|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ITU Logo | INTERNATIONAL TELECOMMUNICATION UNION  **TELECOMMUNICATION STANDARDIZATION SECTOR**  STUDY PERIOD 2017-2020 | | FG-AI4H-D-022 | |
| **ITU-T Focus Group on AI for Health** | |
| **Original: English** | |
| **WG(s):** | | N/A | Shanghai, 2-5 April 2019 | |
| **DOCUMENT** | | | | |
| **Source:** | | Chongqing University | | |
| **Title:** | | Proposal to change the structure of FG-AI4H-C-105: TDD template | | |
| **Purpose:** | | Discussion | | |
| **Contact:** | | Daidi Zhong Chongqing University P.R.China | | Tel: +86-13696454858 E-mail: [daidi.zhong@hotmail.com](mailto:daidi.zhong@hotmail.com) |

|  |  |
| --- | --- |
| **Abstract:** | This document suggests some modifications of the structure of the FG-AI4H Generic topic description document (TDD) template defined in document FG-AI4H-C-105. The purpose is to sort out the relationship between main topic and sub-topic in the aspects of Method. |

# Original motivation of defining topic and sub-topic

According to former discussions, AI4H members understood that there are multiple collectively interested topics that are close to each other, but with minor differences. Or, there can be the case that member-A tries to address a bigger problem with a comprehensive dataset while member-B concentrates on a smaller problem with a subset of the previous dataset, with the common understanding that they both work in the same topic area.

As a result, we agreed to create sub-topics under the topic, in order to allow members to address both a wider problem (based on a comprehensive dataset) or a focused problem (based on a tailored subset). Therefore, from the perspectives of dataset and benchmarking, the subtopic should be the subset of the topic, although reasonable extension/adoption should be allowed. Such understanding corresponds well to the existing classification of clinical problem domains, to which we are targeting.

# Problems in current TDD template

The current TDD template contains a chapters called ‘Method’ to describe the topic for which we are trying to benchmark. With the above understanding, both the full-set and subset of such contents should be described. First of all, as the full-set of a certain topic, the comprehensive details of all aspects in Method should be described. After that, each sub-topic simply just to indicate which items are applicable to this very sub-topic. It could be presented as a cross-link table.

Correspondingly, the Results should contains separate segments for benchmarking results for topic and each sub-topic.

However, the current document structure of TDD template allow such “tree-like” structure.

# Proposal

To insert multiple sections under each sub-item of the chapter ‘Method’; each insertion is for one sub-topic. The same structure can be applied to chapter ‘Results’.

The layout could be like below:

# Method

* Overview of the benchmarking

## AI Input Data Structure

* possible inputs for benchmarking
* ontologies, terminologies
* data format
* Subtopic 1
  + Among the aforementioned ontologies, terminologies and formats, indicate the elements that are applicable to this subtopic.
* Subtopic 2

## AI Output Data Structure

* outputs to benchmark
* ontologies, terminologies
* data format
* Subtopic 1
  + Among the aforementioned ontologies, terminologies and formats, indicate the elements that are applicable to this subtopic.
* Subtopic 2

## Test Data Labels

* label types
* ontologies, terminologies
* data format
* Subtopic 1
  + Among the aforementioned label types, terminologies and formats, indicate the elements that are applicable to this subtopic.
* Subtopic 2

## Scores & Metrics

* which metrics & scores to use for benchmarking
* considering relation to parameters stakeholders need for decision making
* considering scores that providers use
* considering the scope providers designed their solutions for
* considering the state of the art in RCT, statistics, AI benchmarking etc.
* considering bias transparency
* Subtopic 1
  + Among the aforementioned metrics & scores, indicate the elements that are applicable to this subtopic.
* Subtopic 2

## Undisclosed Test Data Set Collection

* raw data acquisition / acceptance
* test data source(s): availability, reliability,
* labelling process / acceptance
* bias documentation process
* quality control mechanisms
* discussion of the necessary size of the test data set for relevant benchmarking results
* specific data governance derived by general data governance document (currently C-004)
* Subtopic 1
  + Among the aforementioned aspects of Undisclosed Test Data Set Collection, indicate the elements that are applicable to this subtopic.
* Subtopic 2

## Benchmarking Methodology and Architecture

* technical architecture
* hosting (IIC, etc.)
* possibility of an online benchmarking on a public test dataset
* protocol for performing the benchmarking (who does what when etc.)
* AI submission procedure including contracts, rights, IP etc. considerations
* Subtopic 1
  + Among the aforementioned aspects of Benchmarking Methodology and Architecture, indicate the elements that are applicable to this subtopic.
* Subtopic 2

## Reporting Methodology

* Report publication in papers or as part of ITU documents
* Online reporting
* public leaderboards vs. private leaderboards
* Credit-Check like on approved sharing with selected stakeholders
* Report structure including an example
* Frequency of benchmarking
* Subtopic 1
  + Among the aforementioned aspects of Reporting Methodology, indicate the elements that are applicable to this subtopic.
* Subtopic 2

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_