AIRIS 2024
Use of AI in Medical Product Development
AI Regulatory & International Symposium
Co-organized by MFDS and U.S. FDA

IEC 63521
MLMD – Performance Evaluation Process

MLMD: Machine Learning-enabled Medical Device

Jonghong Jeon

ETRI
Email: hollobit@etri.re.kr

ETRI
www.etri.re.kr

Contents

- Definition of MLMD
- Why Performance Evaluation ?
- History, Goal, Member, Schedule of PT 63521
- Goal and basic idea of IEC 63521
  - Three pillar model of Performance (SV, TPV, CPV) & Evaluation flow
  - Safety and Effectiveness of ML
- Considerations
  - Align with baseline standards - 62304, 13485, 14971, 62366 …
  - Align with regulatory requirements and International standardization
  - GenAI, Foundation model, LLM
- Conclusion

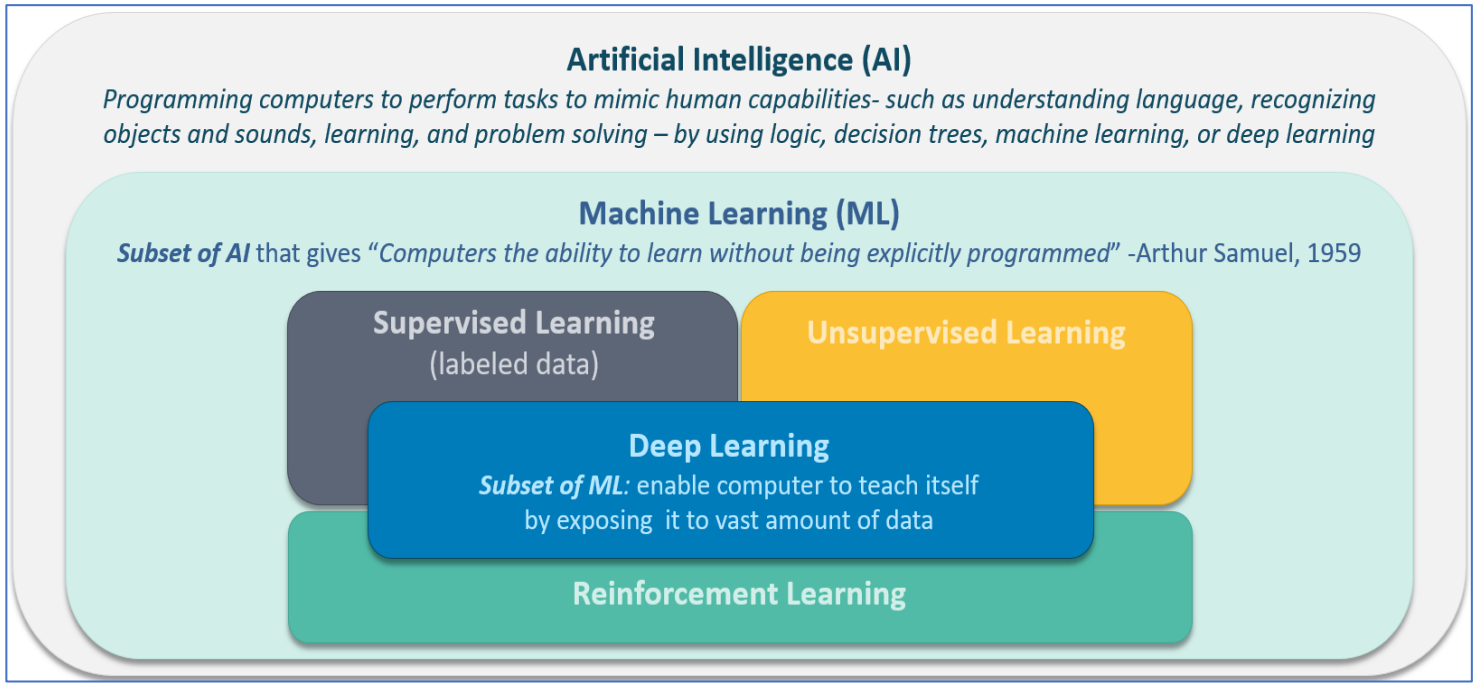
Simple Curriculum Vitae

- **IEC TC62 A, D (Medical equipment, software and systems)**
  - (Convenor) PT 63521 (Project Team 63521) – AI/ML-MD performance evaluation
  - (Project leader) IEC 63521 Machine Learning-enabled Medical Device – Performance Evaluation Process
  - **SNAI (Software Network AI)** Adhoc group member
- **ISO/IEC JTC 1/SC 42 (Artificial Intelligence)**
  - (Project editor) ISO/IEC TS 29119-11 Testing for AI Systems
- **IEC TC124 (Wearable Electronics)**
  - (Convenor) ahG7 (Future Use Cases for Wearable)
  - (Project Leader) IEC 63203-402-2:2024 (Step counting)
- **ISO/IEC JTC 1/WG 12 (3D Printing & Scanning)**
  - (Project Editor) ISO/IEC 3532-2:2024 Medical image-based Modeling – Part2: Segmentation.
  - (Project Editor) ISO/IEC CD 8803 – accuracy and precision evaluation process for modeling from 3D scanned data
  - (Proposed NP) Phantom-based evaluation methods for 3D printing modelling software
- **IMDRF (International Medical Device Regulators Forum) AIMD WG**
  - Chair of Korean mirror committee
- KoSAIM Director of Standardization (1st term: 2018-2020, 2nd term: 2021-2023)
- Head of the Standards and Technology Subcommittee of the Medical AI Technology Standardization Forum

Definition of MLMD

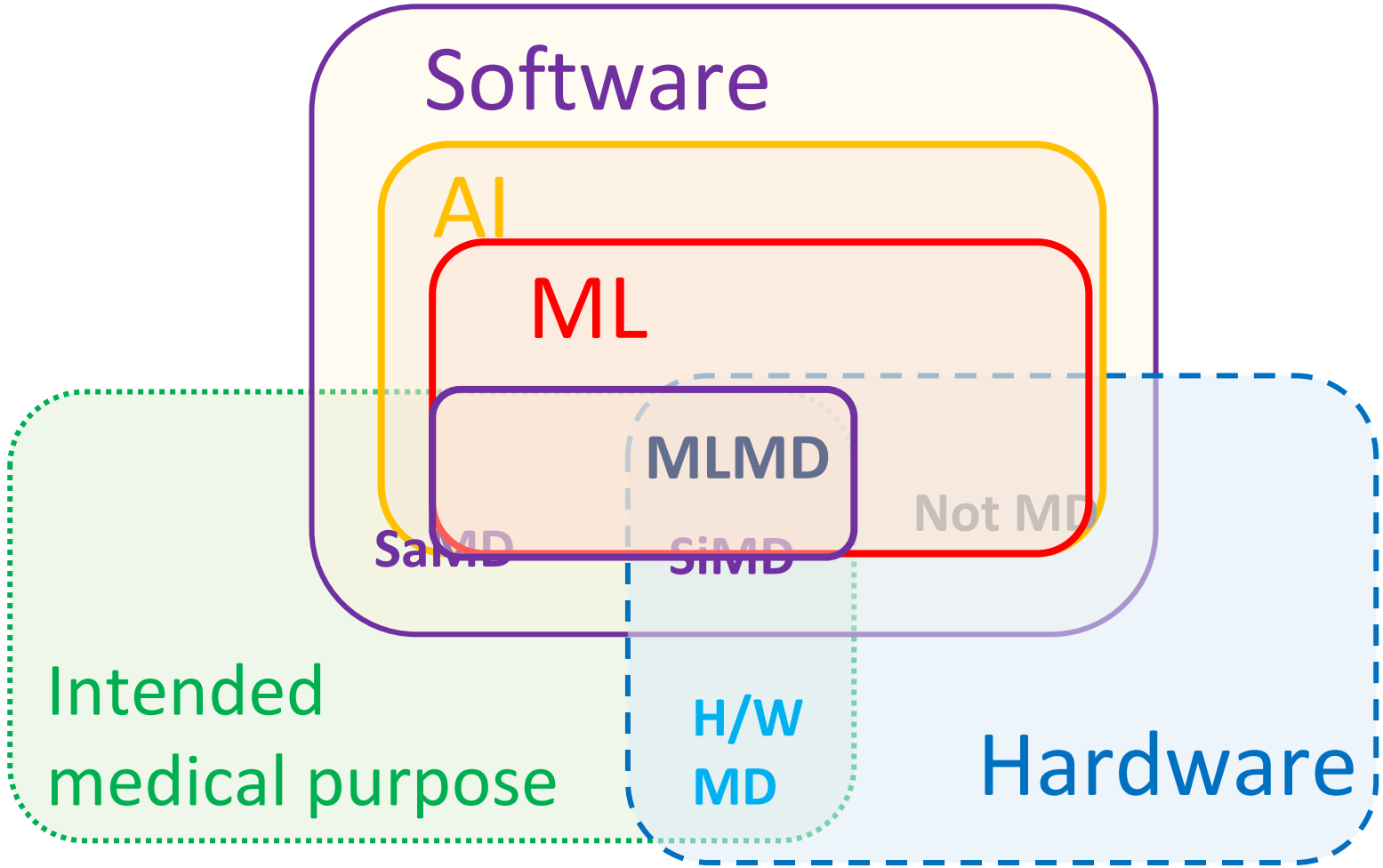
AIRIS 2024
Use of AI in Medical Product Development
AI Regulatory & International Symposium
Co-organized by MFDS and U.S. FDA

- IMRDF's MLMD Definition
  - **MLMD (Machine Learning-enabled Medical Device)**
    - "**A medical device** that *uses **machine learning***, **in part** or **in whole**, to achieve its intended medical purpose"



- "the ability to learn without being explicitly programmed" ➔ learn from **DATA**

