|  |  |  |  |
| --- | --- | --- | --- |
| ITU Logo | INTERNATIONAL TELECOMMUNICATION UNION  **TELECOMMUNICATION STANDARDIZATION SECTOR**  STUDY PERIOD 2017-2020 | | FG-AI4H-D-034 |
| **ITU-T Focus Group on AI for Health** |
| **Original: English** |
| **WG(s):** | | Plenary | Shanghai, 2-5 April 2019 |
| **DOCUMENT** | | | |
| **Source:** | | Chair WG-Operations | |
| **Title:** | | Draft revised Call for Proposals: use cases, benchmarking, and data | |
| **Purpose:** | | Discussion | |
| **Contact:** | | Markus Wenzel | Email: [markus.wenzel@hhi.fraunhofer.de](mailto:markus.wenzel@hhi.fraunhofer.de) |

|  |  |
| --- | --- |
| **Abstract:** | This document contains a proposed update to the Call for Proposals: use cases, benchmarking, and data document B-102, based on the experience gathered from the submissions received at meetings C and D.  [The ITU/WHO Focus Group on "Artificial Intelligence for Health" invites proposals for representative use cases and associated data sets. These data sets will be used to test AI technology for health within a standardized benchmarking framework. If you are interested in getting involved, please prepare a proposal according to the instructions in this document, e-mail it to [tsbfgai4h@itu.int](mailto:tsbfgai4h@itu.int) by 23 May 2019, and present it at meeting E in Geneva, Switzerland, from 30 May to 1 June 2019.] |

# Motivation

Over the last decade, considerable resources have been allocated to exploring the use of AI for health. This has revealed an immense potential but also exposed a weakness. AI algorithms are highly complex and depend on the underlying training data. If AI algorithms are poorly designed and/or the training data are biased or incomplete, errors or problematic results can occur.

An AI algorithm can only be used with complete confidence if it has been quality controlled through a rigorous evaluation against a system of standards. Towards developing such standards, the International Telecommunication Union (ITU) has established the Focus Group on "Artificial Intelligence for Health" (FG-AI4H) in partnership with the World Health Organization (WHO). ITU and WHO have considerable experience in the standardization of information/communication technologies and the health domain, respectively, making FG-AI4H ideally suited for establishing a standard assessment framework of AI for health (cf. our [comment](https://doi.org/10.1016/S0140-6736(19)30762-7) in *The Lancet* and our [Whitepaper](https://www.itu.int/en/ITU-T/focusgroups/ai4h/Documents/FG-AI4H_Whitepaper.pdf)).

The discussions within FG-AI4H have led to a two-step process. First, a *Call for proposals: use cases, benchmarking and data* (i.e. this call) is issued to generate the necessary framework to issue the secondary *Call for AI technology for health*, which is targeted towards evaluating AI applications for the use cases and data that were selected and evaluated as a result of this initial call. This approach provides an open and transparent process that allows participants to become involved with suggesting and selecting use cases and the associated data, and to participate in the testing of AI algorithms at an international/world-wide level.

Hence, with this initial call, FG-AI4H will explore the feasibility of use cases for evaluating AI applications. One necessary condition for the feasibility of a use case is that data need to be identified because AI algorithms can only be benchmarked against actual test data. For a valid evaluation, these test data must remain undisclosed before the test. Furthermore, these data must be of high quality, ethical source, and accompanied by information about their format and properties. Finally, these data should originate from a variety of sources, so that we can ascertain whether an AI algorithm can generalize across different conditions, locations, and/or settings (e.g. across different people, hospitals, and/or measurement devices). Additional information can be found in the documents [C-103](https://www.itu.int/en/ITU-T/focusgroups/ai4h/Documents/FGAI4H-C-103.pdf) (“data acceptance and handling policy”) and [C-104](https://itu.int/en/ITU-T/focusgroups/ai4h/Documents/FGAI4H-C-104-DraftThemClassifScheme.pdf) ("thematic classification scheme") on our [website](https://www.itu.int/go/fgai4h).

Another important condition for a use case to be considered feasible is that the exact health problem(s) that an AI algorithm should solve needs to be specified. The health problem should be relevant (e.g. it impacts a large and diverse part of the global population and/or it solves a problem that is difficult/problematic/expensive for humans) and the AI algorithm should promise tangible improvements relative to the current practices either in terms of quality, efficiency, or both (leading to better care, results, and/or cost effectiveness).

Finally, the benchmarking process needs to be defined and a statistical metric (or metrics) should be selected to evaluate the performance of the AI algorithm.

FG-AI4H has decided to first issue this call for proposals of representative use cases and of associated data sets. By responding to this call, the initial steps are made toward creating the framework that is needed to enable the FG-AI4H to arrive at the call for AI solutions. Hence, responding to this call by providing the information requested below marks the start of the described process.

**Logistics**

To prepare your proposal, please download the document template from  
<https://www.itu.int/en/ITU-T/focusgroups/ai4h/Documents/FG-AI4H-Doc-template.docx>

Update the document header as follows. WG(s): *N/A - Geneva, 30 May – 1 June 2019*. Source: *Your name*. Title: *Title of your use case and data*. Purpose: *Discussion*. Contact: *Insert contact name, contact organization, country, telephone, e-mail*. Abstract: *Summarize your proposal*.

Enter all information requested in sections 1-11 below (using the same outline and section titles), and e-mail your proposal as Word file to [tsbfgai4h@itu.int](mailto:tsbfgai4h@itu.int) by 23:59 (CET) 23 May 2019 at the latest.

**You are required to present this proposal at the FG-AI4H meeting C in Geneva from 30 May to 1 June 2019, answer questions, and provide updates if necessary.** PowerPoint slides can complement your presentation in Lausanne. All proposals will be published on our collaboration site. For meeting attendance, please register on the website, where you will also find information about our workshop at 29 May, just prior to the meeting, during the *[AI for Good Global Summit](https://aiforgood.itu.int/)*.

# Overview

Please give a general overview of the use case and of the health problem that is being addressed.

# Relevance

Please explain the relevance of the health problem.

# Impact

How could an AI algorithm provide an improvement relative to the current practice (e.g. will it lead to better care, lower cost, or higher efficiency)? Furthermore, describe the potential impact of benchmarking AI-based solutions for the use case.

# Existing work

Does the project start from scratch? If not, please provide a brief overview of existing work in the area of the project. Specifically, please explain the current state of the art and how the problem is currently addressed.

# Feasibility

Is the project feasible, based on the current state of the art? Please explain in a few sentences how you see the project progressing from start to finish.

# Data availability

Please describe the data sets that are available for the project. In particular, can you provide data that have not been (and will not be) disclosed? These test data will only be used to evaluate the AI solutions. Can an example subset of the data be made available to AI developers? Please briefly describe the data format and how (if relevant) the data have been labeled/annotated. Do you see legal obstacles to sharing the data with FG-AI4H (data protection and privacy laws, copyright)?

Furthermore, are there open data sets available for training purposes? Would you be able to contribute to an open data set for training purposes?

# Data quality

Please demonstrate that the available data are of high quality.

# Annotation/label quality

Please demonstrate that the annotations/labels of the data are of high quality.

# Data provenance

Please demonstrate that the data have been obtained in a professional and ethical way. Do the data come from a variety of sources (e.g. clinical environments)?

# Benchmarking

Please describe what type of AI algorithm you expect to be benchmarked by your test data. How should the AI algorithm be evaluated (including statistics/metrics)?

# Organizer

Please describe why your organization is interested in this project, and if you have run similar projects/benchmarks/challenges before.

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