

International Telecommunication Union

ITU-T FG-AI4H Deliverable

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

27 September 2023

PRE-PUBLISHED VERSION

DELO

Overview of the FG-AI4H deliverables

ITU-T

Summary

This document provides the overview of the planned deliverables for the ITU-T Focus Group on AI for health (FG-AI4H) to provide a standardization framework on artificial intelligence for health with a combination of horizontal and vertical perspective. The horizontal perspective reviews deliverables DEL1 to DEL9 including the generalized considerations on ethics, regulatory, requirement, data processing, model training, model evaluation, adoption and scale-up, etc. The vertical perspective summarizes use cases in specific domain with corresponding artificial intelligence/ machine learning tasks, for example 24 topic description documents (DEL 10.1-DEL 10.24 as of 2023-09-27) of topic groups in the FG-AI4H.

Keywords

Artificial intelligence, deliverables, health, overview, structure

Change Log

This document contains Version 1 of the Deliverable DEL0 on "*Overview of the FG-AI4H deliverables*" approved on 27 September 2023 via the online approval process for the ITU-T Focus Group on AI for Health (FG-AI4H).

Editor: Shan Xu
China Academy of Information and
Communication Technology(CAICT)
China

Email: xushan@caict.ac.cn

CONTENTS

	Page
1 Scope.....	1
2 References.....	1
3 Terms and definitions	3
4 Abbreviations.....	3
5 Introduction.....	3
6 Deliverables type	4
7 Deliverable structure	4
8 Deliverables status	5
9 Summary of the generalized specifications	8
10 Topic groups summary	12

List of Tables

	Page
Table 1 – Updated list of deliverables (same as S-200).....	5
Table 2 – Summary of generalized documents (DEL 1-9)	9
Table 3 – Summary of Topic Groups (DEL 10.1-10.24)	12

List of Figures

	Page
Figure 1 – FG-A4H Deliverables structure	4

ITU-T FG-AI4H Deliverable DEL0

Overview of the FG-AI4H deliverables

1 Scope

This deliverable provides an overview of the planned deliverables for the ITU-T Focus Group on AI for health (FG-AI4H) to provide a standardization framework on artificial intelligence for health. The reviewed deliverables including 9 deliverables (DEL 1-9) focus on the generalized considerations on ethics, regulatory, requirement, data processing, model training, model evaluation, adoption and scale-up, etc., and 23 topic description documents (DEL 10.1-10.23 as of 2022-06-02) within specific health domains with corresponding AI/ ML benchmarking tasks.

2 References

[ISO/IEC 2382]	ISO/IEC 2382:2015, Information technology — Vocabulary
[IEC 62304]	IEC 62304:2006 + A1:2015, "Medical device software – Software life cycle processes"
[DEL0.1]	Common unified terms in artificial intelligence for health
[DEL1]	Ethics and governance of artificial intelligence for health
[DEL2]	Overview of Regulatory Considerations on Artificial Intelligence for Health
[DEL2.1]	Mapping of IMDRF essential principles to AI for health software
[DEL2.2]	Good practices for health applications of machine learning: Considerations for manufacturers and regulators
[DEL03]	AI4H requirement specifications
[DEL4]	AI software life cycle specification
[DEL5]	Data specification
[DEL5.1]	Data requirements
[DEL5.2]	Data acquisition
[DEL5.3]	Data annotation specification
[DEL5.4]	Training and test data specification
[DEL5.5]	Data handling
[DEL5.6]	Data sharing practices
[DEL6]	AI training best practices specification
[DEL7]	AI for health evaluation considerations
[DEL7.1]	AI4H evaluation process description
[DEL7.2]	AI technical test specification
[DEL7.3]	ML4H trial audits–Iteration 2.0 Playbook (Version 3.0)
[DEL7.4]	Clinical evaluation of AI for health
[DEL7.5]	Assessment platform
[DEL9]	AI4H applications and platforms

[DEL9.1]	Mobile Applications
[DEL9.2]	Cloud-based AI applications
[DEL10]	AI4H use cases: Topic Description Documents
[DEL10.01]	FG-AI4H Deliverable DEL10.01, <i>Cardiovascular disease management (TG-Cardio)</i> .
[DEL10.02]	FG-AI4H Deliverable DEL10.02, <i>Dermatology (TG-Derma)</i> .
[DEL10.03]	FG-AI4H Deliverable DEL10.03, <i>Diagnosis of bacterial infection and anti-microbial resistance (TG-AMR)</i> .
[DEL10.04]	FG-AI4H Deliverable DEL10.04, <i>Falls among the elderly (TG-Falls)</i> .
[DEL10.05]	FG-AI4H Deliverable DEL10.05, <i>Histopathology (TG-Histo)</i> .
[DEL10.06]	FG-AI4H Deliverable DEL10.06, <i>Malaria detection (TG-Malaria)</i> .
[DEL10.07]	FG-AI4H Deliverable DEL10.07, <i>Maternal and child health (TG-MCH)</i> .
[DEL10.08]	FG-AI4H Deliverable DEL10.08, <i>Neurological disorders (TG-Neuro)</i> .
[DEL10.09]	FG-AI4H Deliverable DEL10.09, <i>Ophthalmology (TG-Ophthalmology)</i> .
[DEL10.10]	FG-AI4H Deliverable DEL10.10, <i>Outbreak detection (TG-Outbreaks)</i> .
[DEL10.11]	FG-AI4H Deliverable DEL10.11, <i>Psychiatry (TG-Psy)</i> .
[DEL10.12]	FG-AI4H Deliverable DEL10.12, <i>Radiology (TG-Radiology)</i> .
[DEL10.13]	FG-AI4H DEL10.13, <i>Snakebite and snake identification (TG-Snake)</i> .
[DEL10.14]	FG-AI4H Deliverable DEL10.14, <i>Symptom assessment (TG-Symptom)</i> .
[DEL10.15]	FG-AI4H Deliverable DEL10.15, <i>Tuberculosis (TG-Tuberculosis)</i> .
[DEL10.16]	FG-AI4H Deliverable DEL10.16, <i>Volumetric chest computed tomography (TG-Diagnostic CT)</i> .
[DEL10.17]	FG-AI4H Deliverable DEL10.17, <i>Dental diagnostics and digital dentistry (TG-Dental)</i> .
[DEL10.18]	FG-AI4H Deliverable DEL10.18, <i>AI-based detection of falsified medicine (TG-FakeMed)</i> .
[DEL10.19]	FG-AI4H Deliverable DEL10.19, <i>Primary and secondary Diabetes risk prediction (TG-Diabetes)</i> .
[DEL10.20]	FG-AI4H Deliverable DEL10.20, <i>AI for endoscopy (TG-Endoscopy)</i> .
[DEL10.21]	FG-AI4H Deliverable DEL10.21, <i>AI for Musculoskeletal medicine (TG-MSK)</i> .
[DEL10.22]	FG-AI4H Deliverable DEL10.22, <i>AI for human reproduction and fertility (TG-Fertility)</i> .
[DEL10.23]	FG-AI4H Deliverable DEL10.23, <i>AI in sanitation for public health (TG-Sanitation)</i> .
[DEL10.24]	FG-AI4H Deliverable DEL10.24, <i>AI for point-of-care diagnostics (TG-POC)</i> .