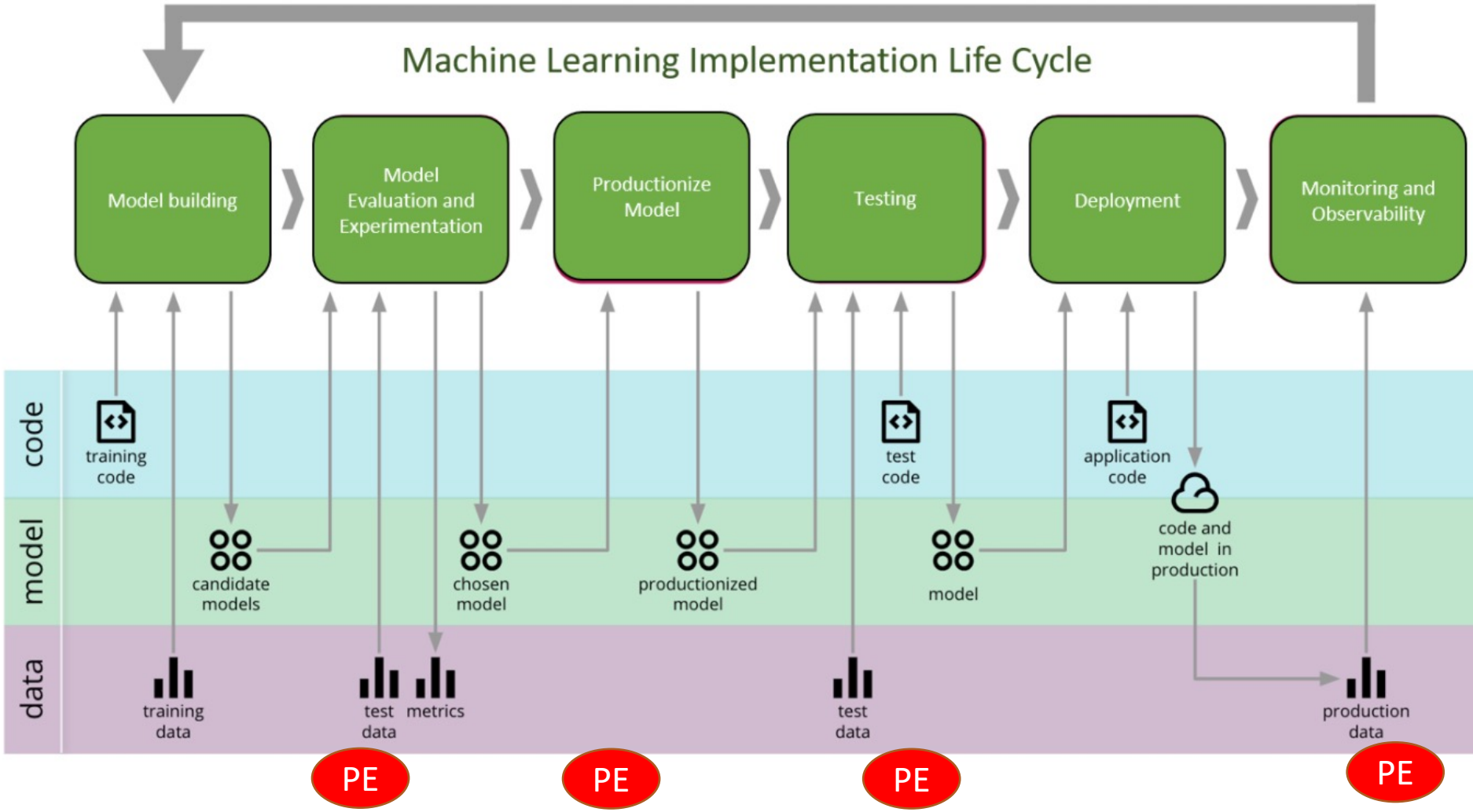
Categorization of MLMD



SaMD: Software As a Medical Devices
SiMD: Software In a Medical Devices (sometimes referred to as “embedded” or “part of”)
MLMD : Machine Learning-enabled Medical Devices

Why Performance evaluation process ?

- **Performance Evaluating of a AI/ML-MD product is an essential part** of implementing an effective AI/ML model



Why Performance evaluation process ?

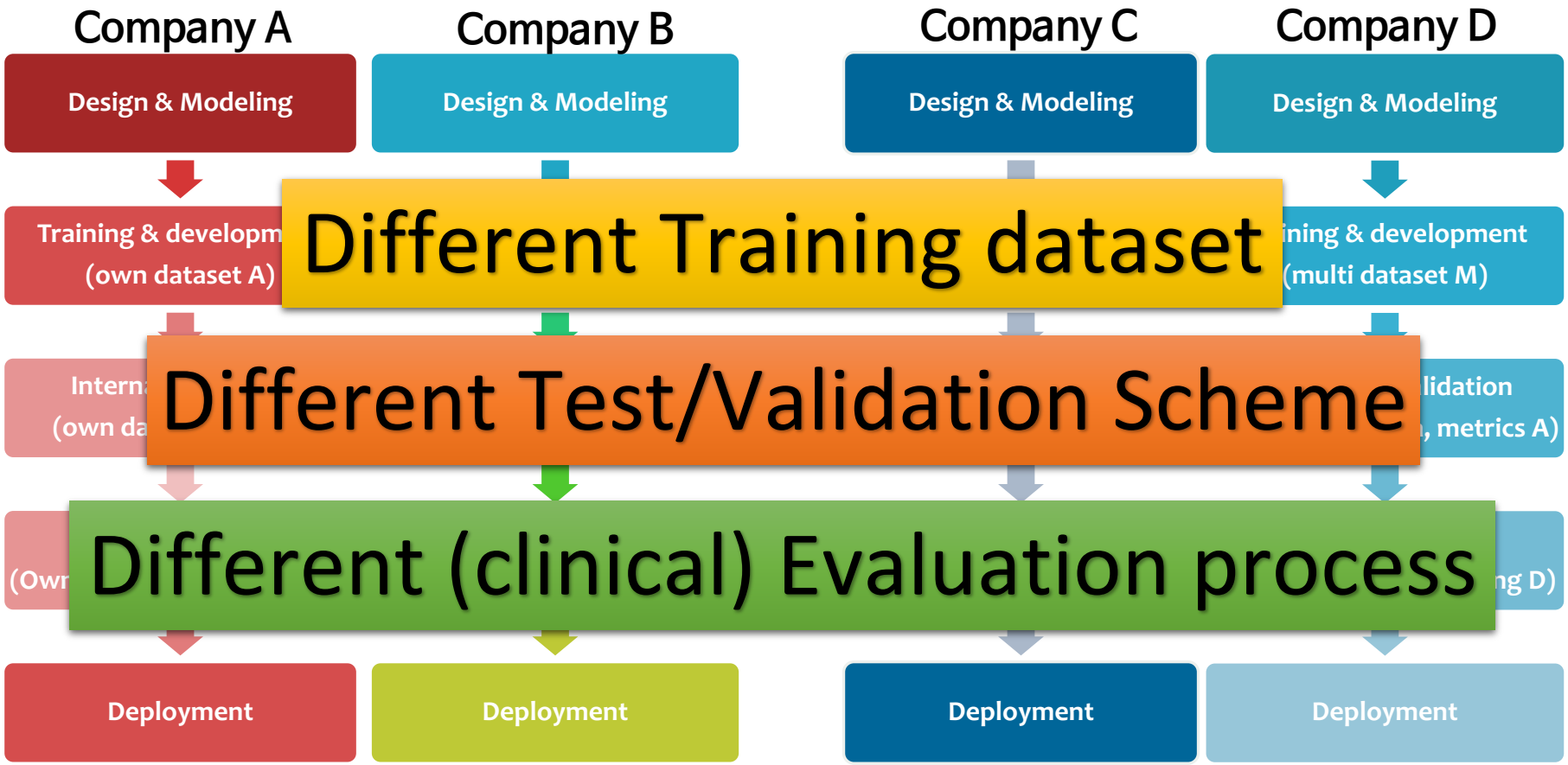
Performance evaluation of a medical device is **a continuous process** by which data are assessed and analysed to demonstrate the scientific validity, technical(analytical) performance and clinical performance of that device for **its intended purpose as stated by the manufacturer.**

The Manufacturer used assessment and analysis of clinical data of a device to **verify the clinical safety, performance, and effectiveness** of the device.

Performance evaluation in machine learning is crucial for several reasons, which underscore its necessity:

- **Task processing ability verification**
  - Performance evaluation allows us to verify how well a model can make task's goal(predictions or classifications). This helps determine if the model is ready for real-world data application.
- **Model Comparison**
  - When multiple models have been developed, performance evaluation is necessary to compare which model performs better. This enables the selection of the best model.
- **Detecting Overfitting**
  - It allows for the detection of overfitting, where a model is too closely fitted to the training data and performs poorly on new data. By comparing the performance of a model on training data versus validation or test data, overfitting can be identified.
- **Hyperparameter Tuning**
  - Performance evaluation is essential during the process of tuning hyperparameters to optimize model performance. By experimenting with various hyperparameter settings and evaluating the outcomes, the optimal configuration can be discovered.
- **Evaluating Generalization Ability**
  - It's important to ensure that a model not only performs well on training data but also on new, unseen data. Performance evaluation helps assess a model's generalization ability.
- **Achieving Business Objectives**
  - Performance evaluation is needed to verify if a machine learning model meets business goals or requirements. For instance, specific accuracy or response time may be required, and its achievement needs to be evaluated.
- **Providing Reliability and Transparency**
  - Systematically evaluating and documenting model performance enhances its reliability and provides transparency to users or stakeholders.

Why Performance evaluation process ?



how can we evaluate by the standardized way?

