INTERNATIONAL TELECOMMUNICATION UNION
TELECOMMUNICATION STANDARDIZATION SECTOR
STUDY PERIOD 2017-2020

FG-AI4H-B-103
ITU-T Focus Group on AI for Health
Original: English

WG(s):	Plenary	New York City, 15-16 November 2018

DOCUMENT

Source:	FG-AI4H
Title:	Updated draft criteria for data to be accepted by the FG-AI4H
Purpose:	Discussion

Contact: Markus Wenzel, Editor
Fraunhofer Heinrich-Hertz-Institut
Germany
Email: markus.wenzel@hhi.fraunhofer.com

Abstract: This document provides criteria for data to be accepted by 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.

Rationale

Artificial Intelligence (AI) can help achieving the important objective of ensuring health for everyone in many ways, worldwide, often at reduced costs and enhanced speed. In the case of modern AI, it is important to notice that practitioners, patients and medical device regulators are confronted with a new kind of machine. While mechanical devices, electronics and software tools from the past have been typically designed from fully understood first principles, it is difficult to anticipate the behaviour of modern AI algorithms, because (1) of the enormous complexity of the algorithms, and (2) because the performance depends not only on the learning algorithm, but also on the underlying training data.

These properties let the users raise doubts about whether they can trust AI models, when they face critical decisions in the health domain. Crucially, these reasonable doubts cannot be resolved at present, because there are no established ways to assess the quality of AI models for health.

The Focus Group on "Artificial Intelligence for Health" will meet this need by demonstrating how the performance of AI solutions for health can be evaluated in a systematic fashion. For this purpose, a benchmarking framework will be developed in a best practice type of approach for representative use cases. Having successfully demonstrated the benefits of benchmarking for selected representative use cases, will allow for expanding the approach to a wider range of use cases. Details of the envisioned benchmarking procedure are presented in the White Paper of the Focus Group.

The core of the benchmarking framework is 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.

Optionally, additional **public training data** may be made publicly available by the Focus Group.

Scope

This document specifies the criteria for data to be accepted by the Focus Group and states the governing principles and rules.

A) General requirements for data to be accepted

- The submission includes a description of the data provenance/source of the data (How was the data collected and/or aggregated? Who has created the labels / ground truths? Who has assessed the data, e.g. with respect to quality?)
- Data to be accepted by the Focus Group have to follow the applicable laws and regulations for data acquisition, processing and sharing, such as privacy laws, copyright laws etc.
- Affirmation that informed written consent to data acquisition, processing (incl. benchmarking) and sharing was obtained (if applicable).
- Affirmation that approval was obtained from local ethics committee where the data was generated (if applicable).
- Adherence to the principles of the Declaration of Helsinki.
- Submitters have to transparently describe potential biases. Bias can - arguably - not be avoided in typical cases of data acquisition and can be expected even in an expert setting (in hospitals, diagnoses and treatment decision are made by experts but might be biased towards reimbursement from health insurances).
- The anonymization/pseudonymization/privacy procedure must be detailed. The Focus Group will follow the best practices from hospitals or other institutions (if applicable).
- Data must belong/correspond to a topics area/use case of interest - to be selected after prioritization among the various possible use cases. Data sets are not limited to any modality (such as images, time series, laboratory tests, "omics", text, electronic health records etc.), but a wide variety is welcome. Exemplary use cases may include AI-based diagnostics, treatment decision making, triage, patient self-management, risk assessment, image segmentation or annotation, early detection, among others. Obviously not all possible use cases can be addressed considering the limited timespan and resources of the Focus Group.
- The submission has to include a description of the data set and of the targeted use case, which must be concise but contain sufficient detail and help the Focus Group to understand data and use case and to decide about their suitability for the project.
    - High-level description of the data acquisition and of the objectives thereof.
    - Detailed protocol that allows for reproducing the data acquisition. With the protocol, independent trusted institutions should be able to acquire more undisclosed validation data for the benchmarking framework.
    - Statement whether the submitter would be willing to help the focus group to record more data, in principle.
    - Type of data (e.g. images, time series, laboratory tests, "omics", text, electronic health records, etc. or combinations thereof)
    - Type of output variables (such as ICD or ICHI codes, triage tags, pixel or voxel labels in a segmentation or annotation task…)
    - Size of the data (file size, number of samples)
    - Structure of the data (database type, file format and content, …)
    - Description of how the data can be loaded (special software required, data loader/importer functions available for common programming languages)?
    - Description of data (pre-)processing methods. Has the raw data been preserved or have the submitters applied any cleaning mechanism? Have they done any data transformation or processing on the collected data?
    - What are their data qualifiers? What is their definition of data quality and quality standards?
    - What is the current ownership of the data?
    - Is their data comparable with other similar data sets?
    - Does the data conform to any standards for health data?

  - If the submitters provide time series patient or clinical data (in the sense of data with a unique identifier collected over certain time intervals, not continuous EKGs data type) are there any gaps or redundancies?
    - Do they have imputations or projections of the data?
    - How was the missing or incomplete data treated if the case?
  - What is the degree of measurement precision of their data?
  - How and where is the data currently hosted/stored?
  - What are their data handling procedures and data governance processes? Knowing these might help the Focus Group in assessing the quality of their data.
  - How do the submitters handle data versioning? What is the frequency of their data updates, if applicable to the particular use case?
  - In case of API access to the data, are there any limitations of their APIs in terms of response time or size of data packages the API endpoints are expected to return?
    - How do they handle the release of a new version of their API?
    - If the APIs are used directly in the tests, will a new release maintain backwards compatibility? This information ensures that the tests will not break.
- Inclusion of contact details and relevant information about the submitter.
- Clear declaration of any potential conflicts of interest.
- Data will be only accepted when the data management procedure of the Focus Group has been specified and when a description is available of what the Focus Group is going to do with the data.

B) Requirements for undisclosed data to be accepted

- For undisclosed data, the general requirements for data apply (cf. section A).
- The undisclosed data are crucial for the benchmarking procedure. Therefore the safe storage has to be assured.
- Clearance for use in benchmarking (under compliance with the relevant laws, e.g. copyright, privacy).
- Data should be suited for benchmarking (to be defined by respective working groups and topic drivers).
- Consent to keep test data undisclosed.
- Description of the measures that guarantee secrecy in the past, present and future.
- Specification of who has had access to the test data in the past, at present, in the future (e.g. published or plan to share with other researchers).
- Sufficient size of data to create a statistically valid output report. Further criteria (heterogeneity, real-world relevance, …) of the data might be considered depending on the use case (to be defined by respective working groups and topic drivers).
- For sensible benchmarking, the following dilemmas will be addressed by the topic drivers per use case / topic area:
  1. Benchmarking is not valid if AI-solutions developed by data donors are tested on their own donated data, because they know the data and associated output variables/labels.
  2. Excluding data donors from benchmarking will considerably reduce the willingness to donate data (that are essential for a reasonable evaluation).
  3. Having a data pool from several sources and testing each AI-solution only on data from other sources (i.e. testing AI-solution developed by x only on data donated by y and z) may tempt data donors that also develop solutions to contribute as "difficult" data (low quality data, wrong annotations, ...) as possible to the data pool, in a competitive setting.

C) Requirements for public data to be accepted

- For public data, the general requirements for data apply (cf. section A).
- Clearance for data publication (under compliance with the relevant laws).
- It is important to notice that public data sharing is not required for implementing the core idea of the benchmarking framework, but merely an optional extra. Data publication is not something new, is not primary objective of the Focus Group, and the corresponding problem space has already been addressed by others.

_____