

# ITU-T Focus Group Deliverable

(03/2023)

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

---

## **FG-AI4H DEL2.1**

**Mapping of IMDRF essential principles to AI for health software**





# ITU-T FG-AI4H Deliverable DEL2.1

## Mapping of IMDRF essential principles to AI for health software

### Summary

AI for health (AI4H) software provides a number of new aspects that have not been considered when developing the regulatory framework for Software as a Medical Device (SaMD), as described by the IMDRF Essential Principles (EPs) in "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.

This document provides a suggested mapping of the EPs to the related aspects of AI4H software. Its purpose is to cover all aspects considered in the regulation of SaMDs and whether, and if yes, how they are applicable to AI4H.

In clause 7.1 the IMDRF EPs are evaluated for their applicability to AI4H. This reduces the number of relevant EPs to six: General (5.1), Clinical Evaluation (5.2), Medical Devices and IVD Medical Devices that Incorporate Software or are Software as a Medical Device (5.8), Labeling (5.10), Protection against the Risks posed by Medical Devices and IVD Medical Devices intended by the Manufacturer for use by Lay Users (5.12) and Performance Characteristics (7.2). The key concepts of these EPs are extracted in clause 7.2 and grouped with related AI4H concepts in clause 7.3. Finally, in clause 7.4 the explicit mapping from EPs to AI4H is presented.

### Keywords

AI for health (AI4H), essential principles (EPs), mapping, software as medical device (SaMD).

### 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 Version 1 of the Deliverable DEL2.1 on "*Mapping of IMDRF essential principles to AI for health software*" approved on 16 March 2023 via the online approval process for the ITU-T Focus Group on AI for Health (FG-AI4H).

|                 |   |  |
|-----------------|---|--|
| <b>Editors:</b> | Luis Oala<br>Fraunhofer Heinrich Hertz Institute Germany        | E-mail: <a href="mailto:luis.oala@dotphoton.com">luis.oala@dotphoton.com</a> |
|                 | Pradeep Balachandran<br>Technical Consultant (eHealth)<br>India | Email: <a href="mailto:pbn.tvn@gmail.com">pbn.tvn@gmail.com</a>              |
|                 | Pat Baird<br>Philips<br>USA                                     | Email: <a href="mailto:pat.baird@philips.com">pat.baird@philips.com</a>      |

© ITU 2023

All rights reserved. No part of this publication may be reproduced, by any means whatsoever, without the prior written permission of ITU.

## Table of Contents

|  | <b>Page</b> |
|--|-------------|
| 1 Scope.....   | 1           |
| 2 References.....  | 1           |
| 3 Definitions .....  | 4           |
| 3.1 Terms defined elsewhere .....                            | 4           |
| 3.2 Terms defined in this Technical Report .....             | 4           |
| 4 Abbreviations and acronyms .....                           | 4           |
| 5 Conventions .....  | 5           |
| 6 Overview.....  | 5           |
| 7 Mapping of IMDRF Essential Principles to AI4H aspects..... | 5           |
| 7.1 Evaluation of applicability .....                        | 5           |
| 7.2 Extraction of key concepts from EPs .....                | 7           |
| 7.3 EPs clusters and AI4H concepts.....                      | 14          |
| 7.4 Mapping IMDRF EPs to AI4H concepts.....                  | 21          |

# ITU-T FG-AI4H Deliverable DEL2.1

## Mapping of IMDRF essential principles to AI for health software

### 1 Scope

AI for health (AI4H) software provides a number of new aspects that have not been considered when developing the regulatory framework for Software as a Medical Device (SaMD), as described by the IMDRF Essential Principles (EPs) in [IMDRF EPs].

This Technical Report provides a suggested mapping of the EPs to the related aspects of AI4H software. Its purpose is to cover all aspects considered in the regulation of SaMDs and whether, and if yes, how they are applicable to AI4H.

### 2 References

- [FG-AI4H DEL0.1] ITU-T FG-AI4H DEL0.1 (2022), *Common unified terms in artificial intelligence for health*, [https://www.itu.int/dms\\_pub/itu-t/opb/fg/T-FG-AI4H-2022-1-PDF-E.pdf](https://www.itu.int/dms_pub/itu-t/opb/fg/T-FG-AI4H-2022-1-PDF-E.pdf).
- [IMDRF-EPs] IMDRF (2018), *Essential Principles of Safety and Performance of Medical Devices and IVD Medical Devices*, IMDRF Good Regulatory Review Practices Group, IMDRF GRRP WG/N47 FINAL.
- [Albarghouthi et al.] Albarghouthi, A. and Vinitzky S. (2019), *Fairness-Aware Programming, FAT\* '19: Conference on Fairness, Accountability, and Transparency*, January Atlanta, GA, USA. ACM, New York, NY, USA, 9 pages. [Doi:10.1145/3287560.3287588](https://doi.org/10.1145/3287560.3287588).
- [Bach et al.] Bach, S. Binder, A., Montavon, G., Müller, K. and Samek, W. (2016), *Analyzing Classifiers: Fisher Vectors and Deep Neural Networks*, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 2912-2920.
- [Bellamy et al.] Bellamy, R. K. et al. (2018), *AI Fairness 360: An extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias*, Preprint at: [arXiv:1810.01943v1](https://arxiv.org/abs/1810.01943v1).
- [Bishop] Bishop, C.M. (2009), *Pattern recognition and machine learning*. Information science and statistics, Springer, New York, NY, corrected at 8th printing 2009 edition. OCLC: 845772798.
- [Breusch] Breusch, T.S. (1978), *Testing for Autocorrelation in Dynamic Linear Models\**, Australian Economic Papers, 17(31):334-355.
- [Denker et al.] Denker, J.S., Schwartz, D. B., Wittner, B.S., Solla, S.A., Howard, R.E., Jackel, L.D. and Hopfield, J.J. (1987), *Large Automatic Learning, Rule Extraction, and Generalization*. Complex Systems, 1.
- [Dobbe et al.] Dobbe, R. and Ames, M. (2019), *Translation Tutorial: Values, Reflection and Engagement in Automated Decision-Making*, Proceedings of ACM Conference on Fairness, Accountability, and Transparency (ACM FAT\* 2019). ACM, New York, NY, USA, Article 4, 2 pages. Retrieved from <https://drive.google.com/file/d/1vhkLBfDUtsd9hXfbXz8QwFo8GYGZyAWP/view>.
- [Duda et al.] Duda, R.O., Hart, P.E. and Stork, D.G. (2000), *Pattern Classification (2<sup>nd</sup> Edition)*, Wiley-Interscience, New York, NY, USA.