History of IEC TC62 PT8 (PT 63521)

AIRIS 2024
Use of AI in Medical Product Development
AI Regulatory & International Symposium
Co-organized by MFDS and U.S. FDA

- Oct 15,2021 – TC62 CIB (62/399/Q, 6 week, by Nov 26)
- Dec 1, 2021 – Established PT8 and registered PWI62-3 (according to 62/409e/RQ)
  - Project Leader: Jonghong Jeon (KR)
- Dec 14, 2021 – Kickoff meeting of PT8
- Sep 8, 2022 – **18th meeting of PT8 (1st phase)**
- Sep 15, 2022 – NP submit to TC62 plenary
- Nov 11, 2022 – CAG presentation and Resolution 62/2022/13
  - TC 62 appreciates the draft for a new work item proposal submitted by PT 8 and requests a further improvement of the document by the end of 2023, in particular by clarifying the terminology and the concepts used. TC 62 hopes to circulate the NP by early 2024.
- May 2, 2023 – kickoff meeting of PT8 (2nd phase)
- July 17, 2023 – **9th meeting of PT8 (2nd phase)**
- July 25, 2023 – 2nd NP submit to TC62 plenary
- Aug 11, 2023 – Start 62/474/NP ballot (until 11/03)
- Nov 3, 2023 – Approved NP, Changed Project Team number to PT 63521
- 1st PT63521 meeting : Nov 29
- 2nd PT63521 meeting : Dec 13
- 3rd PT63521 meeting : Jan 9, 2024 ~ (bi-weekly)

Current Member of PT63521

- 13 nation
  - BE, CH, CN, DE, FR, GB, IN, IT, JP, KR, NG, NL, US
- 33 experts
  - Medtronic, Phillips, Elekta, Varian, Siemens-healthineers, esaote, CFDA, deepeyevision, kakaohealth, ETRI …

TC 62/PT 63521 - Develop IEC 63521 ED1 "Machine Learning-enabled Medical Device – Performance Evaluation Process"

List All | Experts by NC

Function	Last Name	First Name	NC	E-mail address
Member	Anyaoha	Samuel	NG	an uel@gmail.com
Member	Bang	Soo-Young	KR	m .kr
Member	Cabitza	Federico	IT	fe za@unimib.it
Member	Chieregato	Matteo	IT	m gato@poliambulanza.it
Member	Elen	Bart	BE	b o.be
Member	Guidi	Gabriele	IT	g @aou.mo.it
Member	Guraschi	Nicola	IT	ni hi@esaote.com
Member	HAN	Taehwa	KR	ta gmail.com
Member	Ivanovic	Sasa	CH	s @varian.com
Member	Jakob	Thomas	CH	th @varian.com
Project Leader	Jeon	Jong Hong	KR	h re.kr
Member	Keshvari	Jafar	BE	ja @huawei.com
Member	Kim	Hwiyoung	KR	H uhs.ac
Member	KONDO	Yusuke	JP	y- epeyevision.com
Member	Kowalski	Dominik	DE	d lski@brainlab.com
Member	Krantz-Zuppan	Patty	US	p uppan@medtronic.com
Member	LAWAL	Ismaila	NG	is hoo.com
Member	liu	zhongsheng	CN	c mail.com
Member	Matsumoto	Koichiro	JP	k e.auone-net.jp
Member	Meyer	Martin	ISO	m eyer@siemens-healthineers.com
Member	MICHAUD	Cyrille	FR	m 1.io
Member	Moon	In Hyuk	KR	in @gmail.com
Member	Moons	K.	NL	K. @umcutrecht.nl
Member	OKHAREDIA	Taiwo	NG	t hoo.com
Member	Ramakrishnan	S.	IN	s ac.in
Member	Santhanam	Ramya	IN	r nam@siemens-healthineers.com
Member	Sayeed	Abdul	GB	a @elekta.com
Member	Shin	Soo Yong	KR	s akaohealthcare.com
Member	Stoll	Ronny	DE	r vasis.biz
Member	Traverso	Paolo	IT	p @esaote.com
Member	van den Brink	Johan	NL	j .brink@philips.com
Member	Varriale	Rosario	IT	r le@esaote.com
Member	Wang	Hao	CN	d il.com

Total number of experts : 33
Total number of records : 33

Goal of IEC 63521

1. It is **based on the existing standard systems of IEC TC62** and should be avoided as much as possible of redundancy.
2. It should **provide the process (or framework) for verifying the safety/efficacy of AI/ML-MD**.
3. Should be **applicable to various types of AI/ML-MD**
4. Should be able **to support various AI/ML tasks, models, data modalities, evaluation metrics, etc.**
5. It should be **applicable to various types of medical devices (PEMS, SaMD, IVD, Dentistry, Implantable, and medical robots, etc.) that utilize AI/ML technology**
6. Evaluation requirements, methods, standards, etc. should be **established through the process.**
7. The evaluation process should be **able to support conformity to global regulations**

Scope of IEC 63521

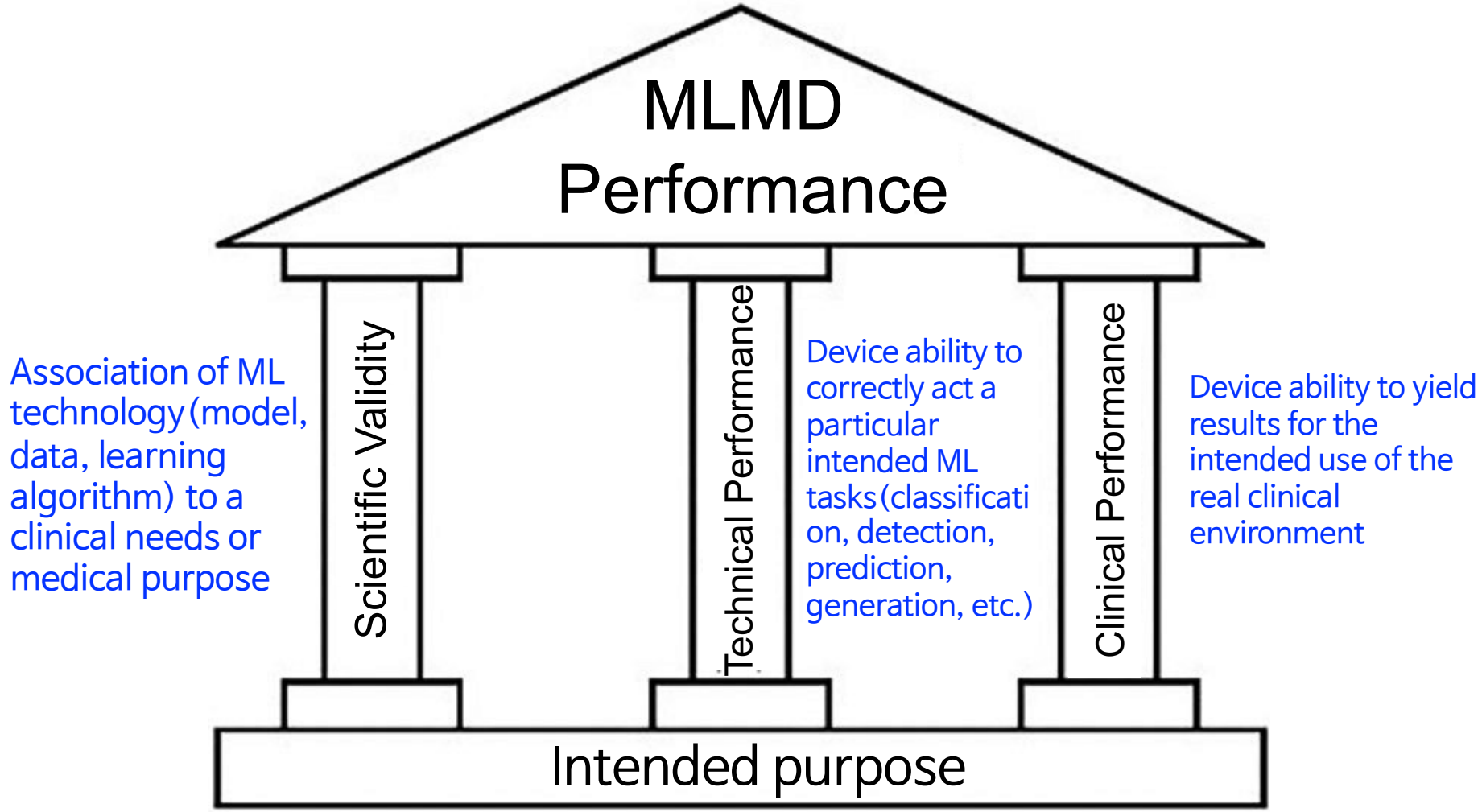
- This document **defines a standardized performance evaluation process** for Machine Learning-enabled Medical Devices. The **set of processes, activities, and tasks** described in this document establishes a common framework for MLMD performance evaluation processes.

- This process is to **assist manufacturers to evaluate** the ML suitability, the technical and clinical performance of the MLMD. It may be used for evaluation the performance to assure expected performance during post-market monitoring.

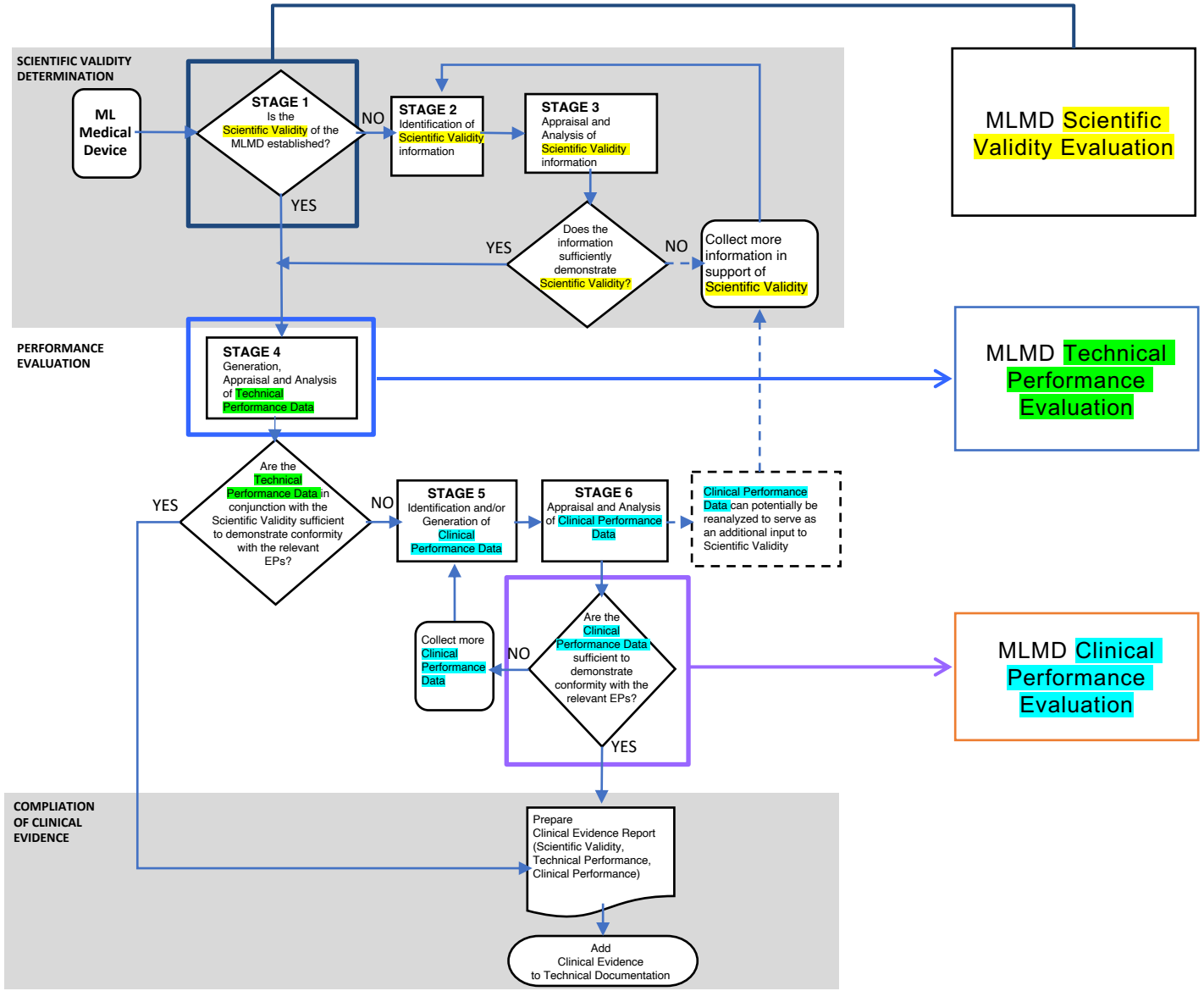
- This document is applicable to the **performance evaluation of all forms of MLMD**, comprising ML components and, where appropriate, **the integrated ML component with non-ML components**. The most important factor to consider performance evaluation of ML is whether it affects the intended purpose and the safety and effectiveness of medical device.

