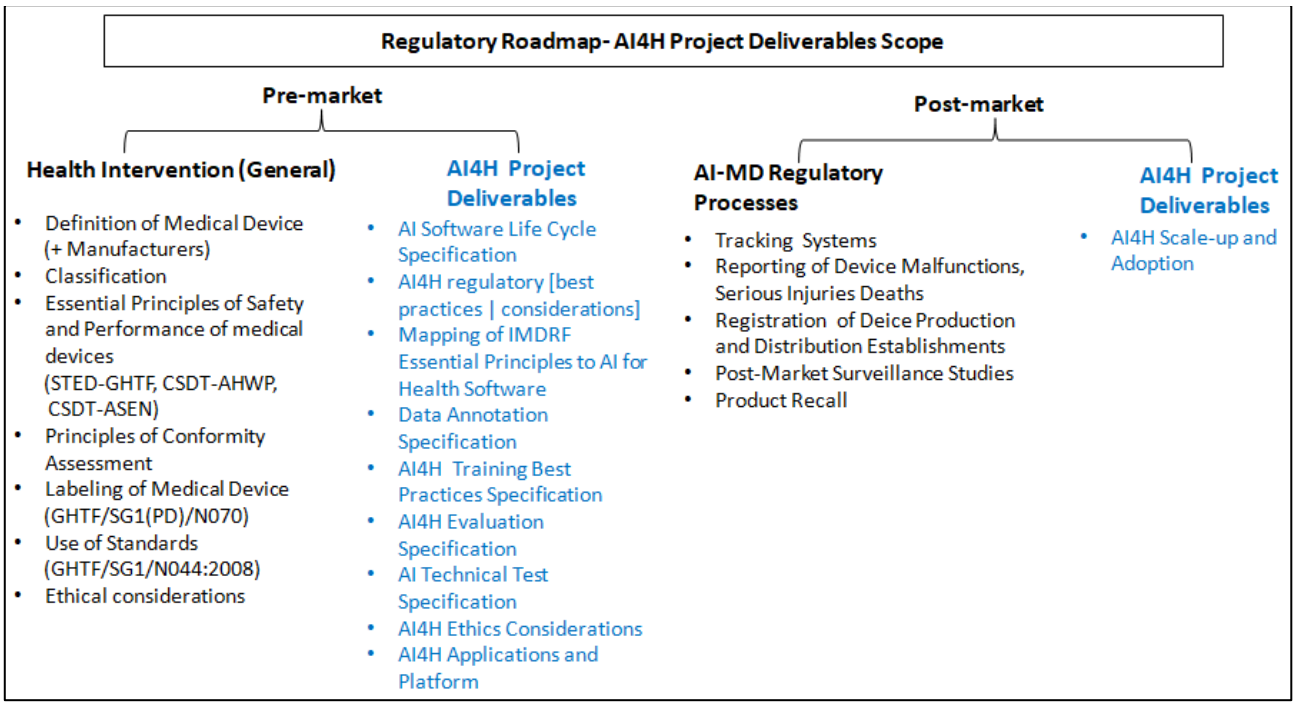

Figure E.2 – Regulatory roadmap-AI4H project deliverables scope

Table E.1 – AI4H project deliverables reference ID

AI4H project deliverable	ITU document reference ID
AI software life cycle specification	FG-AI4H DEL4
AI4H regulatory [best practices | considerations]	FG-AI4H DEL2
Mapping of IMDRF essential principles to AI for health software	FG-AI4H DEL2.1
Data annotation specification	FG-AI4H DEL5.3
AI4H training best practices specification	FG-AI4H DEL6
AI4H evaluation process description	FG-AI4H DEL7.1
AI technical test specification	FG-AI4H DEL7.2
AI4H ethics considerations	FG-AI4H DEL1
AI4H applications and platform	FG-AI4H DEL9
AI4H scale-up and adoption	FG-AI4H DEL8

Bibliography

[b-ISO/IEC 20889] ISO/IEC 20889:2018(en), *Privacy enhancing data de-identification terminology and classification of techniques*.
<https://www.iso.org/obp/ui/#iso:std:iso-iec:20889:ed-1:v1:en>

[b-ISO/IEC 22989] ISO/IEC 22989:2022(en), *Information technology – Artificial intelligence – Artificial intelligence concepts and terminology*.
<https://www.iso.org/obp/ui/#iso:std:iso-iec:22989:ed-1:v1:en>

[b-ISO/IEC 23053] ISO/IEC 23053:2022, *Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML).*
<https://www.iso.org/standard/74438.html>