- [Dwork et al.] Dwork, C., Hardt, M., Pitassi, T., Reingold, O. and Zemel, R. (2012), *Fairness through awareness*, Proceedings of the 3rd Innovations in Theoretical Computer Science Conference (pp. 214-226). ACM.
- [Friedler et al.] Friedler, S. A., Scheidegger, C., Venkatasubramanian, S., Choudhary, S., Hamilton, E.P., and Roth, D. (2018), *A comparative study of fairness-enhancing interventions in machine learning*. Preprint at: [arXiv:1802.04422v1](https://arxiv.org/abs/1802.04422v1).
- [Frost et al.] Frost, N., Sabour, S. and Hinton, G. (2018), *Darccc: Detecting adversaries by reconstruction from class conditional capsules*. arXiv preprint arXiv:1811.06969.
- [Gajane et al.] Gajane, P. and Pechenizkiy, M. (2017), *On formalizing fairness in prediction with machine learning*. Preprint at: [arXiv:1710.03184](https://arxiv.org/abs/1710.03184).
- [Gal et al.] Gal, Y. and Ghahramani, Z. (2016), *Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning*, ICML.
- [Gebru et al.] Gebru, T., Morgenstern, J., Vecchione, B., Wortman Vaughan, J., Wallach, H., Daumé III, H. and Crawford, K. (2018), *Datasheets for datasets*. arXiv preprint arXiv:1803.09010.
- [Hayashi] Hayashi, F. (2000), *Econometrics*. Princeton Univ. Press, Princeton. OCLC: 247253903.
- [Hinton et al.] Hinton, G.E. and Neal, R. M. (1995), *Bayesian Learning for Neural Networks*.
- [Holstein et al.] Holstein, K., Vaughan, J.W., Daumé III, H., Dudík, M. and Wallach, H. (2018), *Improving fairness in machine learning systems: What do industry practitioners need?* ACM CHI Conference on Human Factors in Computing Systems (CHI 2019). Preprint at: [arXiv:1812.05239v2](https://arxiv.org/abs/1812.05239v2).
- [Huber] Huber, P. J. (1981), *Robust statistics*, Wiley, New York.
- [Ilyas et al., 2017] Ilyas, A., Jalal, A., Asteri, E., Daskalakis, C. and Dimakis, A.G. (2017), *The robust manifold defense: Adversarial training using generative models*, arXiv preprint arXiv:1712.09196.
- [Ilyas et al., 2019] Ilyas, A., Santurkar, S., Tsipras, D., Engstrom, L., Tran, B. and Madry, A. (2019), *Adversarial examples are not bugs, they are features*, arXiv preprint arXiv:1905.02175.
- [Madry et al.] Madry, A., Makelov, A., Schmidt, L., Tsipras, D. and Vladu, A. (2017), *Towards deep learning models resistant to adversarial attacks*, arXiv preprint arXiv:1706.06083.
- [Meng et al.] Meng, D. and Chen, H. (2017), *Magnet: a two-pronged defense against adversarial examples*, Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, pages 135–147. ACM.
- [Mitchell et al.] Mitchell, M. et al. (2019), *Model cards for model reporting*. Proceedings of the Conference on Fairness, Accountability, and Transparency (pp. 220-229). ACM.
- [Montavon et al.] Montavon, G., Lapuschkin, S., Binder, A., Samek, W. and Müller, K. (2017), *Explaining nonlinear classification decisions with deep Taylor decomposition*. Pattern Recognition, 65:211-222.

- [Moss] Moss, E.D. (2019), *Translation Tutorial: Toward a Theory of Race for Fairness in Machine Learning*, Proceedings of FAT\* conference (FAT\* Conference). ACM, New York, NY, USA, 2 pages. Retrieved from <https://drive.google.com/file/d/1jKsbnAg7NPotjihvNb8o4YXyjcjF8RTJW/view>.
- [Nix et al.] Nix, D. A. and Weigend, A. S. (1994), *Estimating the mean and variance of the target probability distribution*, Proceedings of 1994 IEEE International Conference on Neural Networks (ICNN'94), Vol. 1, pages 55-60, June.
- [Papernot et al.] Papernot, N., McDaniel, P., Sinha, A. and Wellman, M. (2016), *Towards the science of security and privacy in machine learning*. Preprint at: [arXiv:1611.03814](https://arxiv.org/abs/1611.03814).
- [Paul et al.] Paul, A. Jolley, C., and Anthony, A. (2018), *Reflecting the Past, Shaping the Future: Making AI Work for International Development*. Retrieved from USAID: <https://www.usaid.gov/sites/default/files/documents/15396/AI-ML-in-Development.pdf>.
- [Rasmussen et al.] Rasmussen, C.E. and Williams, C.K.I. (2008), *Gaussian processes for machine learning*. Adaptive computation and machine learning. MIT Press, Cambridge, Mass., 3. print edition. OCLC: 552376743.
- [Russell et al.] Russell, S., Dewey, D. and Tegmark, M. (2015), *Research priorities for robust and beneficial artificial intelligence*. Ai Magazine, 36(4), 105-114.
- [Samangouei et al.] Samangouei, P., Kabkab, M. and Chellappa, R. (2018), *Defense-gan: Protecting classifiers against adversarial attacks using generative models*. International Conference on Learning Representations, Vol. 9.
- [Schott et al.] Schott, L., Rauber, J., Brendel, W. and Bethge, M. (2018), *Robust perception through analysis by synthesis*, arXiv preprint arXiv:1805.09190.
- [Shafer et al.] Shafer, G. and Vovk, V. (2008). *A tutorial on conformal prediction*. Journal of Machine Learning Research, 9 (Mar), 371-421.
- [Shaikh et al.] Shaikh, S., Vishwakarma, H., Mehta, S., Varshney, K.R., Ramamurthy, K.N. and Wei, D. (2017), *An End-To-End Machine Learning Pipeline That Ensures Fairness Policies*. CoRR, abs/1710.06876. Preprint at: [arXiv:1710.06876v1](https://arxiv.org/abs/1710.06876v1).
- [Shen et al.] Shen, S., Jin, G., Gao, K. and Zhang, Y. (2017), *Ape-gan: Adversarial perturbation elimination with gan*. ICLR Submission, available on OpenReview, 4.
- [Shrikumar et al.] Shrikumar, A., Greenside, P. and Kundaje, A. (2017), *Learning Important Features Through Propagating Activation Differences*, ICML.
- [Singh et al.] Singh, G., Gehr, T., Mirman, M., Püschel, M. and Vechev, M. (2018). *Fast and effective robustness certification*. Advances in Neural Information Processing Systems, pp. 10802-10813.
- [Song et al.] Song, Y., Kim, T., Nowozin, S., Ermon, S. and Kushman, N. (2017), *Pixeldefend: Leveraging generative models to understand and defend against adversarial examples*, arXiv preprint arXiv:1710.10766.
- [Srinivasan et al.] Srinivasan, V., Marban, A., Müller, K., Samek, W. and Nakajima, S. (2018), *Counterstrike: Defending deep learning architectures against adversarial samples by langevin dynamics with supervised denoising autoencoder*. arXiv preprint arXiv:1805.12017.