Basic [1] - Three pillar model

Three pillar model of Performance



Basic [2] - Evaluation flow



Basic [3] : Safety and Effectiveness of ML

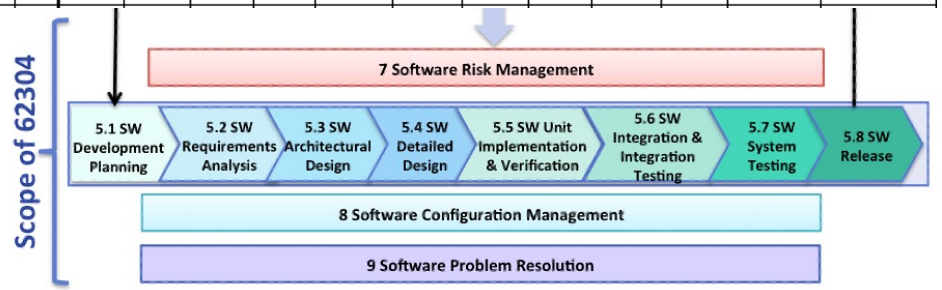
AIRIS 2024
Use of AI in Medical Product Development
AI Regulatory & International Symposium
Co-organized by MFDS and U.S. FDA

1. **Data quality and management**
   - Data plays a crucial role in the performance of machine learning medical devices. Ensure accurate data collection, preprocessing, labeling, and annotation.
2. **Model validation and verification**
   - Ensure the machine learning model serves its intended purpose and provides accurate predictions.
3. **Algorithm transparency and explainability**
   - The decision-making process of the machine learning model should be understandable and transparent.
4. **Bias and fairness**
   - The machine learning model should provide unbiased predictions for different patient groups and clinical situations.
5. **Robustness and generalization**
   - The machine learning model should be able to adapt to different patient populations, clinical situations, and data quality changes.
6. **Personal data protection and data security**
   - Machine learning medical devices must protect patient privacy and ensure data is secure from external threats.
7. **Software and hardware stability**
   - The software and hardware components of machine learning medical devices must be safe from errors, defects, and failures.
8. **User interface and human factors consideration**
   - The user interface of machine learning medical devices should minimize user errors and enable effective device use.
9. **Clinical evaluation and validation**
   - The safety and effectiveness of machine learning medical devices must be validated in real clinical environments.
10. **Benefit and risk**
    - AI should never cause foreseeable or unintentional harm

PE Requirement analysis(...ing)

AIRIS 2024
Use of AI in Medical Product Development
AI Regulatory & International Symposium
Co-organized by MFDS and U.S. FDA

No.	category	Candidate Requirements	Safety & Effectiveness Category											Three pillar model			5. Development process							6. maintenance PROCESS			7. RISK MANAGEMENT PROCESS	8. configuration management PROCESS.
			G	1	2	3	4	5	6	7	8	9	10	SV	TPV	CPV	5.1 development planning	5.2 requirements analysis	5.3 architectural design	5.4 detailed design	5.5 UNIT implementation and verification	5.6 integration and integration testing	5.7 release	6.1 maintenance plan	6.2 Problem and modification analysis	6.3 Modification implementation		
1	ABSTRACT	A summary shall be provided, detailing objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	O											O														
2	Background and objectives	The scientific background and rationale for the study shall be explained, including the medical context (diagnostic or prognostic) and references to existing models.	O											O														
3		Specific objectives or hypotheses shall be clearly stated.	O											O														
4	Background and rationale	Any pre-existing evidence supporting the AI intervention shall be described.										O		O														
5		The intended use of the AI intervention in the clinical pathway shall be explained, including its purpose and intended users (e.g., healthcare professionals, patients, public).										O		O														
6		The rationale for the choice of comparators in the study shall be provided.										O		O														
7	clinical problem	The clinical problem for the model's application shall be detailed.										O		O														
8	cohorts	Characteristics of the cohorts (training and test sets) shall be described in detail.		O		O	O		O			O		O														
9		The representativeness of the cohorts (training and test sets) in real-world clinical settings shall be demonstrated.		O		O	O		O			O		O														
10	Intended use	The targeted medical condition(s) and problem(s), current standard practice, and intended patient population(s) shall be described.			O							O		O														
11	SOTA	The state-of-the-art solution used as a baseline for comparison shall be identified and detailed.			O									O														
12	**Data**	De-identification methods shall be described.		O					O						O													
13		The sufficiency of data for training the algorithm shall be evaluated.		O		O	O	O							O													
14		Data sources shall be identified.		O		O	O		O						O													
15		Data pre-processing steps shall be described.		O		O	O		O						O													
16		The quality of data used for training the algorithm shall be assessed.		O			O	O							O													
17		The extent of data accuracy and bias-free nature shall be evaluated.		O			O								O													
18		Standardization and interoperability of data shall be confirmed.		O		O		O							O													
19	**Data collection methods**	Plans for assessing and collecting trial data shall be outlined, including processes to promote data quality.		O		O		O							O													
20	**data independence**	Independence between training and test sets shall be confirmed.		O	O	O	O								O													
21	**Data monitoring**	Composition and role of the Data Monitoring Committee shall be detailed, including its independence and conflict of interest policies.		O		O	O								O													
22	**Data Partitions**	Intended sample size and determination method shall be stated.		O			O								O													
23		Data assignment to partitions and proportions shall be specified.		O			O								O													
24	Ground Truth	Definition of the ground truth reference standard shall be sufficiently detailed for replication.		O		O	O	O							O													
25		Rationale for choosing the reference standard shall be provided.		O		O	O								O													
26		Source of ground-truth annotations and annotator qualifications shall be described.		O		O	O								O													
27		Annotation tools and measurement of variability shall be detailed.		O		O	O								O													
28		Methods for mitigating variability and resolving discrepancies shall be outlined.		O		O	O								O													
29	models	Information on evaluated models and selection criteria shall be provided.		O	O	O	O	O							O													
30	transformation	Data transformations prior to model application shall be described.		O		O	O								O													
31	Confidentiality	Methods for collecting, sharing, and maintaining personal information about participants to protect confidentiality shall be detailed.							O						O													
32	Consent or ascent	The process for obtaining informed consent or assent from potential trial participants or authorized surrogates shall be specified.							O						O													
33	Declaration of interests	Financial and other competing interests of principal investigators for the overall trial and each study site shall be disclosed.							O						O													
34	Research ethics approval	Plans for seeking research ethics committee/institutional review board approval shall be outlined.							O						O													