NOTE: Literature references are listed in the bibliography.

3 Terms and definitions

This document uses the following terms defined elsewhere:

3.1 Algorithm [IMDRF/SaMD-N41]: a finite set of instructions (or rules) that defines a sequence of operations for solving a particular computational problem for all problem instances for a problem set.

3.2 Application [ITU-T H.764]: A functional implementation realized as software running in one or spread over several interplaying hardware entities.

3.3 Artificial intelligence [ISO/IEC 2382]: Branch of computer science devoted to developing data processing systems that perform functions normally associated with human intelligence, such as reasoning, learning, and self-improvement.

3.4 Bias [WHO AI-EG]: A threat to inclusiveness and equity that represents a departure, often arbitrary, from equal treatment.

3.5 Machine learning [ISO/IEC 2382]: Automatic learning, process by which a functional unit improves its performance by acquiring new knowledge or skills, or by reorganizing existing knowledge or skills.

4 Abbreviations

AHG	Ad hoc group
AI	Artificial Intelligence
DEL	Deliverable
FG	Focus Group
FG-AI4H	Focus Group on artificial intelligence for health
ITU	International Telecommunication Union
ML	Machine Learning
TDD	Topic description document
TG	Topic groups
WG	Working group
WHO	World Health Organization

5 Introduction

The ITU/WHO Focus Group on artificial intelligence for health (FG-AI4H) was established by ITU-T Study Group 16 at its meeting in Ljubljana, Slovenia, 9-20 July 2018 and continued into operation until September 2023. This group was committed to establish a standardized assessment framework for the evaluation of AI-based methods for health, diagnosis, triage or treatment decisions. The deliverables planned for the FG-AI4H included nine deliverables (DEL 1-9) that focused on the generalized considerations on ethics, regulatory, requirement, data processing, model training, model evaluation, adoption and scale-up, etc., and 24 topic groups (DEL 10.1-10.24 as of 2023-07-05) within specific health domains with corresponding artificial intelligence (AI)/machine learning (ML) benchmarking tasks. This document was developed to provide a compiled overview of the FG-AI4H deliverables. It was also meant to be used as a quick guide for to understand the FG-AI4H activities.

6 Deliverables type

According to utilization scenario, the planned deliverables can be divided into two types:

- Generalized specifications (DEL 1-9): focus on the generalized specifications including ethics, regulatory, requirement, data, training, evaluation, application, etc. Each part is interconnected to form a life cycle process of AI-based methods for health.
- Topic description documents (DEL 10.1-10.24): focus on the use cases in specific health domains with corresponding AI/ML tasks. Each case can be regarded as an example of a whole process recommended by generalized specifications (DEL 1-9) and profiled in a specific application scenario.

7 Deliverable structure

Figure 1 gives an overall structure of all deliverables from DEL 1-9 and DEL 10.1-10.24. The arrows between DEL 1-9 indicate sequential considerations from the perspective of AI implementation, which includes demand finding, solution (data preparation, model development and evaluation aspects), and finally scale-up and adoption to population. Topic groups (TGs), shown at the bottom, consider the specific health domains with corresponding AI/ML tasks in the recommended process of the above generalized specifications. They provide the practical verification on specific health scenarios involved with a number of AI for health tasks and data modalities.

Four pillars can be used to categorize the various deliverables, as indicated by colours in Figure 1:

- Ethics (DEL1, blue)
- Regulatory (DEL2, orange)
- Clinical assessment and use cases (DEL7.4, DEL10 and DEL10.1 to DEL10.24, green)
- Technical (all other deliverables, white)