- [Sundararajan et al.] Sundararajan, M., Taly, A. and Yan, Q. (2017), *Axiomatic Attribution for Deep Networks*, ICML.
- [Wachter et al.] Wachter, S., Mittelstadt, B. and Russell, C. (2017), *Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR*, Harv. JL & Tech., 31, 841.
- [White] White, H. (1980), *A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity*, *Econometrica*, 48(4):817–838.
- [Williams] Williams, C.K.I. (1996), *Computing with Infinite Networks*, NIPS.
- [Yona et al.] Yona, G. and Rothblum, G. (2018), *Probably Approximately Metric-Fair Learning*, International Conference on Machine Learning, pp. 5666-5674. Retrieved from <http://proceedings.mlr.press/v80/yona18a.html>.
- [Zeiler et al.] Zeiler, M.D. and Fergus, R. (2014), *Visualizing and Understanding Convolutional Networks*, ECCV.
- [Zheng et al.] Zheng, S., Song, Y., Leung, T. and Goodfellow, I. (2016), *Improving the Robustness of Deep Neural Networks via Stability Training*. arXiv:1604.04326 [cs], April. arXiv: 1604.04326.

### 3 Definitions

#### 3.1 Terms defined elsewhere

The terms defined in [FG-AI4H DEL0.1] are applicable to this Technical Report.

#### 3.2 Terms defined in this Technical Report

This Technical Report defines the following terms:

- 3.2.1 AI software lifecycle:** Software lifecycle concepts conveyed to AI software.
- 3.2.2 delay of the response:** Time delay of the AI software to provide an output after the input has been provided.
- 3.2.3 effectiveness:** A score that captures the performance of an AI determined over a test set.
- 3.2.4 out-of-sample input:** An input to the AI that is not part of the training set.
- 3.2.5 robustness:** Insensitivity of AI to small deviations from training conditions.
- 3.2.6 training set:** A dataset used for training AI software.
- 3.2.7 test set:** A dataset used for testing. The data set is undisclosed to the AI software provider.
- 3.2.8 uncertainty:** A score estimating the combined uncertainty present in inputs, as well as the AI software prediction.

### 4 Abbreviations and acronyms

This Technical Report uses the following abbreviations and acronyms:

|      |                         |
|------|-------------------------|
| AI   | Artificial Intelligence |
| AI4H | AI for Health           |



|       |   |
|-------|---|
| EP    | Essential Principle                           |
| FAT   | Factory Acceptance Testing                    |
| IMDRF | International Medical Device Regulators Forum |

## 5 Conventions

Applicable: The IMDRF EP is directly relevant to AI

Not applicable: The IMDRF EP is not directly relevant to AI

## 6 Overview

The mapping from the IMDRF Essential Principles (EPs) to concepts of AI4H software was produced using the following steps. First, the IMDRF EPs were reviewed by the authors to determine whether they may be applicable to AI4H. The result of this evaluation is summarized in the second row of Table 1. As is visible some EPs, such as 5.3 Chemical, Physical, and Biological Properties, may not be relevant for AI4H. Second, all applicable EPs were screened for keywords and concepts to enable the clustering of themes that reoccur throughout, such as risk management and specification of intended use. The original IMDRF EPs text, alongside the extracted keywords, are documented in clause 7.2. Third, related IMDRF EPs' key concepts were clustered into groups, for example one such group is 'Risk and Alarms'. For each of these groups relevant concepts and aspects from AI4H are clustered together. This clustered collection of AI4H concepts and aspects is documented in clause 7.3. Finally, the mapping between the EPs and AI4H concepts is presented in clause 7.4. This mapping may easily be extended by assigning new AI4H concepts to the EPs' clusters in clause 7.3 and updating the map accordingly in clause 7.4.

## 7 Mapping of IMDRF Essential Principles to AI4H aspects

### 7.1 Evaluation of applicability

The IMDRF Essential Principles comprise the following topics listed below. Their applicability to AI4H is summarized in Table 1.

- 5.0 Essential Principles Applicable to all Medical Devices and IVD Medical Devices
  - 5.1 General
  - 5.2 Clinical Evaluation
  - 5.3 Chemical, Physical, and Biological Properties
  - 5.4 Sterilization and Microbial Contamination
  - 5.5 Considerations of Environment and Conditions of Use
  - 5.6 Protection against Electrical, Mechanical and Thermal Risks
  - 5.7 Active Medical Devices and IVD Medical Devices and Medical Devices Connected to Them
  - 5.8 Medical Devices and IVD Medical Devices that Incorporate Software or are Software as a Medical Device
  - 5.9 Medical Devices and IVD Medical Devices with a Diagnostic or Measuring Function

- 5.10 Labelling
  - 5.11 Protection against Radiation
  - 5.12 Protection against the Risks posed by Medical Devices and IVD Medical Devices intended by the Manufacturer for use by Lay Users
  - 5.13 Medical Devices and IVD Medical Devices Incorporating Materials of Biological Origin
- 6.0 Essential Principles Applicable to Medical Devices other than IVD Medical Devices
- 6.1 Chemical, Physical and Biological Properties
  - 6.2 Protection against Radiation
  - 6.3 Particular Requirements for Implantable Medical Devices
  - 6.4 Protection against the Risks Posed to the Patient or User by Medical Devices Supplying Energy or Substances
  - 6.5 Medical Devices Incorporating a Substance Considered to be a Medicinal Product/Drug
- 7.0 Essential Principles Applicable to IVD Medical Devices
- 7.1 Chemical, Physical and Biological Properties
  - 7.2 Performance Characteristics