IEC 63521 : Focused on Safety and Effectiveness of ML

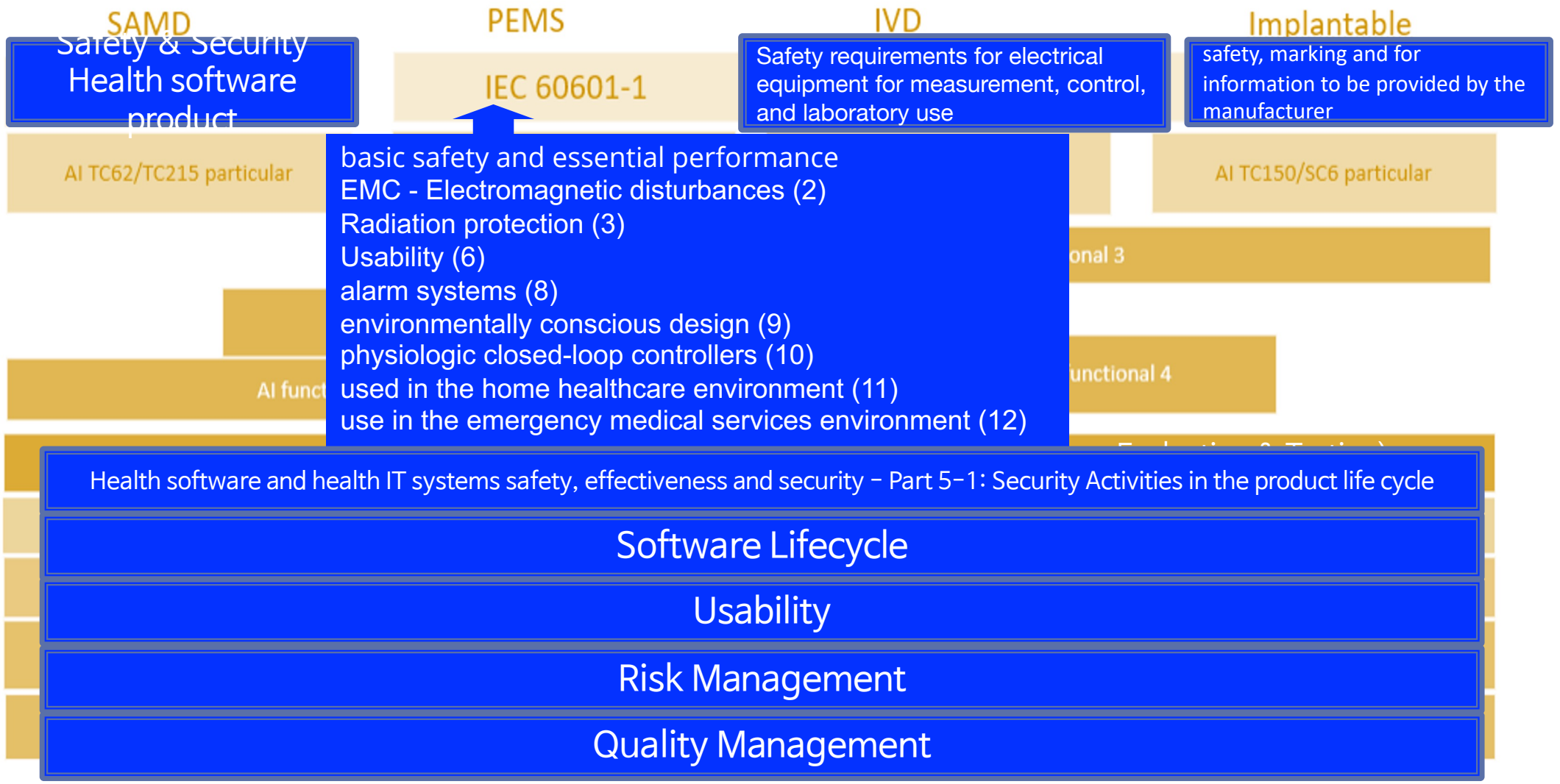


Figure 1 in the 4th SNAIG Report (62/432/INF)

IEC 63521 : Focused on Safety and Effectiveness of ML

basic safety and essential performance
EMC - Electromagnetic disturbances (2)
Radiation protection (3)
Usability (6)
alarm systems (8)
environmentally conscious design (9)
physiologic closed-loop controllers (10)
used in the home healthcare environment (11)
use in the emergency medical services environment (12)

SAMD

Safety & Security Health software product

IVD

Safety requirements for electrical equipment for measurement, control, and laboratory use

Implantable

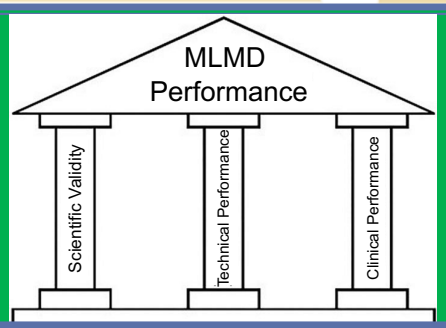
safety, marking and for information to be provided by the manufacturer

AI TC62/TC215 particular

AI TC62 particular

AI IVD particular

AI TC150/SC6 particular



Safety & effectiveness of ML

Health software and health IT systems safety, effectiveness and security – Part 5-1: Security Activities in the product life cycle

Software Lifecycle

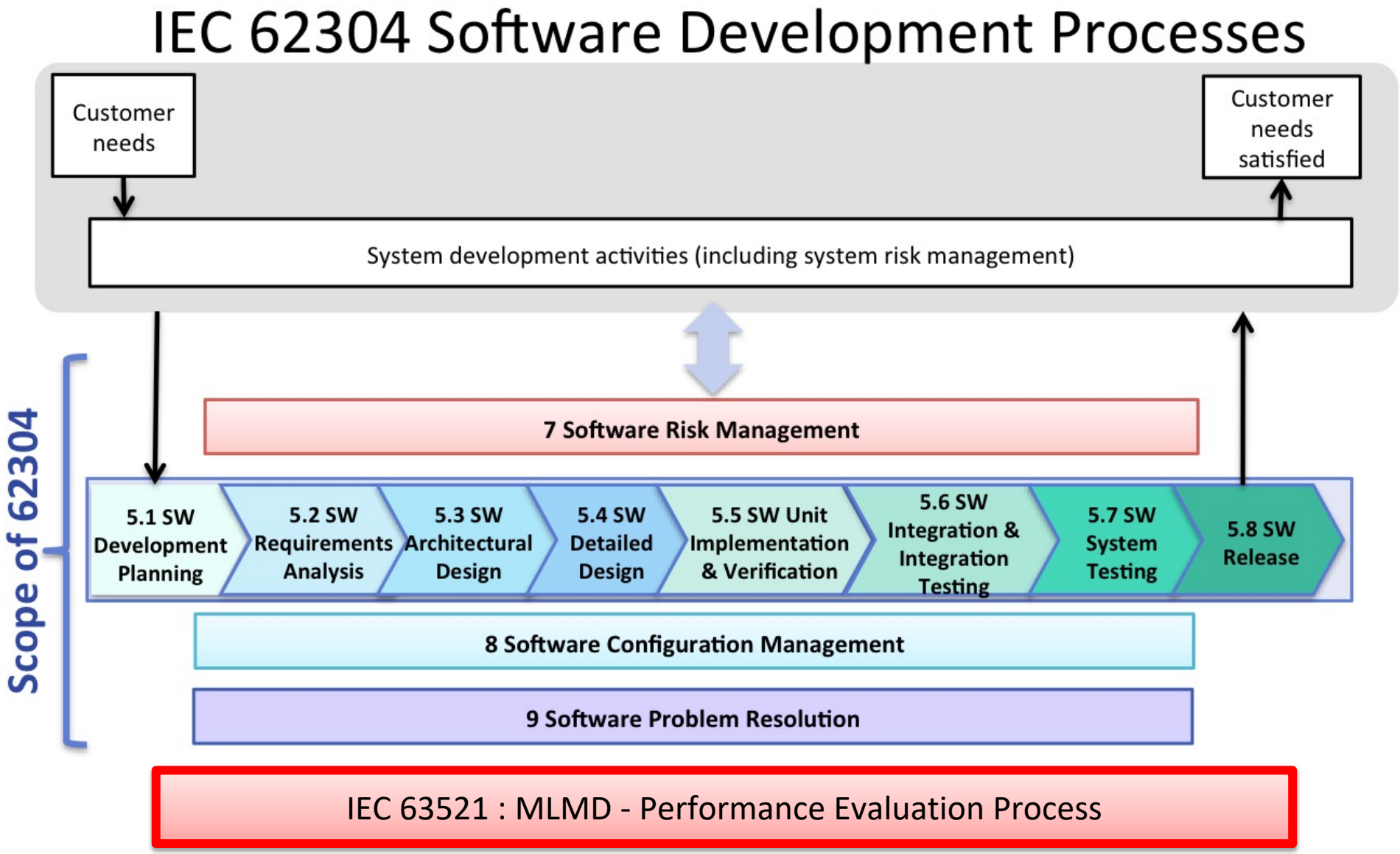
Usability

Risk Management

Quality Management

Figure 1 in the 4th SNAIG Report (62/432/INF)

Considerations 1: Align with IEC 62304 process



Considerations 2: IEC TC62's AI standardization roadmap

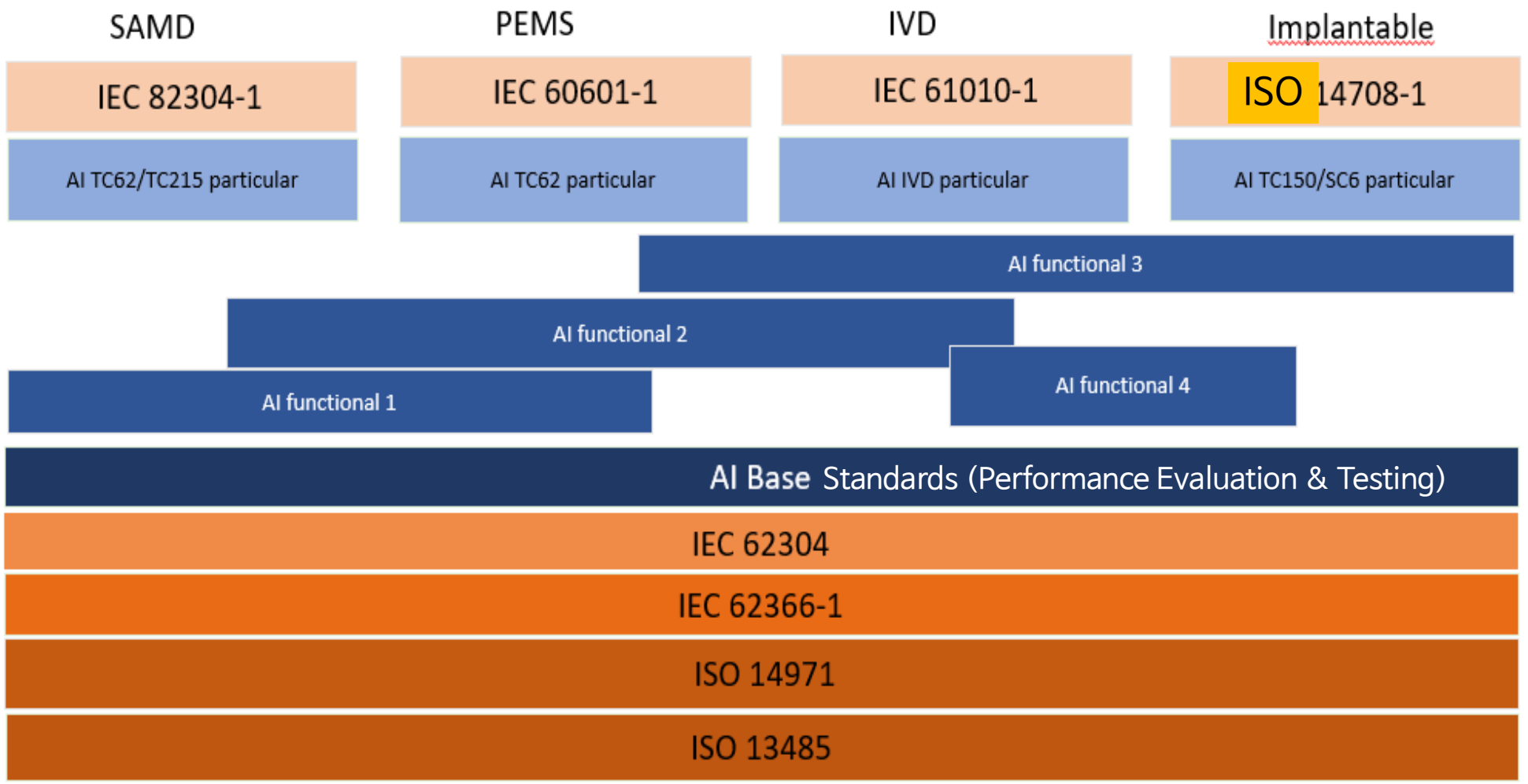


Figure 1 in the 4th SNAIG Report (62/432/INF)