[b-ISO/IEC 25010] ISO/IEC 25010:2011, *Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – System and software quality models*.
<https://www.iso.org/standard/35733.html>

[b-ISO/IEC Guide 2] ISO/IEC Guide 2:2004, *Standardization and related activities – General vocabulary*.
<https://www.electropedia.org/iev/iev.nsf/display?openform&ievref=901-05-01>

[b-ISO/IEC Guide 51] ISO/IEC Guide 51:2014, *Safety aspects – Guidelines for their inclusion in standards*.
<https://www.iso.org/standard/53940.html>

[b-ISO/IEC TR 9126-2] ISO/IEC TR 9126-2:2003(en), *Software engineering – Product quality – Part 2: External metrics*.
<https://www.iso.org/obp/ui/#iso:std:iso-iec:tr:9126:-2:ed-2:v1:en>

[b-ISO/IEC TR 9126-4] ISO/IEC TR 9126-4:2004, *Software engineering – Product quality – Part 4: Quality in use metrics*.
<https://www.iso.org/standard/39752.html>

[b-ISO/IEC TR 24028] ISO/IEC TR 24028:2020(en), *Information technology – Artifical intelligence – Overview of trustworthiness in artificial intelligence*.
<https://www.iso.org/obp/ui/#iso:std:iso-iec:tr:24028:ed-1:v1:en>

[b-ISO/TR 31004] ISO/TR 31004:2013, *Risk management – Guidance for the implementation of ISO 31000*.
<https://www.iso.org/standard/56610.html>

[b-ISO 2911] ISO 2911:2004(en), *Sweetened condensed milk – Determination of sucrose content – Polarimetric method*.
<https://www.iso.org/obp/ui/#iso:std:iso:2911:ed-2:v1:en>

[b-ISO 7396-2] ISO 7396-2:2007, *Medical gas pipeline systems – Part 2: Anaesthetic gas scavenging disposal systems*.
<https://www.iso.org/standard/41945.html>

[b-ISO 9000] ISO 9000:2000, *Quality management systems – Fundamentals and vocabulary*.
<https://www.iso.org/standard/29280.html>

[b-ISO 13407] ISO 13407:1999(en), *Human-centred design processes for interactive systems*.
<https://www.iso.org/obp/ui/#iso:std:iso:13407:ed-1:v1:en>

[b-IEC 31010] IEC 31010:2019, Risk management – Risk assessment techniques. <https://www.iso.org/standard/72140.html>

[b-IEC 60050-351] IEC 60050-351:2013, *International Electrotechnical Vocabulary (IEV) – Part 351: Control technology*. <https://webstore.iec.ch/publication/194>

[b-IEEE 610] IEEE Xplore 610.12-1990 – *IEEE Standard Glossary of Software Engineering Terminology*. <https://ieeexplore.ieee.org/document/159342>

[b-EU-Regulation 207/2012] Office Journal of the European Union, *COMMISSION REGULATION (EU) No 207/2012 of 9 March 2012 on electronic instructions for use of medical devices*. <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2012:072:0028:0031:en:PDF>

[b-21 CFR 801] 21 CFR 801 (2022), *Labeling – FDA*. <https://www.govinfo.gov/content/pkg/CFR-2022-title21-vol8/pdf/CFR-2022-title21-vol8-part801.pdf>

[b-21 CFR 814.20] 21 CFR 814.20 (2006), *Premarket Approval of Medical Devices – FDA*. <https://www.govinfo.gov/content/pkg/CFR-1996-title21-vol8/pdf/CFR-1996-title21-vol8-sec814-20.pdf>

[b-21 CFR 820.30] 21 CFR 820.30 (2017), *Subpart C–Design Controls – FDA*. <https://www.govinfo.gov/content/pkg/CFR-2017-title21-vol8/pdf/CFR-2017-title21-vol8-sec820-30.pdf>

[b-21 CFR 820.120] 21 CFR 820.120 (2022), *Device labelling – FDA*. <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?fr=820.120>

[b-21 CFR 820.170] 21 CFR 820.170 (2023), *Installation – FDA*. <https://www.ecfr.gov/current/title-21/chapter-I/subchapter-H/part-820/subpart-L/section-820.170>

[b-21 CFR 822] 21 CFR 822 (2023), *Postmarket Surveillance – FDA*. <https://www.ecfr.gov/current/title-21/chapter-I/subchapter-H/part-822>