**Table 1 – Overview IMDRF EP sections applicability to AI4H, existing guidance or standards**

| EP Section      | 5.1 | 5.2 | 5.3 | 5.4 | 5.5 | 5.6 | 5.7 | 5.8 | 5.9 | 5.10 | 5.11 | 5.12 | 5.13 | 6.1 | 6.2 | 6.3 | 6.4 | 7.1 | 7.2 |
|-----------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|------|------|------|-----|-----|-----|-----|-----|-----|
| Applicable      | ✓   | ✓   | X   | X   | X   | X   | X   | ✓   | X   | ✓    | X    | ✓    | X    | X   | X   | X   | X   | X   | ✓   |
| Guidance exists | ✓   | ✓   | X   | X   | X   | X   | X   | ✓   | X   | ✓    | X    | X    | X    | X   | X   | ✓   | X   | X   | X   |
| Standards exist | ✓   | ✓   | ✓   | ✓   | ✓   | ✓   | ✓   | ✓   | ✓   | ✓    | ✓    | ✓    | ✓    | ✓   | ✓   | (✓) | ✓   | ✓   | ✓   |

Legend: ✓: applies to; X: does not apply to

## 7.2 Extraction of key concepts from EPs

In the following tables the key concepts of the IMDRF EPs are extracted. Each table corresponds to an EP and has three columns: the first column lists the EP subsection number, the second column lists the corresponding full text of the EP and the third column displays the extracted key concepts for that EP.

### 7.2.1 5.0 Essential Principles Applicable to all Medical Devices and IVD Medical Devices

#### 7.2.1.1 EP 5.1 General

| EP #  | EP text  | 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 lifecycle        |
|       | a) establish and document a risk management plan covering each medical device and IVD medical device;  | Risk management plan   |
|       | b) identify and analyse the known and foreseeable hazards associated with each medical device and IVD medical device;  | Identify and analyse hazards   |
|       | c) estimate and evaluate the risks associated with, and occurring during, the intended use and during reasonably foreseeable misuse;   | Risk; intended use; foreseeable misuse   |
|       | 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   |
|       | 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  |
|       | 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   |

| EP #  | EP text   | EP key concepts   |
|-------|---|---|
| 5.1.3 | <p>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:</p>   | Risk control measures; safety principles compliance; state of the art; risk control |
|       | a) eliminate or appropriately reduce risks through safe design and manufacture;   | Safe design   |
|       | 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   |
|       | 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  |
|       | 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  |
|       | 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   |

| EP #  | EP text   | EP key concepts    |
|-------|---|--------------------|
| 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 |

#### 7.2.1.2 EP 5.2 Clinical Evaluation

| EP #  | EP text   | EP key concepts   |
|-------|---|---|
| 5.2.1 | Where appropriate and depending on jurisdictional legal/regulatory requirements, a clinical evaluation may be required. A clinical evaluation should assess clinical data to establish that a favorable benefit-risk determination exists for the medical device and IVD medical device in the form of one or more of the following: <ul style="list-style-type: none"> <li>• clinical investigation reports (for IVDs, clinical performance evaluation reports)</li> <li>• published scientific literature/reviews</li> <li>• clinical experience</li> </ul>   | 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; wellbeing; pre-study protocol review; informed consent; leftover specimen |

#### 7.2.1.3 EP 5.3 Chemical, Physical, and Biological Properties

This EP is not applicable to AI4H.

#### 7.2.1.4 EP 5.4 Sterilization and Microbial Contamination

This EP is not applicable to AI4H.

#### 7.2.1.5 EP 5.5 Considerations of Environment and Conditions of Use

This EP is not applicable to AI4H.

#### 7.2.1.6 EP 5.6 Protection against Electrical, Mechanical, and Thermal Risks

This EP is not applicable to AI4H.

### 7.2.1.7 EP 5.7 Active Medical Devices and IVD Medical Devices and Medical Devices Connected to Them

This EP is not applicable to AI4H.

### 7.2.1.8 EP 5.8 Medical Devices and IVD Medical Devices that Incorporate Software or are Software as a Medical Device

| EP #  | EP text   | 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 lifecycle (e.g., rapid development cycles, frequent changes, the cumulative effect of changes); risk management (e.g., changes to system, environment, and data); information security (e.g., safely implement updates); verification; validation; change management process |
| 5.8.3 | Software that is intended to be used in combination with mobile computing platforms should be designed and developed taking into account the platform itself (e.g., size and contrast ratio of the screen, connectivity, memory, etc.) and the external factors related to their use (varying environment as regards level of light or noise).  | Mobile computing platforms; size; contrast ratio of the screen; connectivity; memory; external factors related to their use (varying environment as regards level of light or noise)   |
| 5.8.4 | Manufacturers should set out minimum requirements concerning hardware, IT networks characteristics and IT security measures, including protection against unauthorized access, necessary to run the software as intended.   | Minimum requirements; hardware; IT networks characteristics; IT security measures; protection against unauthorized access  |

| EP #  | EP text  | EP key concepts                                       |
|-------|--|---|
| 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 |

#### 7.2.1.9 EP 5.9 Medical Devices and IVD Medical Devices with a Diagnostic or Measuring Function

This EP is not applicable to AI4H.

#### 7.2.1.10 EP 5.10 Labeling

| EP #   | EP text  | 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 |

#### 7.2.1.11 EP 5.11 Protection against Radiation

This EP is not applicable to AI4H.

#### 7.2.1.12 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 #   | EP text   | 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  |
|        | 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  |

| EP #   | EP text   | EP key concepts  |
|--------|---|--|
|        | 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                      |
|        | 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                            |
|        | 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                                     |

### 7.2.1.13 EP 5.13 Medical Devices and IVD Medical Devices Incorporating Materials of Biological Origin

This EP is not applicable to AI4H.

### 7.2.2 EP 6.0 Essential Principles Applicable to Medical Devices Other Than IVD Medical Devices

This EP is not applicable to AI4H.

### 7.2.3 EP 7.0 Essential Principles Applicable to IVD Medical Devices

#### 7.2.3.1 EP 7.1 Chemical, Physical and Biological Properties

This EP is not applicable to AI4H.