IEC TC62's AI standardization roadmap - extend

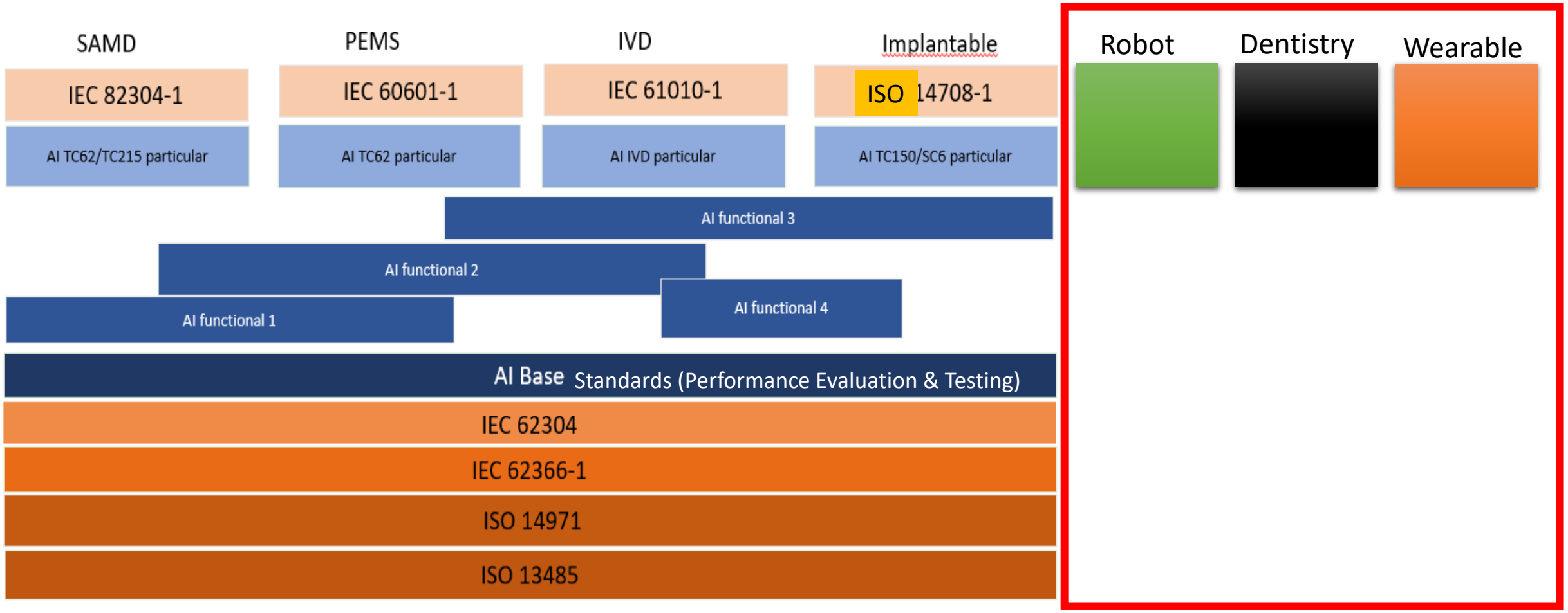


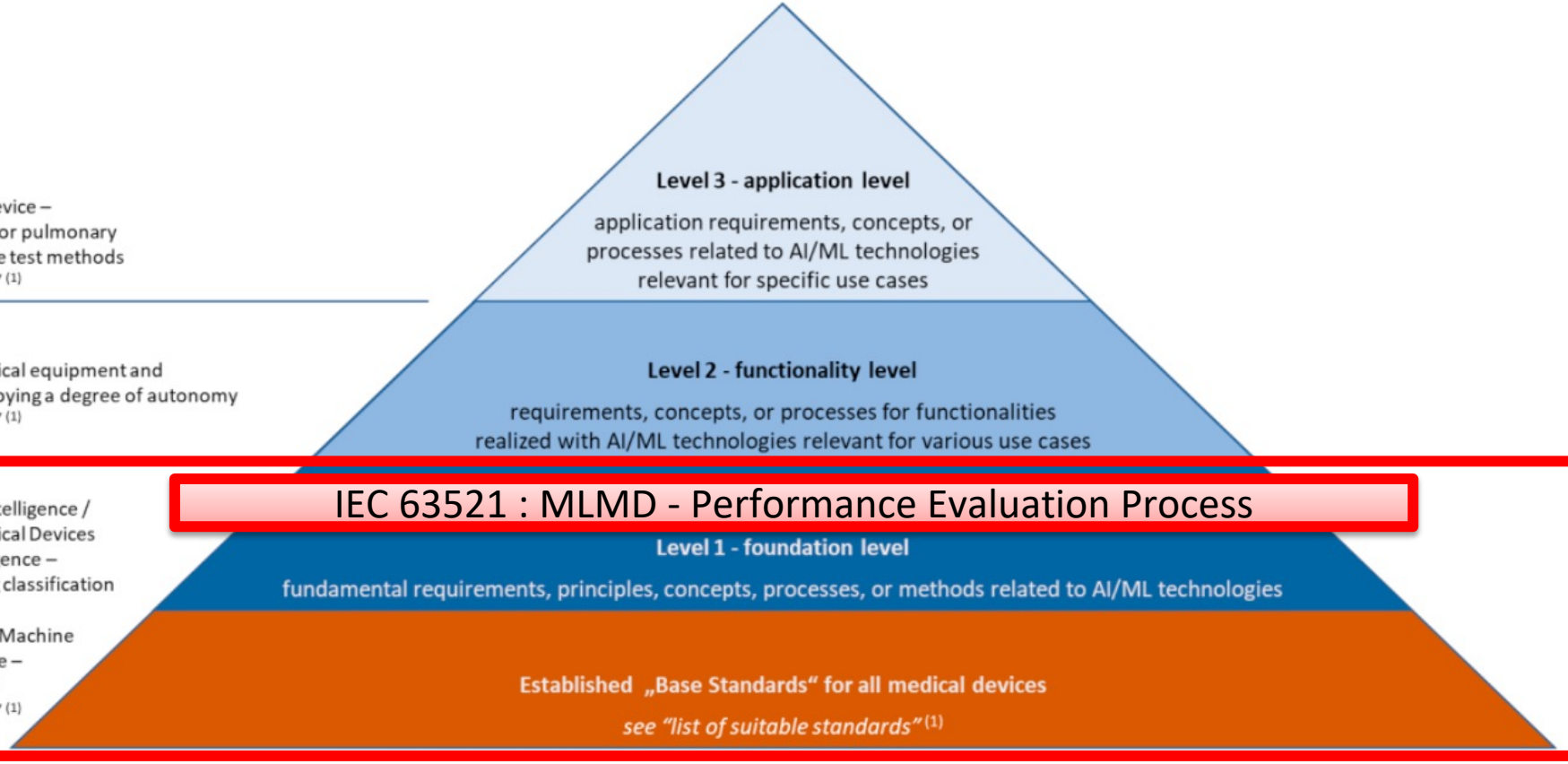
Figure 1 in the 4th SNAIG Report (62/432/INF)

How to role as Foundational level standard

Level 3 - application level
- **IEC 63524** AI enabled medical device – Computer assisted analysis SW for pulmonary images – Algorithm performance test methods
- … *see "list of suitable standards"* (1)

Level 2 - functionality level
- **IEC TR 60601-4-1** Medical electrical equipment and medical electrical systems employing a degree of autonomy
- … *see "list of suitable standards"* (1)

Level 1 - foundation level
- **IEC 63450** Testing of Artificial Intelligence / Machine Learning-enabled Medical Devices
- **ISO/IEC TS 4213** Artificial intelligence – Assessment of machine learning classification performance
- **IEC 63521** Artificial Intelligence/Machine Learning-enabled Medical Device – Performance Evaluation Process
- … *see "list of suitable standards"* (1)