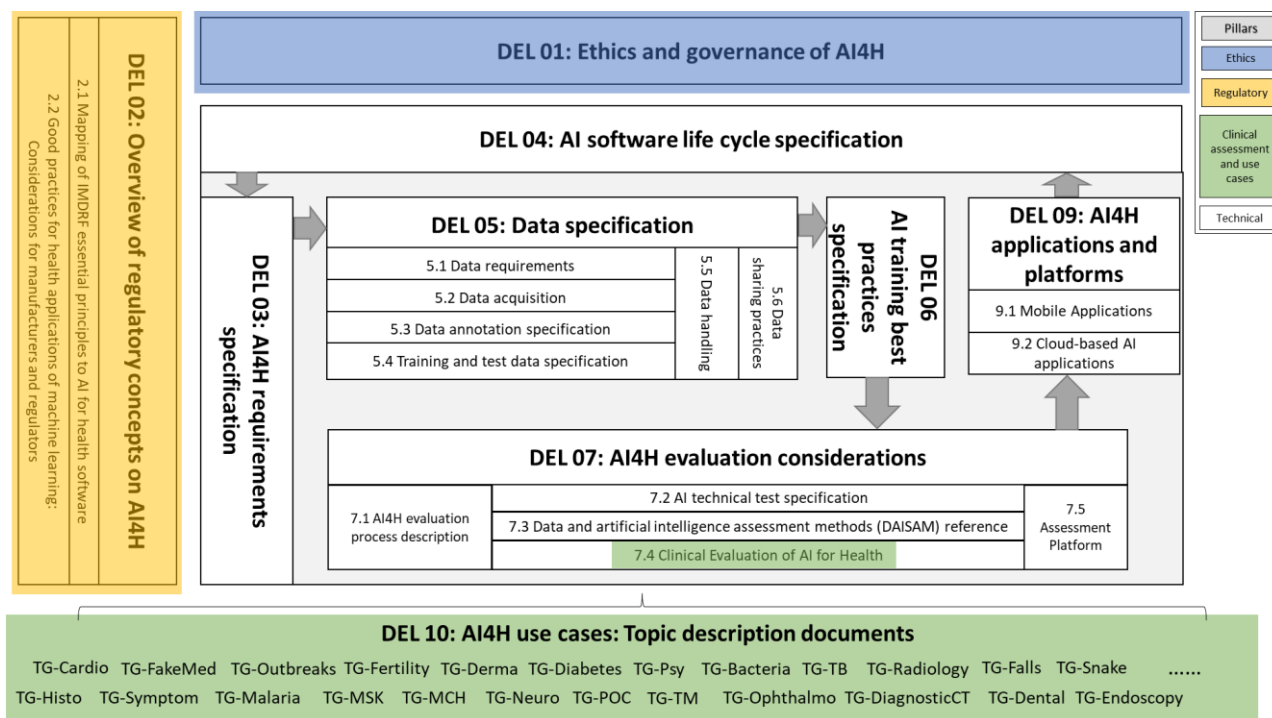


Figure 1 – FG-A4H Deliverables structure

8 Deliverables status

Table 1 shows the status of FG-AI4H deliverables at its life end, with the updates from FG-AI4H meetings, management input and feedback from editors. The status column indicates the status of each deliverable as of 2023-09-27:

- X: Archived
- P: Published (in the ITU website)
- A: Approved, under publication

Table 1 – Updated list of deliverables (same as S-200)

No.	Deliverable	Updated initial draft editor	Availability*	Status
0	Overview of the FG-AI4H deliverables	Shan Xu (CAICT, China)	S-046	A
0.1	<i>Common unified terms in artificial intelligence for health</i>	Markus Wenzel (Fraunhofer HHI, Germany)	P-201 (agreed at P)	P
1	<i>AI4H ethics considerations</i>	Andreas Reis (WHO)	O-201 (agreed at O)	P
2	<i>Overview of regulatory considerations on artificial intelligence for health</i>	Shada Alsalamah (WHO)	P-202 (agreed at P)	P
2.1	<i>Mapping of IMDRF essential principles to AI for health software</i>	Luis Oala (Dotphoton, Switzerland), Pradeep Balachandran (Technical Consultant eHealth, India), Pat Baird (Philips, USA), Thomas Wiegand (Fraunhofer HHI, Germany)	R-047 (agreed online, 2023-03-16)	A
2.2	<i>Good practices for health applications of machine learning: Considerations for manufacturers and regulators</i>	Pradeep Balachandran (India) and Christian Johner (Johner Institut, Germany)	P-203 (agreed at P)	P
3	<i>AI4H requirement specifications</i>	Pradeep Balachandran (India)	R-049 (agreed online, 2023-03-16)	A
4	<i>AI software life cycle specification</i>	Pat Baird (Philips, USA)	R-044 (Agreed at R)	A
5	<i>Data specification</i>	Marc Lecoultre (MLlab.AI, Switzerland)	G-205	X
5.1	<i>Data requirements</i>	[Marc Lecoultre (MLlab.AI, Switzerland)]**	R-066 (agreed online, 2023-03-16)	A
5.2	<i>Data acquisition</i>	Rajaraman (Giri) Subramanian (Calligo Tech, India), Vishnu Ram (India)	G-205-A02	X
5.3	<i>Data annotation specification</i>	Shan Xu (CAICT, China), Harpreet Singh (ICMR, India), Sebastian Bosse (Fraunhofer HHI, Germany)	R-067 (agreed online, 2023-03-16)	A
5.4	<i>Training and test data specification</i>	Luis Oala (Dotphoton, Switzerland), Pradeep Balachandran (India)	R-066 (agreed online, 2023-03-16)	A
5.5	<i>Data handling</i>	Marc Lecoultre (MLlab.AI, Switzerland)	R-048 (agreed online, 2023-03-16)	A

No.	Deliverable	Updated initial draft editor	Availability*	Status
5.6	Data sharing practices	Ferath Kherif (CHUV, Switzerland), Banusri Velpandian (ICMR, India), WHO Data Team	L-044	X
6	AI training best practices specification	Xin Ming Sim and Stefan Winkler (AI Singapore)	R-051 (agreed online, 2023-03-16)	A
7	AI for health evaluation considerations	Markus Wenzel (Fraunhofer HHI, Germany)	R-042-RI (agreed online, 2023-03-16)	A
7.1	AI4H evaluation process description	Yu (Ursula) Zhao (WHO)	G-207-A01	X
7.2	AI technical test specification	Auss Abbood (Robert Koch Institute, Germany)	R-069 (agreed online, 2023-03-16)	A
7.3	Data and artificial intelligence assessment methods (DAISAM) reference	Luis Oala (Dotphoton, Switzerland)	P-032 (L-052)	X
7.4	Clinical evaluation of AI for health	Naomi Lee (NICE, UK), Eva Weicken (Fraunhofer HHI, Germany), Shubhanan Upadhyay (ADA Health, Germany)	R-202 (agreed at R), Checklist: R-063	A
7.5	Assessment platform	Luis Oala (Dotphoton, Switzerland), Marc Lecoultre and Steffen Vogler (Bayer AG, Germany)	I-037	X
8	AI4H scale-up and adoption	Sameer Pujari (WHO), Yu Zhao and Javier Elkin [Previously: Robyn Whittaker (New Zealand)]	– (O-056)	X
9	AI4H applications and platforms	Manjeet Chalga (ICMR, India)	L-050 (P-055)	X
9.1	Mobile applications	Khondaker Mamun (UIU, Bangladesh), Manjeet Chalga (ICMR, India)	Q-047	X
9.2	Cloud-based AI applications	Khondaker Mamun (UIU, Bangladesh)	I-049	X
10	AI4H use cases: Topic description documents	Eva Weicken (Fraunhofer HHI, Germany)	R-050 (agreed online, 2023-03-16)	A
10.1	Cardiovascular disease management (TG-Cardio)	Benjamin Muthambi (Watif Health, South Africa)	S-006-A01	X
10.2	Dermatology (TG-Derma)	Harsha Jayakody (Flash Health, Sri Lanka), Ivy Lee (American Academy of Dermatology, USA)	S-007-A01	A
10.3	Diagnosis of bacterial infection and anti-microbial resistance (TG-Bacteria)	Nada Malou (MSF, France)	S-008-A01	X
10.4	Falls among the elderly (TG-Falls)	Pierpaolo Palumbo (University of Bologna, Italy)	S-012-A01	A
10.5	Histopathology (TG-Histo)	Frederick Klauschen (LMU Munich & Charité Berlin, Germany)	S-013-A01	X
10.6	Malaria detection (TG-Malaria)	Rose Nakasi (Makerere University, Uganda)	S-014-A01	A