#### 7.2.3.2 EP 7.2 Performance Characteristics

| EP #  | EP text  | 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                                 |
|       | a) The analytical performance can include, but is not limited to: <ul style="list-style-type: none"> <li>a. Traceability of calibrators and controls</li> <li>b. Accuracy of measurement (trueness and precision)</li> <li>c. Analytical sensitivity/Limit of detection</li> </ul>   | Traceability of calibrators and controls; accuracy of measurements (trueness and precision); analytical sensitivity/limit of detection; |



| EP #  | EP text  | EP key concepts   |
|-------|--|---|
|       | d. Analytical specificity<br>e. Measuring interval/range<br>f. Specimen stability  | analytical specificity; measuring interval/range; specimen stability  |
|       | 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     |
|       | 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 a 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:<br>a) intended user, for example, lay user, laboratory professional;<br>b) intended use environment, for example, patient home, emergency units, ambulances, healthcare centers, laboratory;<br>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 | Performance evaluation; intended use<br>Intended user<br>Intended use environment<br>Relevant population; appropriate representation; ethnicity; gender; genetic diversity; representative population; prevalence rates |

| <b>EP #</b> | <b>EP text</b>   | <b>EP key concepts</b> |
|-------------|--|------------------------|
|             | 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. |                        |

### 7.3 EPs clusters and AI4H concepts

This clause clusters the IMDRF EPs to groups of AI4H concepts; section numbers indicated in parenthesis are those in [IMDRF EPs].

The following are unassigned EP key concepts: State of the art (7.2.1), Software (5.8.1), Software as a medical device (5.8.1), State of the art (5.1.3), State of the art (5.8.2), Valid result (5.12.3).

#### A) Performance

##### a) Analytical performance

*Associated EP key concepts:* {Accuracy (5.8.1), Accuracy (5.12.2), Accuracy of measurements (trueness and precision) (7.2.1), Analytical performance (7.2.1), Analytical sensitivity/Limit of detection (7.2.1), Analytical specificity (7.2.1), Measuring interval/range (7.2.1), Performance (5.1.1), Performance (5.1.2), Performance (5.8.1), Performance characteristics (7.2.1), Quality (5.1.2), Reliability (5.8.1), Specimen stability (7.2.1), Traceability of calibrators and controls (7.2.1), Validation (5.8.2), Validation (7.2.1), Validation (7.2.1), Verification (5.8.2), Verification (5.12.3), Calibrators (7.2.2), Control materials (7.2.2), Precision (5.8.1)}

*Associated AI4H aspects:*

##### 1) Two-class classification metrics

- i. Accuracy
- ii. F1-score
- iii. Sensitivity
- iv. Specificity

##### 2) Multiclass classification metrics

- i. Top-N accuracy

##### 3) Regression metrics

- i. Mean squared error

##### b) Benefit-risk

*Associated EP key concepts:* {Patient benefits (5.1.1), Well-being (5.2.2), Benefit-risk determination (5.2.1)}

*Associated AI4H aspects:*

##### 1) For further study

- i. For further study

##### c) Clinical performance

*Associated EP key concepts:* {Prevalence rates (7.2.4), Relevant population (7.2.4), Representative population (7.2.4), Clinical evaluation (5.2.1), Clinical experience (5.2.1), Clinical investigation report (5.2.1), Clinical performance (7.2.1), Clinical performance (7.2.1), Diagnostic/clinical sensitivity (7.2.1), Diagnostic/clinical specificity (7.2.1), Expected values in normal and affected populations. (7.2.1), Gender (7.2.4), Genetic diversity (7.2.4), Health (5.1.1), Likelihood ratios (7.2.1), Negative predictive value (7.2.1), Positive predictive value (7.2.1), Published scientific literature (5.2.1), Performance (5.1.1), Performance (5.1.2), Performance characteristics (7.2.1), Quality (5.1.2), Patient benefits (5.1.1), Well-being (5.2.2), Ethnicity (7.2.4)}

*Associated AI4H aspects:*

##### 1) For further study

- i. For further study

#### **d) Measurements**

*Associated EP key concepts:* {Calibrators (7.2.2), Control materials (7.2.2), Reference materials of higher order (7.2.2), Reference measurement procedures (7.2.2), Standardized units (7.2.3), Traceability of values (7.2.2), Accuracy of measurements (trueness and precision) (7.2.1)}

*Associated AI4H aspects:*

- 1) see B) Risk and Control b) Data quality

### **B) Risk and control**

#### **a) Control**

*Associated EP key concepts:* {Update control measures (5.1.2), Updating (5.1.2), Validation (5.8.2), Validation (7.2.1), Validation (7.2.1), Verification (5.8.2), Verification (5.12.3), Control procedures (7.2.1)}

*Associated AI4H aspects:*

- 1) **Cross-validation [Bishop]**

- 2) **Statistical tests**

- i.  $t$ -test,  $F$ -test [Hayashi]
- ii. Serial autocorrelation [White] [Breusch]

- 3) **Information criteria**

- i. AIC [31]
- ii. BIC [32]
- iii. Occam-weighted likelihood [33]

- 4) **Robustness validation**

- i. Adversarial vulnerability [Madry et al.] [Singh et al.] [Ilyas et al., 2019]

- 5) **Out of sample testing**

- 6) **Attribution methods ("Explainable AI (XAI)")**

- i. Gradient\*input [Shrikumar et al.]
- ii. Integrated gradients [Sundararajan et al.]
- iii. Layer-wise relevance propagation [Bach et al.]/deep Taylor decomposition [Montavon et al.]
- iv. Perturbation-based attribution [Zeiler et al.]

#### **b) Data quality**

*Associated EP key concepts:* {Ethnicity (7.2.4), Prevalence rates (7.2.4), Relevant population (7.2.4), Representative population (7.2.4), Appropriate representation (7.2.4), Reference materials of higher order (7.2.2), Reference measurement procedures (7.2.2), Standardized units (7.2.3), Traceability of values (7.2.2), Stability (5.1.8)}

*Associated AI4H aspects:*

- 1) **Data diversity [Gebru et al.]**

- 2) **Preprocessing**

- i. Zero centering
- ii. PCA [Duda et al.]
- iii. Whitening [Duda et al.]