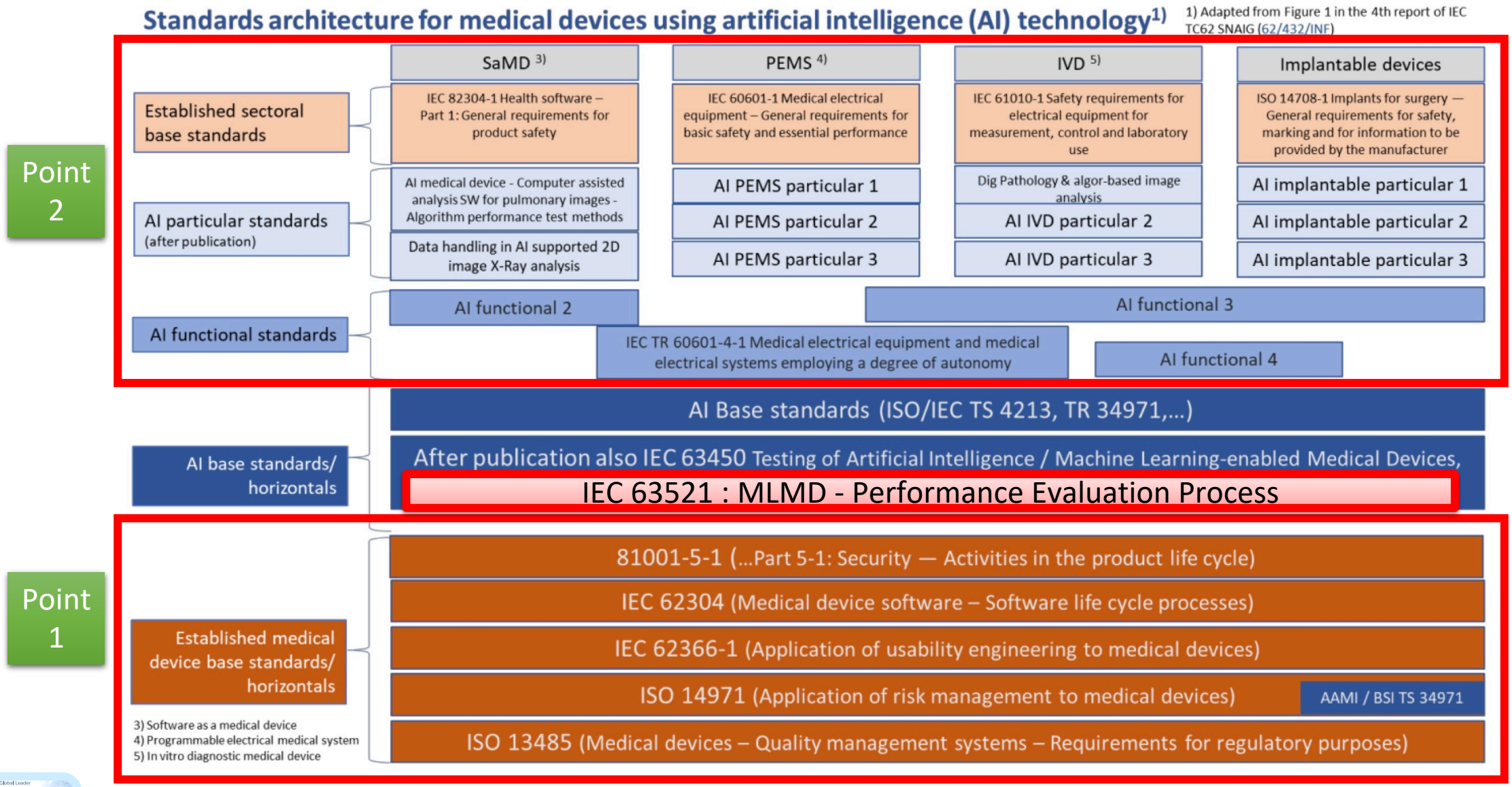


Established general safety standards for medical devices

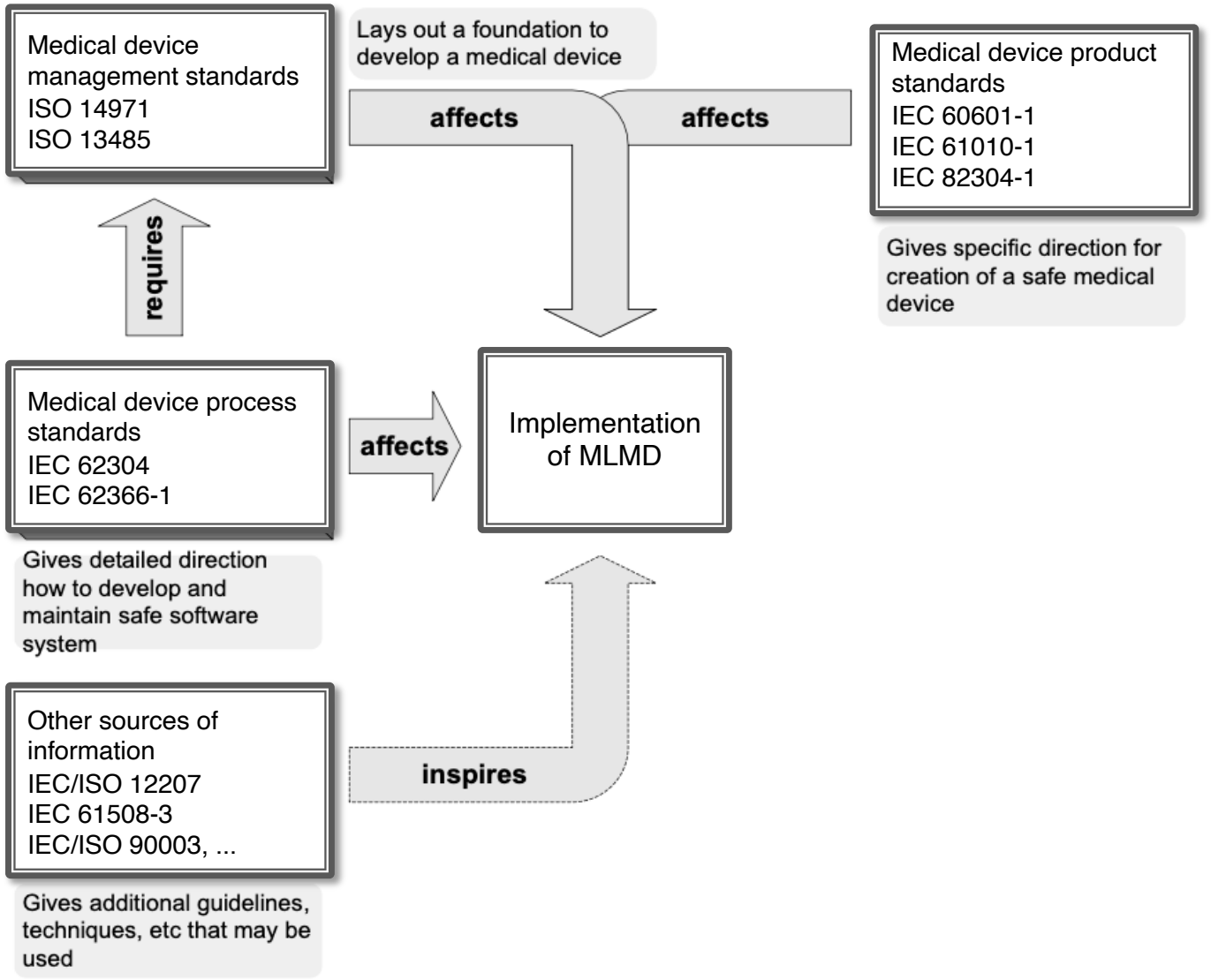
IEC 82304-1 Health software – Part 1: General requirements for product safety	IEC 60601-1 Medical electrical equipment – General requirements for basic safety and essential performance	IEC 61010-1 Safety requirements for electrical equipment for measurement, control and laboratory use	ISO 14708-1 Implants for surgery — General requirements for safety, marking and for information to be provided by the manufacturer
SaMD	PEMS	IVD	Implantable devices

(1) IEC TC 62 supporting document "Standards suitable for use in the development of AI/ML standards in the medical area"