No.	Deliverable	Updated initial draft editor	Availability*	Status
10.7	Maternal and child health (TG-MCH)	Alexandre Chiavegatto Filho (University of São Paulo, Brazil)	S-015-A01	A
10.8	Neurological disorders (TG-Neuro)	Marc Lecoultré (MLlab.AI, Switzerland) and Ferath Kherif (CHUV, Switzerland)	S-016-A01	A
10.9	Ophthalmology (TG-Ophthalmo)	Arun Shroff (MedIndia)	S-017-A01	A
10.10	Outbreak detection (TG-Outbreaks)	Auss Abbood and Alexander Ullrich (Robert Koch Institute, Germany); Khahlil Louisy and Alexander Radunsky (Institute for Technology & Global Health, ITGH, US)	S-018-A01	A
10.11	Psychiatry (TG-Psy)	Nicolas Langer (ETH Zurich, Switzerland)	S-019-A01	X
10.12	AI for radiology (TG-Radiology)	Darlington Ahiale Akogo (minoHealth AI Labs, Ghana)	S-023-A01	A
10.13	Snakebite and snake identification (TG-Snake)	Rafael Ruiz de Castaneda (UniGE, Switzerland)	S-020-A01	X
10.14	Symptom assessment (TG-Symptom)	Henry Hoffmann (Ada Health, Germany) and Martin Cansdale (Healthily, UK)	S-021-A01	A
10.15	Tuberculosis (TG-TB)	Manjula Singh (ICMR, India)	S-022-A01	A
10.16	Volumetric chest CT (TG-DiagnosticCT)	Kuan Chen (Infervision, China)	S-009-A01	X
10.17	Dental diagnostics and digital dentistry (TG-Dental)	Falk Schwendicke and Joachim Krois (Charité Berlin, Germany); Tarry Singh (deepkapha.ai, Netherlands)	S-010-A01	A
10.18	Falsified Medicine (TG-FakeMed)	Franck Verzeffé (TrueSpec-Africa, DRC)	S-011-A01	X
10.19	Primary and secondary diabetes prediction (TG-Diabetes)	Andrés Valdivieso (Anastasia.ai, Chile)	S-024-A01	X
10.20	AI for endoscopy (TG-Endoscopy)	Jianrong Wu (Tencent Healthcare, China)	S-025-A01	A
10.21	AI for musculoskeletal medicine (TG-MSK)	Peter Grinbergs (EQL, UK), Mark Elliott (Warwick University, UK)	S-026-A01	A
10.22	AI for human reproduction and fertility (TG-Fertility)	Susanna Brandi , Eleonora Lippolis , (Merck KGaA, Darmstadt, Germany)	S-027-A01	X
10.23	AI for traditional medicine (TG-TM)	Saketh Ram Thrigulla (Ministry of Ayush, India)	S-028-A01	A
10.24	AI for point-of care diagnostics (TG-POC)	Nina Linder (University of Helsinki, Finland)	S-029-A01	A

NOTES

* Approved deliverables are indicated in *italics*. The document numbers indicated reflect the status as of the end of the Meeting R. Some links provided are to slide sets; these slide sets are not meant to be the deliverable documents, but rather a status update concerning progress of the respective deliverable. Documents in parenthesis are status updates, not a deliverable text.

** Acting editor

Two deliverables were initially considered but did not proceed:

No.	Deliverable	Updated initial draft editor	Reference	Status
–	Open Code Initiative reference software implementation	Marc Lecoultré (MLlab.AI, Switzerland)	K-043	X

No.	Deliverable	Updated initial draft editor	Reference	Status
–	Risk management in AI for health	Pat Baird (Philips, USA)	K-034	X

Other documents already approved were:

No.	Deliverable	Editor(s)	Reference	Status
AHG-DT4HE Output 1	Guidance on digital technologies for COVID health emergency	Shan Xu (CAICT, China), Ana Riviere-Cinnamond (PAHO)	K-042 (<i>agreed K</i>)	A
TG-Dental Output 1	Artificial intelligence in dental research: A checklist for authors and reviewers	Falk Schwendicke , Joachim Krois (Charité Berlin, Germany)	M-004 (<i>agreed M</i>)	A
TG-Dental Output 2	Artificial intelligence for oral and dental healthcare: Core education curriculum	Falk Schwendicke , Joachim Krois (Charité Berlin, Germany)	R-033 (<i>agreed R</i>)	A
TG-Dental Output 3	Ethical considerations on artificial intelligence in dentistry: A framework and checklist	Falk Schwendicke , Joachim Krois (Charité Berlin, Germany)	S-037	A

9 Summary of the generalized specifications

A summary of the generalized specifications is compiled here and provide a general standardization framework on ethics, regulatory, requirement, data processing, model training, model evaluation, adoption and scale-up, etc. on AI for health. These key messages were useful to help each DEL editor avoid possible conflicts or overlap in document scope, therefore facilitating collaboration and adaption between different use cases.