- 3) **Normalization**

- i. Standardization

- ii. Min-max scaling

#### 4) Expert labels

#### 5) Data collection procedure

### c) Risk and alarms

*Associated EP key concepts:* {Risks that cannot be eliminated (5.1.3), Identify and analyze hazards (5.1.2), Acceptable risks (5.1.1), Continuous, iterative risk management (5.1.2), Continuous, iterative risk management (5.1.2), Continuous, iterative risk management (5.1.2), Risk (5.1.2), Risk (5.1.9), Risk control (5.1.2), Risk control (5.1.3), Risk control measures (5.1.3), Risk elimination (5.1.2), Risk management (e.g., changes to system, environment, and data) (5.8.2), Risk management plan (5.1.2), Risk management system (5.1.2), Risk of error (5.12.2), Risk reduction (5.1.5), Risk reduction (5.1.5), Risk reduction (5.8.1), Risk reduction (5.12.2), Risk reduction (5.12.2), Side-effects (5.1.9), Single fault conditions (5.8.1), Failure (5.12.3), Foreseeable misuse (5.1.2), Alarms (5.1.3), Warning (5.12.3), Alarms (5.1.3)}

*Associated AI4H aspects:*

#### 1) Uncertainty quantification

- i. Gaussian Process-based [Denker et al.] [Hinton et al.] [Williams]
- ii. Aleatoric uncertainty [Nix et al.]
- iii. Conformal prediction [Shafer et al.]

#### 2) Outlier detection

- i. Generative methods [Meng et al.] [Frost et al.]
- ii. Bayesian uncertainty [Rasmussen et al.] [Gal et al.]

### d) Safety

*Associated EP key concepts:* {Precision (5.8.1), Control procedures (7.2.1), Stress resistance (5.1.6), Near-patient testing (5.12.2), Near-patient testing (5.12.3), Performance (5.12.3), Failure (5.12.3), Foreseeable misuse (5.1.2), Protection against unauthorized access (5.8.4), Protection against unauthorized access (5.8.5), Safe design (5.1.3), Safety (5.1.1), Safety (5.1.2), Safety (5.2.2), Safety (5.8.1), Safety (5.10.1), Safety (5.12.2), Safety principles compliance (5.1.3), Alarms (5.1.3), Warning (5.12.3), Accuracy (5.8.1), Update control measures (5.1.2), Updating (5.1.2), Expected life of device (5.1.6), Perform as intended (5.1.1), Principles of development life cycle (e.g., rapid development cycles, frequent changes, the cumulative effect of changes) (5.8.2), Identify and analyze hazards (5.1.2), Cybersecurity (5.8.5)}

*Associated AI4H aspects:*

#### 1) Robust training

- i. Adversarial training [Madry et al.]
- ii. Generative methods [Samangouei et al.] [Ilyas et al., 2017] [Shen et al.] [Song et al.] [Schott et al.] [Srinivasan et al.]
- iii. Stability training [Zheng et al.]

### C) Usability and documentation

#### a) Documentation

*Associated EP key concepts:* {Handling (5.12.2), Information [Manual] (5.10.1), Instructions (5.12.1), Instructions (5.12.2), Residual risk information for user (5.1.4), Performance (5.10.1), Training (5.12.2), User training (5.1.3), User understanding (7.2.3), Expected life of device (5.1.6), Intended usage environment (5.1.5), Easy to apply (5.12.1), Risks that cannot be eliminated (5.1.3), Cybersecurity (5.8.5)}

*Associated AI4H aspects:*

- 1) **Datasheets for datasets [Gebru et al.]**
  - 2) **Model cards for ML models [Mitchell et al.]**
- b) Ease of understanding**
- Associated EP key concepts:* {Easily understood (5.10.1), Easy to understand (5.12.1)}
- Associated AI4H aspects:*
- 1) **see B) Risk and control a) Control 6) Attribution methods ("Explainable AI (XAI)")**
  - 2) **Counterfactual explanations [Wachter et al.]**
- c) Intended use**

*Associated EP key concepts:* {Appropriate representation (7.2.4), Intended usage environment (5.1.5), Self-testing (5.12.1), Self-testing (5.12.2), Self-testing (5.12.3), Usage variations (user technique, usage environment) (5.12.1), Perform as intended (5.1.1), Intended use (5.1.2), Intended use (5.1.6), Intended use (5.12.1), Intended use (5.12.3), Intended use (7.2.1), Intended use (7.2.4), Intended use environment (7.2.4), Intended conditions of use (5.1.1), Performance (5.8.1), Handling (5.12.2), Information [Manual] (5.10.1), Instructions (5.12.1), Instructions (5.12.2), Residual risk information for user (5.1.4), Performance evaluation (7.2.4), Easily understood (5.10.1)}

*Associated AI4H aspects:*

- 1) **Specification for inputs**
    - i. Data set coverage with respect to training data
  - 2) **see C) Usability and documentation a) Documentation 1) Datasheets for datasets**
  - 3) **see C) Usability and documentation a) Documentation 2) Model cards for ML models**
- d) Intended user**

*Associated EP key concepts:* {Easy to apply (5.12.1), Performance evaluation (7.2.4), Consider user knowledge (5.1.5), Intended user (7.2.4), Lay user (5.12.1), Lay user (5.12.2), Lay users (5.12.3), Performance (5.10.1), Training (5.12.2), User training (5.1.3), User understanding (7.2.3), Easily understood (5.10.1), Easy to understand (5.12.1), Self-testing (5.12.1), Self-testing (5.12.2), Self-testing (5.12.3), Usage variations (user technique, usage environment) (5.12.1), Interpretation of results (5.12.2), Near-patient testing (5.12.2), Near-patient testing (5.12.3), Performance (5.12.3), Near-patient testing (5.12.2)}

*Associated AI4H aspects:*

- 1) **For further study**
    - i. For further study
- e) Interpretability**
- Associated EP key concepts:* {Interpretation of results (5.12.2), Numerical values (7.2.3)}
- Associated AI4H aspects:*
- 1) **see B) Risk and control a) Control 6) Attribution methods ("Explainable AI (XAI)")**

**D) Lifecycle**

- a) Change management**

*Associated EP key concepts:* {Change management process (5.8.2), Stability (5.1.8)}

*Associated AI4H aspects:*

**1) Evolution of the AI algorithm**

- i. For further study

**b) Lifecycle**

*Associated EP key concepts:* {Stability (5.1.8), Principles of development life cycle (e.g., rapid development cycles, frequent changes, the cumulative effect of changes) (5.8.2), MD life cycle (5.1.2), Shelf life (5.1.8)}

*Associated AI4H aspects:*

**1) AI software lifecycle**

- i. AI lifecycle and failure modes [Russell et al.]