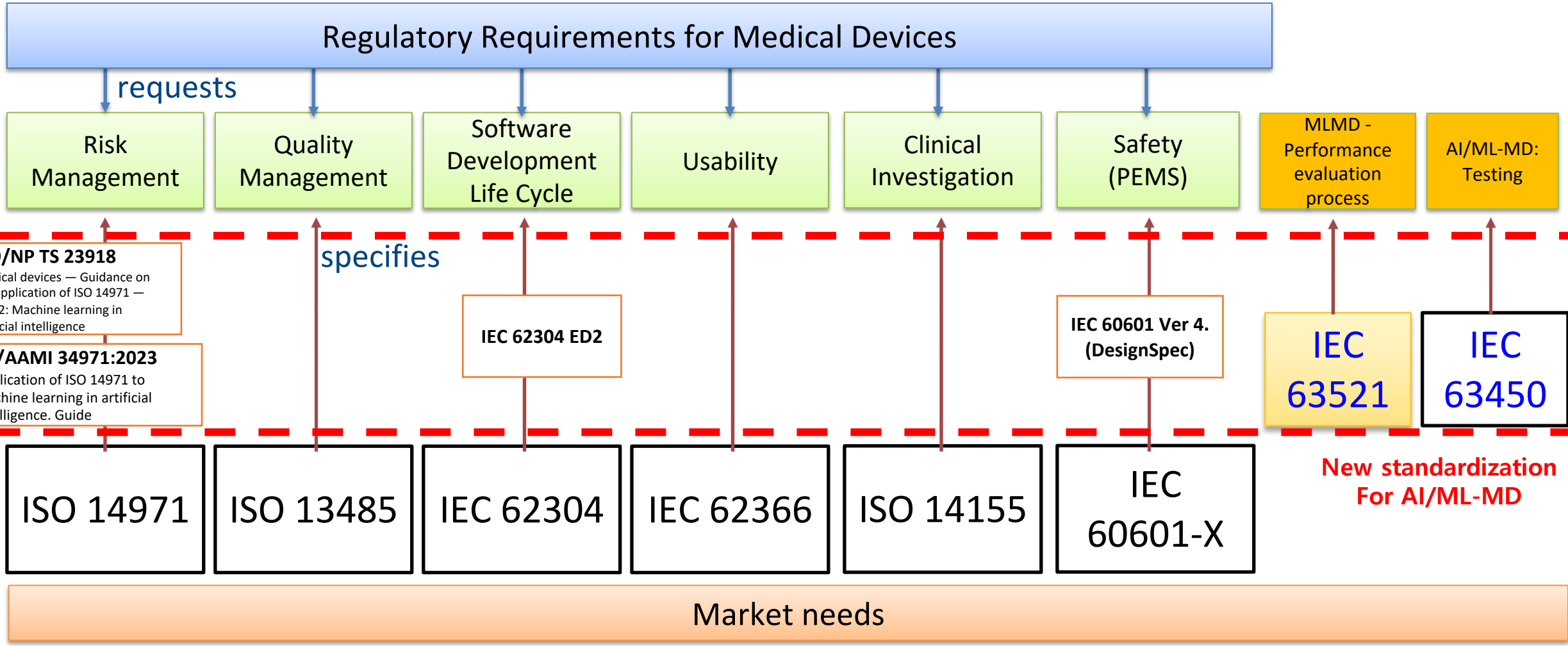
Align with other standards



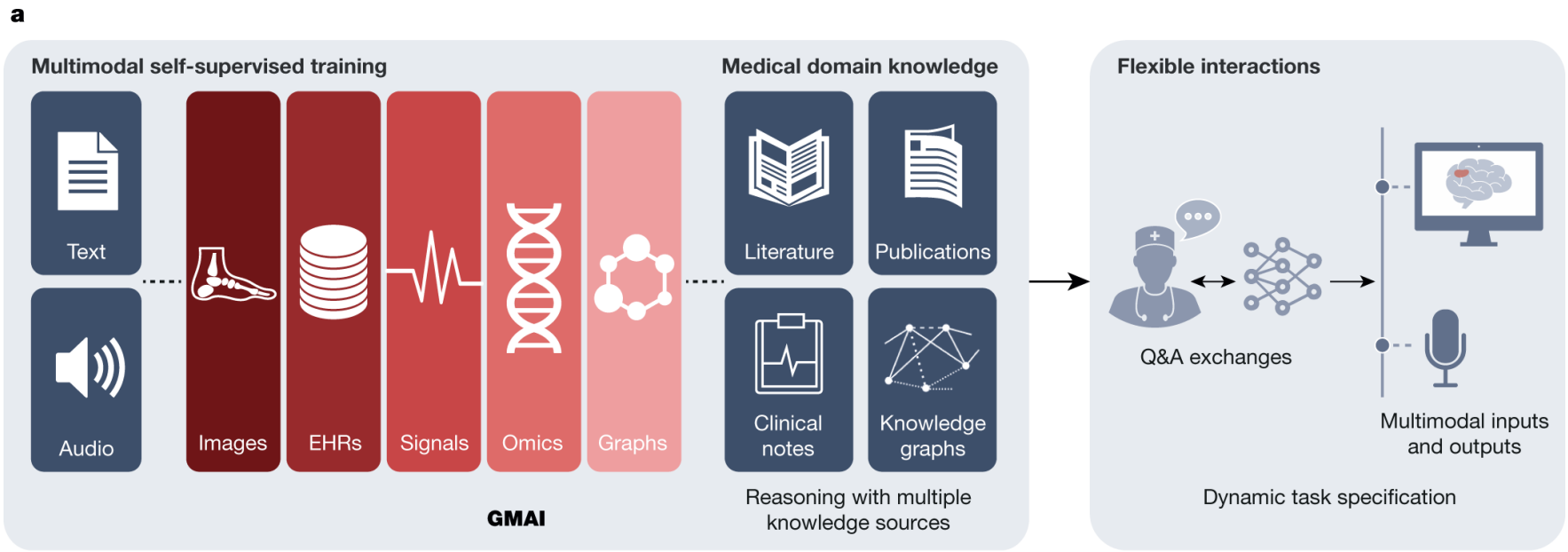
Align with other standards : point 1

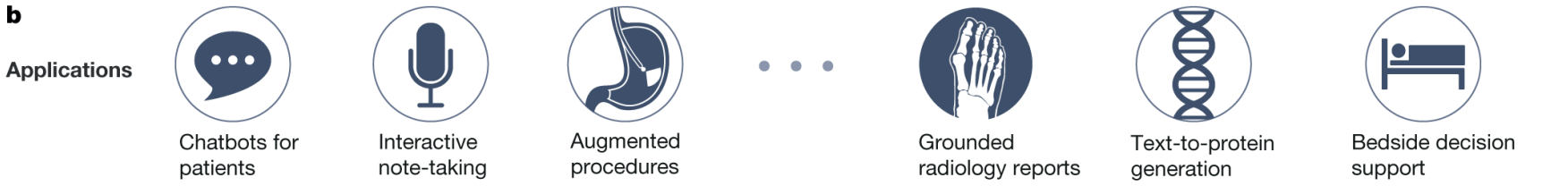


Considerations 3: Regulation requirements and International standardization

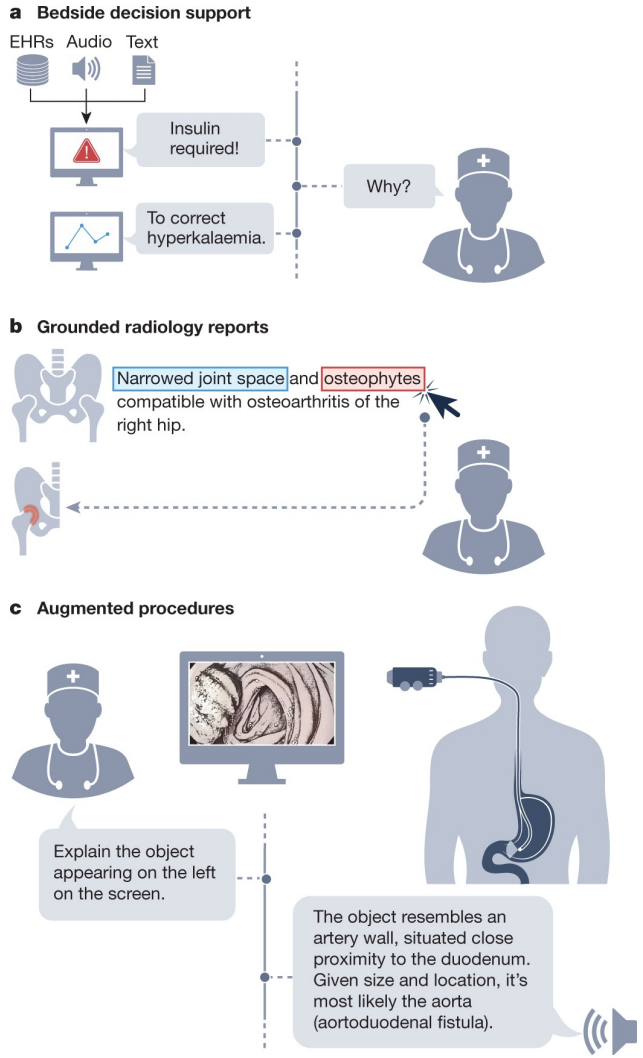


Considerations 4: GenAI, Foundation model, LLM





Regulations: Application approval; validation; audits; community-based challenges; analyses of biases, fairness and diversity



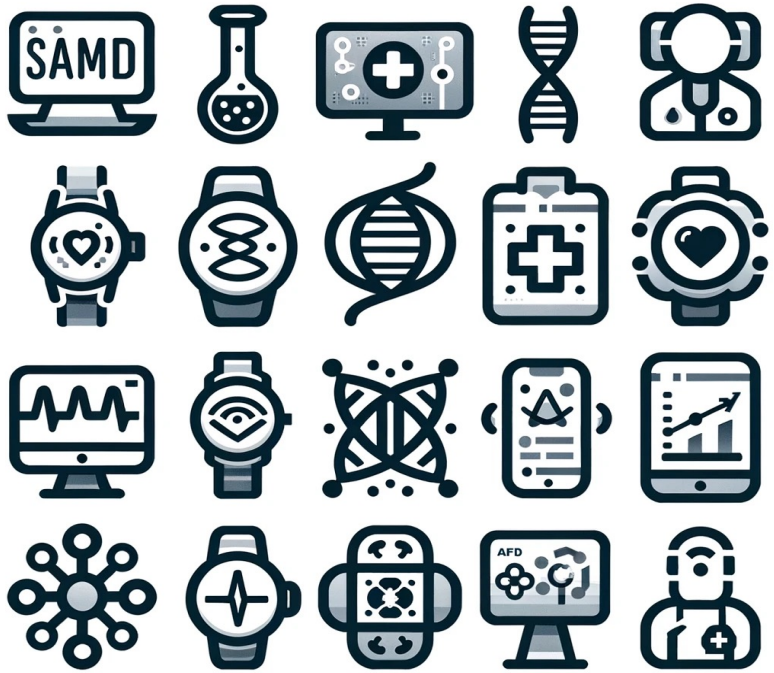
Source: https://www.nature.com/articles/s41586-023-05881-4

Conclusions

- History, Goal, Member, Schedule of PT 63521
- Why Performance Evaluation ?
- Three pillar model of Performance
  - **Scientific Validity requirements,**
  - **Technical (analytical) performance requirements**
  - **Clinical performance requirements**
- Current major discussion topics of PT 63521
  - Safety and Effectiveness of ML
  - TPLC & MLLC & performance evaluation
  - V-model or W-model
  - Align with baseline standards - 62304, 13485, 14971, 62366 …
  - TC 62's AI standardization roadmap & considerations
- Future considerations
  - Global harmonization & AI regulation/law

IEC 63521

Machine Learning-enabled Medical Device – Performance Evaluation Process



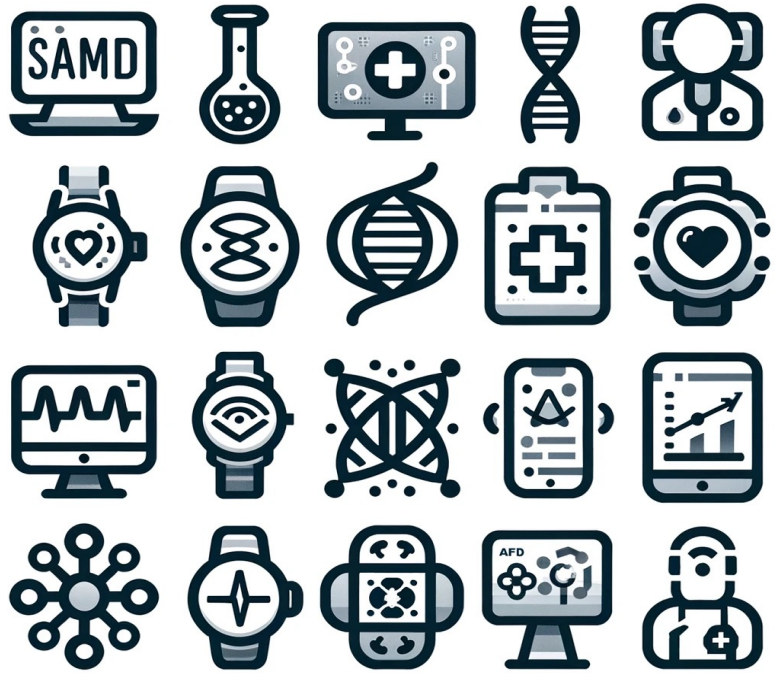
Conclusions

AIRIS 2024
Use of AI in Medical Product Development
AI Regulatory & International Symposium
Co-organized by MFDS and U.S. FDA

Please join us

IEC 63521

Machine Learning-enabled Medical Device – Performance Evaluation Process



Jonghong Jeon (hollobit@etri.re.kr)
+82-42-860-5333

https://www.linkedin.com/in/hollobit