Table 2 – Summary of generalized documents (DEL 1-9)

Deliverable		Scope	Last update
1- AI4H ethics considerations		The rapidly developing field of AI raises a number of ethical, legal and social concerns, e.g. regarding equitable access, privacy, appropriate uses and users, liability and bias and inclusiveness. These issues are trans-national in nature, as capturing, sharing and using data generated and/or used by these technologies goes beyond national boundaries. Many questions remain unanswered concerning the ethical development and use of these technologies, including how low- and middle-income countries will benefit from AI developments. This document is to develop a harmonised ethics guidance for the design and implementation of AI in global health.	2022/06/02
2- Overview of regulatory considerations on artificial intelligence for health		This document is aimed as a general, high-level, and nonexclusive overview of key regulatory considerations' topic areas delivered by the WG-RC on AI for health. It highlights some of the key regulatory principles and concepts, such as risk/benefit assessments and considerations for the evaluation and monitoring of the performance of AI solutions.	2022/09/22
2.1	Mapping of IMDRF essential principles to AI for health software	This document 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 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.	2023/03/06
2.2	Good practices for health applications of machine learning: Considerations for manufacturers and regulators	This document recommends a set of good machine learning practice guidelines to the manufacturers and regulators of data driven Artificial Intelligence 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.	2022/09/22
3- AI4H requirements specification		This document is to define the System Requirement Specifications (SyRS) that explains the informational, functional, behavioural and operational aspects a generic AI for health (AI4H) system. SyRS serves as the basis and helps to create system design, system verification and validation plans and procedures. System requirements analysis methodology follows a collaborative team-oriented approach, involving all the working groups and topic groups of AI4GH FG, to help the project team identify, control and track various requirements and changes to those requirements during the AI4H system development lifecycle.	2023/03/16
4-AI software life cycle specification		This deliverable includes the following considerations: a) Identification of all standards and best practices that are relevant for the AI for health software life cycle. Similar to other software life cycle processes, the AI software life cycle process needs to be specified. b) Summary and critical review of the identified documents including a discussion of their limits/gaps and need for action. c) Identification of life cycle steps that are specific/characteristic for AI for health software, such as training and test procedures based on data that potentially need to be annotated. d) Specification of the AI for health software life cycle and definition of best practices for the different life cycle steps in one document (under consideration of a, b, and c). Overview and examples of best practices	2023/03/14

Deliverable		Scope	Last update
5-Data specification		This document combines a set of six separate deliverables as umbrella, which address six important aspects related to data specification when used for artificial intelligence (AI) and machine learning (ML) models/methods for health purposes. Each editor proposed an initial outline (=Table of Contents), defined the objectives of the future deliverable, and collected a bibliography of existing literature and material relevant for the development of the respective document. A short call for participation, the expertise profile of potential contributors, a time plan, and a brief characterisation of the target audience serve as preface.	2020/06/17
5.1	Data requirements	This document lists acceptance criteria for data submitted to the FG-AI4H and states the governing principles and rules. These principles are crucial because the core of the benchmarking framework for AI for health methods will be an undisclosed test data set – per use case of each topic area to be defined – that will not be made accessible to the AI developers.	2023/03/16
5.2	Data acquisition	This document presents a framework for public healthcare data acquisition and management model based on standard protocol for its easy adoption by any country or international health organizations. This paper assumes basic digitization of electronic health record (EHR) at basic health facilities. There is a gap in developing an integrated and comprehensive framework that addresses the use of EHR in a standardized way for public health, privacy issue by anonymizing patient specific information, fusing multiple records with slight changes in the same information, augmenting a broad spectrum of contextual data, and so on.	2020/05/19
5.3	Data annotation specification	This document is committed to give a general guideline of data annotation specification, including definition, background and goals, framework, standard operating procedure, scenario classifications and corresponding criteria, as well as recommended metadata, etc. A questionnaire is attached to seek input and collaboration with topic groups in FG-AI4H regarding data annotation.	2023/03/16
5.4	Training and test data specification	This document is intended to guide the target audience with a systematic way of preparing technical requirements specification for datasets used in training and testing of machine ML models This document explains the best practices of data quality assurance aimed at minimizing the data error risks during the training and test data preparation phase of machine learning process lifecycle. The training and test data requirement specifications follow the data integrity, data security and data safety norms of the AI data governance lifecycle process.	2023/03/16
5.5	Data handling	This document outlines how data will be handled, once they are accepted. Health data are one of the most valuable and sensitive types of data. Handling this kind of data is often associated with a strict and factual framework defined by data protection laws. There are two major issues that the data handling policy should address: (a) compliance with regulations dealing with the use of personal health data; and (b) non-disclosure of the undisclosed test data held by FG-AI4H for the purpose of model evaluation.	2023/03/16
5.6	Data sharing practices	This document aims to provide an overview of the existing best practices for data sharing of health-related data, including the requirement to enable secure data sharing and issues related to data governance. The documents described established solutions and novel approaches based on distributed and federated environments.	2021/05/19
6-AI training best practices specification		This document aims to provide best practices for training and documentation so as to facilitate maximum performance and transparency. This document provides a review of the different aspects of AI model training pipeline. The first part discusses the best practices for data pre-processing aspects, while the second part discusses the best practices for AI model training aspects.	2023/03/16