**E) Dependencies**

**a) External factors**

*Associated EP key concepts:* {External factors related to their use (varying environment as regards level of light or noise) (5.8.3), Stress resistance (5.1.6)}

*Associated AI4H aspects:*

**1) For further study**

- i. For further study

**b) Technical interfaces**

*Associated EP key concepts:* {Cybersecurity (5.8.5), Connectivity (5.8.3), Contrast ratio of the screen (5.8.3), Hardware (5.8.4), Information security (e.g., safely implement updates) (5.8.2), IT networks characteristics (5.8.4), IT security measures (5.8.4), Memory (5.8.3), Minimum requirements (5.8.4), Mobile computing platforms (5.8.3), Size (5.8.3), Reliability (5.8.1), Protection against unauthorized access (5.8.4), Protection against unauthorized access (5.8.5)}

*Associated AI4H aspects:*

**1) Compression of AI4H models**

**2) Response time**

**3) Memory**

- i. RAM
- ii. Storage

**4) Compute**

- i. GPU
- ii. CPU

**5) Networking**

- i. Cloud
- ii. Hardware resources sharing

**6) Operating system**

**7) Displays**

**8) Sensors for input data**

**F) Ethical compliance**

**a) Ethical compliance**

*Associated EP key concepts:* {Declaration of Helsinki (5.2.2), Ethical principles (5.2.2), Informed consent (5.2.2), Leftover specimen (5.2.2), Pre-study protocol review (5.2.2), Rights (5.2.2)}

*Associated AI4H aspects:*

**1) Factory Acceptance Testing (FAT) optimization objectives ("FAT training")**

- i. Fairness policies [Shaikh et al.]
- ii. Value translation [Dobbe et al.]
- iii. Fairness-aware programming [Albarghouthi et al.]
- iv. Fairness in ML systems [Holstein et al.]
- v. Metric-fair learning [Yona et al.]
- vi. Fairness enhancing methods for ML [Friedler et al.]

**2) FAT validation**

- i. Fairness through awareness [Dwork et al.]
- ii. Privacy in machine learning [Papernot et al.]
- iii. Formal fairness [Gajane et al.]
- iv. Fairness and race [Moss]
- v. AI fairness evaluation toolkits [Bellamy et al.]

**3) Data acceptance and handling**

- i. Privacy

**4) Patient consent**



## 7.4 Mapping IMDRF EPs to AI4H concepts

| EP #  | EP key concepts   | AI4H concepts  |
|-------|---|--|
| 5.1.1 | Performance; Intended conditions of use; Safety; Perform as intended; Acceptable risks; Patient benefits; Health  | Two-class classification metrics (A.a.1), Multiclass classification metrics (A.a.2), Regression metrics (A.a.3), Specification for inputs (C.c.1), see C.a.1 (C.c.2), see C.a.2 (C.c.3), Robust training (B.d.1), Uncertainty quantification (B.c.1), Outlier detection (B.c.2)  |
| 5.1.2 | Risk management system; Quality; Safety; Performance; Continuous, iterative risk management; MD life cycle; Risk management plan; Identify and analyse hazards; Risk; Intended use; Foreseeable misuse; Risk elimination; Risk control; Continuous, iterative risk management; Continuous, iterative risk management; Update control measures | Uncertainty quantification (B.c.1), Outlier detection (B.c.2), Two-class classification metrics (A.a.1), Multiclass classification metrics (A.a.2), Regression metrics (A.a.3), Robust training (B.d.1), AI software lifecycle (D.b.1), Cross-validation (B.a.1), Statistical tests (B.a.2), Information criteria (B.a.3), Robustness validation (B.a.4), Out of sample testing (B.a.5), Attribution methods (B.a.6), Specification for inputs (C.c.1), see C.a.1 (C.c.2), see C.a.2 (C.c.3) |
| 5.1.3 | Risk control measures; Safety principles compliance; State of the art; Risk control; Safe design; Alarms; Risks that cannot be eliminated; Alarms; User training  | Uncertainty quantification (B.c.1), Outlier detection (B.c.2), Robust training (B.d.1), Datasheets for data sets (C.a.1), Model cards for ML models (C.a.2), Specification for inputs (C.c.1), see C.a.1 (C.c.2), see C.a.2 (C.c.3)  |
| 5.1.4 | Residual risk information for user  | Uncertainty quantification (B.c.1), Outlier detection (B.c.2), Robust training (B.d.1), Datasheets for data sets (C.a.1), Model cards for ML models (C.a.2), Specification for inputs (C.c.1), see C.a.1 (C.c.2), see C.a.2 (C.c.3)  |
| 5.1.5 | Risk reduction; Risk reduction; Intended usage environment; Consider user knowledge   | Uncertainty quantification (B.c.1), Outlier detection (B.c.2), Robust training (B.d.1), Datasheets for data sets (C.a.1), Model cards for ML models (C.a.2), Specification for inputs (C.c.1), see C.a.1 (C.c.2), see C.a.2 (C.c.3), Uncertainty quantification (B.c.1), Outlier detection (B.c.2)   |
| 5.1.6 | Stress resistance; Intended use; Expected life of device  | Specification for inputs (C.c.1), see C.a.1 (C.c.2), see C.a.2 (C.c.3), Robust training (B.d.1), Datasheets for data sets (C.a.1), Model cards for ML models (C.a.2)   |
| 5.1.8 | Stability; Shelf life   | AI software lifecycle (D.b.1), Evolution of the AI algorithm (D.a.1), Data diversity (B.b.1), Preprocessing (B.b.2), Normalization (B.b.3), Expert labels (B.b.4), Data collection procedure (B.b.5)   |
| 5.1.9 | Risk; Side-effects  | Uncertainty quantification (B.c.1), Outlier detection (B.c.2)  |