[b-AAMI TIR57] AAMI TIR57:2016 / (R)2019, *Principles for medical device security – Risk management*. <https://store.aami.org/s/store#/store/browse/detail/a152E000006j60WQAQ>

[b-Ethics AI] Report / Study (2019), Ethics guidelines for trustworthy AI. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

[b-FDA Cybersecurity] FDA Guidance document (2022), *Cybersecurity in Medical Devices: Quality System Considerations and Content of Premarket Submissions*. <https://www.fda.gov/media/119933/download>

[b-FDA Digital health] FDA – *Guidances with Digital Health Content.* <https://www.fda.gov/medical-devices/digital-health-center-excellence/guidances-digital-health-content>

[b-FDA HFE] FDA HFE (2016), *Applying Human Factors and Usability Engineering to Medical Devices*. <https://www.fda.gov/media/80481/download>

[b-FDA OTS] FDA OTS (2019), *Off-The-Shelf Software Use in Medical Devices*. <https://www.fda.gov/media/71794/download>

[b-FD&C] FD&C Act 522, *Postmarket Surveillance Under Section 522 of the Federal Food, Drug, and Cosmetic Act*. <https://www.fda.gov/media/81015/download>

[b-HIPAA] HIPAA (1996), *Health Insurance Portability and Accountability Act of 1996*. <https://www.cdc.gov/phlp/publications/topic/hipaa.html>

[b-MEDDEV 2.7/1] MEDDEV 2.7/1 (2016), *Clinical Evaluation: A Guide For Manufacturers And Notified Bodies Under Directives 93/42/EEC and 90/385/EEC*. <https://www.google.com/url?sa=i&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=0CAMQw7AJahcKEwjI6vWZ-Zj9AhUAAAAAHQAAAAQAg&url=https%3A%2F%2Fec.europa.eu%2Fdocsroom%2Fdocuments%2F17522%2Fattachments%2F1%2Ftranslations%2Fen%2Frenditions%2Fnative&psig=AO>

[b-Moons, 2019] Moons, K.G.M., Wolff, R.F., Riley, R.D., Whiting, P.F., Westwood, M., Collins, G.S., Reitsma, J.B., Kleijnen, J. and Mallett, S. (2019), *PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration*. Ann Intern Med. <https://pubmed.ncbi.nlm.nih.gov/30596876/>

[b-OECD] OECD (2002), *Guidelines for th e Security of Information Systems and Networks: Towards a Culture of Security*. https://www.oecd.org/sti/ieconomy/oecdguidelinesforthesecurityofinformationsystemsandnetworkstowardsacultureofsecurity.htm

[b-OECD PF] Organisation for economic co-operation and development (2013), The OECD Privacy Framework. <https://www.oecd.org/sti/ieconomy/oecd_privacy_framework.pdf>

[b-PSO NAVIGATOR] PSO NAVIGATOR™ (2015), Wrong-Record, Wrong-Data Errors with Health IT Systems. ECRI Institute. <https://www.ecri.org/Resources/In_the_News/PSONavigator_Data_Errors_in_Health_IT_Systems.pdf>

[b-UL Standard 2900-1] ANSI/CAN/UL (2017), *Standard for Software Cybersecurity for Network-Connectable Products, Part1: General Requirements*. <https://www.shopulstandards.com/ProductDetail.aspx?productId=UL2900-1_1_S_20170705>

[b-UL Standard 2900-2-1] UL Standard (2017), *Software Cybersecurity for Network-Connectable Products, Part 2-1: Particular Requirements for Network Connectable Components of Healthcare and Wellness Systems*. <https://www.shopulstandards.com/ProductDetail.aspx?UniqueKey=33295>

[b-XAVIER] XAVIER (2018), *Perspectives and Good Practices for AI and Continuously Learning Systems in Healthcare*. <https://www.healthcareproducts.org/wp-content/uploads/2022/04/AI_WhitePaper_GoodPractices.pdf>

[b-XAVIER University] XAVIER University (2019), *Building Explainability and Trust for AI in Healthcare*. <https://www.healthcareproducts.org/wp-content/uploads/2022/04/AI_Whitepaper_BuildingExplainability_final3.pdf>

[b-Whiting, 2011] Whiting, P.F., Rutjes, A.W.S., Westwood, M.E, Mallett, S., Deeks, J.J., Reitsma, J.B., Leeflang, M.M.G., Sterne, J.A.C., and Bossuyt, P.M.M. (2011), *QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies*. Ann Intern Med. <https://pubmed.ncbi.nlm.nih.gov/22007046/>