Deliverable		Scope	Last update
7-AI for health evaluation considerations		This introduction with considerations on the evaluation of AI for health sets the scene for the five related documents DEL07.1-5. In this document, an overview of the deliverables DEL7.1-5 is given, preliminary considerations on the evaluation process are being made, characteristics of health AI validation and evaluation that are novel are identified, and the concept of standardized model benchmarking is introduced. Moreover, requirements for a benchmarking platform are considered in detail and best practices for the health AI model assessment are collected from selected sources.	2023/03/16
7.1	AI4H evaluation process description	The AI4H evaluation process description serves as overview of the state of the art of AI evaluation principles and methods and a forward-looking initiator for the evaluation process of AI4H. This process description includes a review of existing evaluation principles and methods, evaluation need and solutions specific for AI4H. It will also look into ethics and risks aspects of AI4H evaluation. Furthermore, based on the fundamentals of AI, the description will gain insights on the direction of how the current evaluation methods evolve towards the concept of REAL AI.	2020/05/20
7.2	AI technical test specification	This document specifies how an AI can and should be tested in silico. Among other aspects, best practices for test procedures known from (but not exclusively) AI challenges will be reviewed in this document. Important testing paradigms that are not exclusively related to AI applications should be mentioned too.	2023/05/04
7.3	Data and artificial intelligence assessment methods (DAISAM) reference	This document provides a summary of how to understand and identify algorithmic bias at different stages of the AI-based product that may have critical implications when the algorithm is applied in a real-world clinical setting. The aim is to train the most accurate model for each group without harming any minority group of patients. Furthermore, methods to mitigate bias according to the problem at hand are provided. These guidelines aim to provide a framework for technologists that build health related AI based products to investigate the presence of algorithmic bias.	2022/09/15
7.4	Clinical evaluation of AI for health	This document is to outline the current best practices, the principles and outstanding issues for further considerations related to clinical evaluation of AI health technologies. It serves as the output document of the WHO/ITU Focus Group on AI for Health (FG-AI4H) Working group on Clinical Evaluation of AI for Health (WG-CE).	2023/03/16
7.5	Assessment platform	This document demonstrates the development of the assessment platform that can be used to perform health AI evaluation for the different topic groups.	2020/05/30
8-AI4H scale-up and adoption			–
9-AI4H applications and platforms		This document contains a discussion on development of AI tool for Health using Mobile Applications & Cloud-based AI applications. This document describes type of mobile applications and the development of App based system for disease surveillance in the health sector.	2021/05/20
9.1	Mobile applications	This document contains a draft set of rules for development of AI tool for Health using Mobile Applications, their testing and benchmarking. It is to prepare the rules for development of AI tool for Health using Mobile Applications, and discuss the regulatory/ethical rules for Mobile Apps with AI for Healthcare.	2023/03/09
9.2	Cloud-based AI applications	This document contains a draft set of rules for development of Cloud-based AI applications, their testing and benchmarking. It is to discuss on technology, security and legal issues related to cloud-based AI tools, and to provide a forum for open communication among various stakeholders.	2020/05/21

10 Topic groups summary

To provide a quick overview of the specific health domains with corresponding AI/ML tasks considered in FG-AI4H, a summary table of all Topic Description Documents (TDD) is given below. Key messages includes health domain, task classification, gold standard, input data type, testing/training dataset, data annotation, algorithm, evaluation, etc. These information listed below were obtained from seven responses of a questionnaire to all TG drivers (<https://forms.gle/3fYrm3SZSrNQu3eeA>) .

Table 3 – Summary of Topic Groups (DEL 10.1-10.24)

Topic Groups (Examples)	Domain (Cardiovascular/ Dermatology/ Histopathology/ etc.)	Task (Classification/ detection/ segmentation/ prediction/etc.)	Gold Standard (state-of-the-art task intervention method)	Input data type (Text/ Image/ video/ audio/ numerical/etc.)	Testing/ Training dataset (Public dataset/ Collected by myself/etc.)	Data annotation (Procedure/ annotator number/ tool/ etc.)	Algorithm (specific model used in this TG)	Evaluation (Metrics used in this TG)
TG-Bacteria	Diagnoses of bacterial infection and anti-microbial resistance	Classification	clinical microbiologists with 4 to 5 years of specialization	2D Image	Self-built	Manual annotation	Models were trained using Tensorflow in Python via Google Colab; Artificial Neural Networks/ANNs	Accuracy
TG-Cardio	cardiovascular disease	prediction	clinical CVD risk scoring tools/calculators (WHO, 2019)	Quantitative & qualitative data (structured)	De-identified retrospective secondary data from healthcare/EMR & research data repositories	Structured data are used, thus simple R programming is used to recode structured data to required standardized labels.	Support Vector Machines/SVM; Random Forest/RF; & Artificial Neural Networks/ANNs	Accuracy of each risk prediction; Kappa statistic
TG-Dental	Dental diagnostics and digital dentistry	Classification/ detection/ segmentation/ prediction	Histology, Cross-image validation, human annotations	2D Image, 3D Image, Video, Text	Self-built	Custom made tool	ML/DP models; ML approach based on a multicenter dataset of panoramic radiographs	Accuracy, Sensitivity/recall, Specificity, Positive predictive

Topic Groups (Examples)	Domain (Cardiovascular/ Dermatology/ Histopathology/ etc.)	Task (Classification/ detection/ segmentation/ prediction/etc.)	Gold Standard (state-of-the-art task intervention method)	Input data type (Text/ Image/ video/ audio/ numerical/etc.)	Testing/ Training dataset (Public dataset/ Collected by myself/etc.)	Data annotation (Procedure/ annotator number/ tool/ etc.)	Algorithm (specific model used in this TG)	Evaluation (Metrics used in this TG)
							(OPGs) and electronic medical records (EMR)	value/precision, Negative predictive value, Area under the curve, F1- score/Dice score, Intersection over Union, AFROC, among others
TG-Derma	Dermatology	Classification	Not mentioned	2D Image	Public dataset (EDRA, ISIC, Dermofit, AICOS) and private data	Manual annotation	Not mentioned	Sensitivity; Specificity; F1- score
TG-Diabetes	Primary and secondary diabetes prediction	Classification/ detection/ prediction/ segmentation	Biological markers	2D Image	Publicly dataset (EyePACS), MESSIDOR dataset, The DIARETDB dataset, Diabetic retinopathy datasets from Kaggle, undisclosed data	Manual annotation	Not mentioned	Accuracy; Cohen's Kappa and Qua Diabetic- retinopathyatic Weighted Kappa
TG-DiagnosticCT	Volumetric chest computed tomography	Classification/ detection/ segmentation	Cross annotation by doctors, Pathological report	2D Image	Public dadset; Self-built	Cross annotation	slice-aware network (SANet)	Nodule-based sensitivity, false positive rate, free-response ROC, location ROC, AFORC,F-score, time spent reading the cases; varying sizes, varying densities, varying shapes; Nodule characterization, Subtle case performances