| EP #  | EP key concepts   | AI4H concepts   |
|-------|---|---|
| 5.2.1 | Clinical evaluation; Benefit-risk determination; Clinical investigation report; Published scientific literature; Clinical experience  |   |
| 5.2.2 | Ethical principles; Declaration of Helsinki Rights; Safety; Well-being; Pre-study protocol review; Informed consent; Leftover specimen  | FAT optimization objectives ("FAT training") (F.a.1), FAT validation (F.a.2), Data acceptance and handling (F.a.3), Patient consent (F.a.4), Robust training (B.d.1)  |
| 5.8.1 | Electronic programmable systems; Software; Software as a medical device; Accuracy; Reliability; Precision; Safety; Performance; Single fault conditions; Risk reduction   | Two-class classification metrics (A.a.1), Multiclass classification metrics (A.a.2), Regression metrics (A.a.3), Robust training (B.d.1), Compression of AI4H models (E.b.1), Response time (E.b.2), Memory (E.b.3), Compute (E.b.4), Networking (E.b.5), Operating system (E.b.6), Displays (E.b.7), Sensors for input data (E.b.8), Uncertainty quantification (B.c.1), Outlier detection (B.c.2)   |
| 5.8.2 | 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 | AI software lifecycle (D.b.1), Robust training (B.d.1), Uncertainty quantification (B.c.1), Outlier detection (B.c.2), Compression of AI4H models (E.b.1), Response time (E.b.2), Memory (E.b.3), Compute (E.b.4), Networking (E.b.5), Operating system (E.b.6), Displays (E.b.7), Sensors for input data (E.b.8), Two-class classification metrics (A.a.1), Multiclass classification metrics (A.a.2), Regression metrics (A.a.3), Cross-validation (B.a.1), Statistical tests (B.a.2), Information criteria (B.a.3), Robustness validation (B.a.4), Out of sample testing (B.a.5), Attribution methods (B.a.6), Evolution of the AI algorithm (D.a.1) |
| 5.8.3 | 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)  | Compression of AI4H models (E.b.1), Response time (E.b.2), Memory (E.b.3), Compute (E.b.4), Networking (E.b.5), Operating system (E.b.6), Displays (E.b.7), Sensors for input data (E.b.8)  |
| 5.8.4 | Minimum requirements; Hardware; IT networks characteristics; IT security measures; Protection against unauthorized access   | Compression of AI4H models (E.b.1), Response time (E.b.2), Memory (E.b.3), Compute (E.b.4), Networking (E.b.5), Operating system (E.b.6), Displays (E.b.7), Sensors for input data (E.b.8), Robust training (B.d.1)   |
| 5.8.5 | Cybersecurity; Protection against unauthorized access   | Compression of AI4H models (E.b.1), Response time (E.b.2), Memory (E.b.3), Compute (E.b.4), Networking (E.b.5), Operating system (E.b.6), Displays (E.b.7), Sensors for input data (E.b.8), Robust training (B.d.1), Datasheets for data sets (C.a.1), Model cards for ML models (C.a.2)  |

| <b>EP #</b> | <b>EP key concepts</b>  | <b>AI4H concepts</b>  |
|-------------|---|---|
| 5.10.1      | Information [Manual]; Safety; Performance; Easily understood  | Datasheets for data sets (C.a.1), Model cards for ML models (C.a.2), Specification for inputs (C.c.1), see C.a.1 (C.c.2), see C.a.2 (C.c.3), Robust training (B.d.1), see B.a.6 (C.b.1), Counterfactual explanations (C.b.2)  |
| 5.12.1      | Lay user; Self-testing; Intended use; Usage variations (user technique, usage environment); Instructions; Easy to understand; Easy to apply   | Specification for inputs (C.c.1), see C.a.1 (C.c.2), see C.a.2 (C.c.3), Datasheets for data sets (C.a.1), Model cards for ML models (C.a.2), see B.a.6 (C.b.1), Counterfactual explanations (C.b.2)   |
| 5.12.2      | Lay user; Self-testing; Near-patient testing; Safety; Accuracy; Instructions; Risk reduction; Training; Risk reduction; Risk of error; Handling; Interpretation of results  | Specification for inputs (C.c.1), see C.a.1 (C.c.2), see C.a.2 (C.c.3), Robust training (B.d.1), Two-class classification metrics (A.a.1), Multiclass classification metrics (A.a.2), Regression metrics (A.a.3), Uncertainty quantification (B.c.1), Outlier detection (B.c.2), see B.a.6 (C.e.1)  |
| 5.12.3      | Warning; Failure; Valid result  | Specification for inputs (C.c.1), see C.a.1 (C.c.2), see C.a.2 (C.c.3), Robust training (B.d.1), Two-class classification metrics (A.a.1), Multiclass classification metrics (A.a.2), Regression metrics (A.a.3), Cross-validation (B.a.1), Statistical tests (B.a.2), Information criteria (B.a.3), Robustness validation (B.a.4), Out of sample testing (B.a.5), Attribution methods (B.a.6), Uncertainty quantification (B.c.1), Outlier detection (B.c.2) |
| 7.2.1       | Performance characteristics; Analytical performance; Clinical performance; Validation; State of the art; Traceability of calibrators and controls; Accuracy of measurements (trueness and precision); Analytical sensitivity/Limit of detection; Analytical specificity; Measuring interval/range; Specimen stability; Clinical performance; Diagnostic/clinical sensitivity; Diagnostic/clinical specificity; Positive predictive value; Negative predictive value; Likelihood ratios; Expected values in normal and affected populations.; Validation; Control procedures; Intended use | Two-class classification metrics (A.a.1), Multiclass classification metrics (A.a.2), Regression metrics (A.a.3), See B.b (A.d.1), Cross-validation (B.a.1), Statistical tests (B.a.2), Information criteria (B.a.3), Robustness validation (B.a.4), Out of sample testing (B.a.5), Attribution methods (B.a.6), Robust training (B.d.1), Specification for inputs (C.c.1), see C.a.1 (C.c.2), see C.a.2 (C.c.3)   |
| 7.2.2       | Calibrators; Control materials; Traceability of values; Reference measurement procedures; Reference materials of higher order   | See B.b (A.d.1), Two-class classification metrics (A.a.1), Multiclass classification metrics (A.a.2), Regression metrics (A.a.3), Data diversity (B.b.1), Preprocessing (B.b.2), Normalization (B.b.3), Expert labels (B.b.4), Data collection procedure (B.b.5)  |

| EP #  | EP key concepts   | AI4H concepts  |
|-------|---|--|
| 7.2.3 | Numerical values; Standardized units; User understanding  | see B.a.6 (C.e.1), See B.b (A.d.1), Data diversity (B.b.1), Preprocessing (B.b.2), Normalization (B.b.3), Expert labels (B.b.4), Data collection procedure (B.b.5), Specification for inputs (C.c.1), see C.a.1 (C.c.2), see C.a.2 (C.c.3) |
| 7.2.4 | Performance evaluation; Intended use; Intended user; Intended use environment; Relevant population; Appropriate representation; Ethnicity; Gender; Genetic diversity; Representative population; Prevalence rates | Specification for inputs (C.c.1), see C.a.1 (C.c.2), see C.a.2 (C.c.3), Data diversity (B.b.1), Preprocessing (B.b.2), Normalization (B.b.3), Expert labels (B.b.4), Data collection procedure (B.b.5)                                     |

---