Topic Groups (Examples)	Domain (Cardiovascular/ Dermatology/ Histopathology/ etc.)	Task (Classification/ detection/ segmentation/ prediction/etc.)	Gold Standard (state-of-the-art task intervention method)	Input data type (Text/ Image/ video/ audio/ numerical/etc.)	Testing/ Training dataset (Public dataset/ Collected by myself/etc.)	Data annotation (Procedure/ annotator number/ tool/ etc.)	Algorithm (specific model used in this TG)	Evaluation (Metrics used in this TG)
TG-Endoscopy	Endoscopy	Classification/ detection/ segmentation	Pathological report, Cross annotation by doctors	2D Image, Video	Public dataset, self-built	Cross annotation, Self- built annotation tool	deep CNN based on SegNet architecture	True positive (TP) True negative (TN), False- positive (FP), False-negative (FN), Recall (REC) or Sensitivity (SENS), Precision (PREC), Specificity (SPEC), Accuracy (ACC), Matthews correlation coefficient (MCC), F1 score (F1), DICE coefficient (DICE), Jaccard coefficient or Intersection over Union (IoU), Polyp detection rate (PDR), Adenoma detection rate (ADR), Detection Latency, Average precision (AP), mAP, Runtime
TG-FakeMed	AI-based detection of falsified medicine	Classification/ detection/ prediction	Not mentioned	2D Image, Text	Self-built	Python/pip Labelling	the Random Forest algorithm	relation to parameters stakeholders need for decision making, scores that providers use, the scope providers designed their solutions for, the state of the art in RCT, statistics, AI benchmarking etc., bias transparency
TG-Falls	Falls among the elderly	Prediction	Systematic reviews report, algorithm	table and additional text files	public dataset, Anonymised Microfile Dataset	Manual annotation	Supervised Learning	The AUC, The sensitivity and specificity at a cut-off maximising the Youden index, The Brier score
TG-Fertility	Human reproduction and fertility	Identification, diagnosis, and treatment decision support	Relies solely in human judgment based on currently identified biomarkers	Demographics, genomics, behaviours and lifestyle, medical history, prior fertility	Not mentioned	Not mentioned	Not mentioned	Not mentioned

Topic Groups (Examples)	Domain (Cardiovascular/ Dermatology/ Histopathology/ etc.)	Task (Classification/ detection/ segmentation/ prediction/etc.)	Gold Standard (state-of-the-art task intervention method)	Input data type (Text/ Image/ video/ audio/ numerical/etc.)	Testing/ Training dataset (Public dataset/ Collected by myself/etc.)	Data annotation (Procedure/ annotator number/ tool/ etc.)	Algorithm (specific model used in this TG)	Evaluation (Metrics used in this TG)
				treatments, fertility treatment protocols, cycle data, and gamete and embryo development				
TG-Histo	Histopathology	Detection	Histopathology	2D Image	Self-built	Web-based annotation tool	Patch-wise classification	The detection performance (accuracy, F1 score, area under the curve of the receiver operating characteristic etc.) and the quantification error (e.g., the root mean square error)
TG-Malaria	Malaria detection	Detection	Manual microscopy	2D Image	Self-bult	Manual annotation	Image processing algorithm based on a pre-trained model of Faster Convolutional Neural Network (Faster R-CNN) and then integrated it with web-based technology; Mobile-Aware	ROC accuracy, precision, recall, specificity F1 scores, specificity, sensitivity, mean Average Precision (mAP), average precision

Topic Groups (Examples)	Domain (Cardiovascular/ Dermatology/ Histopathology/ etc.)	Task (Classification/ detection/ segmentation/ prediction/etc.)	Gold Standard (state-of-the-art task intervention method)	Input data type (Text/ Image/ video/ audio/ numerical/etc.)	Testing/ Training dataset (Public dataset/ Collected by myself/etc.)	Data annotation (Procedure/ annotator number/ tool/ etc.)	Algorithm (specific model used in this TG)	Evaluation (Metrics used in this TG)
							Deep Learning Algorithms	
TG-MCH	Maternal and child health	Prediction; warning	AI applications	Numeric variable, coded variable	Public dataset	Not mentioned	lightgbm (LGBM), xgboost (XGB), adaboost, catboost and random forest	AUC-ROC, precision, recall, accuracy, and F1-score
TG-MSK	Musculoskeletal medicine	Prevention, diagnosis, prognosis and treatment	Prediction modelling	EQ-5D-3L (VAS), Oxford Hip and Knee Score (Q score)	Self-built	Manual annotation	Not mentioned	Classifier's output value, Classifier certainty/uncertainty, Recall / sensitivity / true positive rate, Precision / positive predictive value
TG-Neuro	Neurological disorders	Classification/ detection/ prediction	Post-mortem pathology evaluation, and biological markers.	2D Image, 4D Image, clinical scores, genetics and biomarkers (e.g. csf)	Public dataset, self-built.	Manual annotation	Machine learning algorithm	Test accuracy, Clinical sensitivity, Clinical specificity, Clinical precision
TG-Ophthalmology	Ophthalmology	Classification/ detection/ segmentation/	Pathological report, Cross annotation by doctors	2D Image, 3D Image, Text	Public dataset, self-built	Cross annotation, Self-built annotation tool	Neural networks and deep learning for image recognition and classification; Decision Tree	Sensitivity or Recall or True Positive Rate, Specificity or True negative rate, Precision or Positive Predictive Value, F1-Score,

Topic Groups (Examples)	Domain (Cardiovascular/ Dermatology/ Histopathology/ etc.)	Task (Classification/ detection/ segmentation/ prediction/etc.)	Gold Standard (state-of-the-art task intervention method)	Input data type (Text/ Image/ video/ audio/ numerical/etc.)	Testing/ Training dataset (Public dataset/ Collected by myself/etc.)	Data annotation (Procedure/ annotator number/ tool/ etc.)	Algorithm (specific model used in this TG)	Evaluation (Metrics used in this TG)
								Accuracy, AUC (Area Under Receiver Operating Curve or ROC)
TG-Outbreaks (merged with TG-Sanitation)	Outbreak detection	Detection, prediction	Incorporated syndromic surveillance systems and valuable external data sources	Outbreak cases per week	Public dataset, and private data	Manual annotation	CUSUM, regression models	Sensitivity, Specificity, Precision, Negative predictive value, F1, ROC/AUC, Matthews Correlation Coefficient, Scaled probability of detection (POD), the Scaled POD, the Scaled POD, the average time before detection, absolute delays in detection
TG-POC	point-of care diagnostics	Diagnostic	Visual inspection with acetic acid (VIA), cytology- based screening tests, polymerase chain reaction (PCR)	2D Image	Self-built (from TG-POC)	Self-control	Image-based deep learning algorithm; an algorithm based on deep convolutional neural networks	Diagnostic accuracy (sensitivity, specificity, positive and negative predictive value), Cost-efficiency
TG-Psy	Psychiatry	Classification	Using classification schemes such as DSM-V and ICD-10	2D image (resting state EEG data)	Public dataset	Manual annotation	Deep learning, interpretable DNNs	Accuracy
TG-Radiology	Radiology	Detection	Using Deep Learning to detect diseases	2D image or 3D image, DICOM	Public dataset	Not mentioned	Convolutional Neural Networks,	Classification Accuracy, Classification

Topic Groups (Examples)	Domain (Cardiovascular/ Dermatology/ Histopathology/ etc.)	Task (Classification/ detection/ segmentation/ prediction/etc.)	Gold Standard (state-of-the-art task intervention method)	Input data type (Text/ Image/ video/ audio/ numerical/etc.)	Testing/ Training dataset (Public dataset/ Collected by myself/etc.)	Data annotation (Procedure/ annotator number/ tool/ etc.)	Algorithm (specific model used in this TG)	Evaluation (Metrics used in this TG)
			via medical images	or some other format			Transfer Learning, DenseNet 121, Bayesian Optimization, Strong Augmentations,	Error, Patient Level Accuracy & Image Level Accuracy, Pixel Accuracy, Exact Match Ratio (EMR), Example- Based Accuracy, Macro Averaged Accuracy, Micro Averaged Accuracy, Sensitivity, Specificity, Geometric mean (G-Mean), Precision, Recall, F-Measure, Fbeta- Measure, Matthews Correlation Coefficient (MCC), Macro Averaged Precision, Micro Averaged Precision, Macro Averaged Recall, Micro Averaged Recall, Negative Predictive Value (NPV), Receiver Operating Characteristic (ROC) Curve, Precision-Recall Curve, Average Precision (AP), Mean Average Precision (mAP),

Topic Groups (Examples)	Domain (Cardiovascular/ Dermatology/ Histopathology/ etc.)	Task (Classification/ detection/ segmentation/ prediction/etc.)	Gold Standard (state-of-the-art task intervention method)	Input data type (Text/ Image/ video/ audio/ numerical/etc.)	Testing/ Training dataset (Public dataset/ Collected by myself/etc.)	Data annotation (Procedure/ annotator number/ tool/ etc.)	Algorithm (specific model used in this TG)	Evaluation (Metrics used in this TG)
								Logarithmic loss or Cross-entropy, Brier Score, Brier Skill Score, Intersection Over Union (IoU), Hamming Loss, α - Evaluation Score
TG-Snake	Snakebite and snake identification	Classification	Snake expert (herpetologist) identification	2D Image	Public dataset, self-built.	Expert identification, crowdsourcing	AICrowd direct ingestion of algorithms & live leaderboard top algorithm available via API	F1 and logloss, scores that providers use, the scope providers designed their solutions for, the state of the art in RCT, statistics, AI benchmarking, bias transparency
TG-Symptom	Symptom assessment	Classification	Average doctor opinion.	Text, semantically structured cases.	Self-built.	a new case- creation tool	Vacant	Intended use of the device, Intended users, Risk classification, Point of Information and comparison to Standard of Care Intended/stated benefits to the user, clinical workflow and health system

Topic Groups (Examples)	Domain (Cardiovascular/ Dermatology/ Histopathology/ etc.)	Task (Classification/ detection/ segmentation/ prediction/etc.)	Gold Standard (state-of-the-art task intervention method)	Input data type (Text/ Image/ video/ audio/ numerical/etc.)	Testing/ Training dataset (Public dataset/ Collected by myself/etc.)	Data annotation (Procedure/ annotator number/ tool/ etc.)	Algorithm (specific model used in this TG)	Evaluation (Metrics used in this TG)
TG-TB	Tuberculosis	Detection, identification, segmentation	Doctors-experts marking	2D Image	Not mentioned	Manual annotation, Cross annotation	Deep fully convolutional neural network (FCN) model	AUC, accuracy, sensitivity, specificity
TG-TM	Traditional medicine	Classification, prediction, clustering, or segmentation	One on basis (whole system approach) involving continuous interaction between the subject and the TM practitioner.	2D Image, text, semantically structured cases.	Not mentioned	Not mentioned	Adaboost (AB), artificial neural network (ANN), back propagation (BP), classification tree (CT), convolutional neural network (CNN), damped least- squares (DLS), decision tree (DT), deep learning (DL), k nearest neighbor (kNN), least absolute shrinkage and selection operator (LASSO), levenberg– marquardt algorithm (LMA), linear discriminant analysis (LDA), logistic regression (LR), machine learning (ML), multiple linear regression (MLR), multivariate pattern analysis (MVPA), naïve bayes (NB), naïve Bayesian (NB), neural network (BP), pattern identification algorithm (PI), probabilistic neural network method (PNN), quantitative	Not mentioned

Topic Groups (Examples)	Domain (Cardiovascular/ Dermatology/ Histopathology/ etc.)	Task (Classification/ detection/ segmentation/ prediction/etc.)	Gold Standard (state-of-the-art task intervention method)	Input data type (Text/ Image/ video/ audio/ numerical/etc.)	Testing/ Training dataset (Public dataset/ Collected by myself/etc.)	Data annotation (Procedure/ annotator number/ tool/ etc.)	Algorithm (specific model used in this TG)	Evaluation (Metrics used in this TG)
							structure-activity relationship (QSAR), random ensemble with asymmetric learning (REAL), random forest (RF), self- organizing map(SOM), support vector machine (SVM), synthetic minority oversampling technique (SMOTE), takagisugeno-kang fuzzy system (TSK), term frequency- inverse document frequency (TIFD), the deformable template(DT)	

* NOTE: Colour codes indicate TGs' status reflected in FG-AI4H website, as "Under online approval process" (green) and "archived" (blue). as of 2023-09-27.