|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ITU Logo | INTERNATIONAL TELECOMMUNICATION UNION  **TELECOMMUNICATION STANDARDIZATION SECTOR**  STUDY PERIOD 2017-2020 | | **FG-AI4H-C-106** | |
| **ITU-T Focus Group on AI for Health** | |
| **Original: English** | |
| **WG(s):** | | N/A | Lausanne, 22-25 January 2019 | |
| **DOCUMENT** | | | | |
| **Source:** | | FG-AI4H | | |
| **Title:** | | Call for proposals of open-source software that enables the FG-AI4H to run the benchmarking procedure on computing infrastructure of the United Nations [similar to challenge platforms from machine learning or data science] | | |
| **Purpose:** | | Discussion | | |
| **Contact:** | | TSB | | Email: [tsbfgai4h@itu.int](mailto:tsbfgai4h@itu.int) |

|  |  |
| --- | --- |
| **Abstract:** | Call for proposals of open-source software that enables the FG-AI4H to run the benchmarking procedure on computing infrastructure of the United Nations. The intended benchmarking platform is similar to challenge platforms from machine learning or data science. |

The WHO/ITU Focus Group on "AI for Health" (FG-AI4H) will evaluate artificial intelligence (AI) technology that could serve for the betterment of health. For this purpose, the FG-AI4H is creating an evaluation community around selected use cases, and will establish an AI benchmarking platform.

In this call, we invite you to propose open-source software that enables the FG-AI4H to run the benchmarking procedure on computing infrastructure of the United Nations. The intended platform can be compared very well with popular challenge platforms from the machine learning and data science communities. On the United Nations servers, the AI to-be-evaluated generates output variables y from undisclosed test data x. The output y=f(x) will be compared with the “ground truth” - typically labels or annotations given by experts. This benchmarking with agreed upon metrics allows for evaluating the AI technology, without unclosing and inspecting the technology itself. Importantly (and in contrast to some data science challenges), not only the ground truth, but also the test data have to remain undisclosed on the servers.

# Requirements for the software platform

The software platform is open-source for reasons of transparency and trust.

The platform enables the FG-AI4H to benchmark AI with undisclosed test data, by running the submissions on the servers, to create the predictions/output variables, and to compute the benchmarking metrics (as described above).

Containers (like Docker) allow for running software with various, different dependencies on the servers. AI technology is submitted in containers and not as readable source code.

An API makes it possible to control and monitor the different steps of the benchmarking procedure, and to interact with the test data and to issue the output variables and benchmarking measures.

Compute-intensive submissions using different software dependencies can be evaluated, too.

Sensitive test data (health data) as well as the submitted AI technology to-be-evaluated (intellectual property/trade secret) have to remain undisclosed on the United Nations server and must be protected with the highest possible security standards. Therefore, the software should automatically create a completely closed environment to run the containers (AI solutions) such that the container code is not able to transfer the data out to any remote location.

The frontend allows for testing submissions in a “dry run” before the actual benchmarking. A leader board enables comparisons of the (potentially anonymized) submissions.

**Procedure**

To propose software, please complete the document template found at : <https://www.itu.int/en/ITU-T/focusgroups/ai4h/Documents/FG-AI4H-Doc-template.docx>. First, update the document header as follows. WG(s): N/A - place, date. Source: Your name. Title: Proposal of open-source benchmarking software: <name of software>. Purpose: Discussion. Contact: Insert contact name, contact organization, country, telephone, e-mail. Abstract: Short summary of main text.

In the main text, introduce the software and explain how it meets the requirements detailed above. This document will be published on our collaboration site. Please email the completed document template to [tsbfgai4h@itu.int](mailto:tsbfgai4h@itu.int)

# About the ITU/WHO Focus Group on "Artificial Intelligence for Health"

***Motivation:*** Ubiquitous digital health data and powerful artificial intelligence (AI) algorithms are revolutionizing healthcare and -research. However, while AI algorithms can assist with many tasks in health including diagnosis, decision-making, or early detection, these algorithms are highly complex and depend on the underlying training data. When AI algorithms are poorly designed and/or the training data are biased or incomplete, errors or problematic results can occur. An AI algorithm should only be used with high 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 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.

***Management:*** FG-AI4H is chaired by Prof. Dr. Thomas Wiegand (Fraunhofer HHI/TU Berlin, Germany), with the vice-chairs: Shan Xu (CAICT, China), Stephen Ibaraki (ACM and REDDS Capital, U.S.A.), Ramesh Krishnamurthy (WHO, Switzerland), Naomi Lee (The Lancet, U.K.), Sameer Pujari (WHO, Switzerland), and Marcel Salathé (EPFL, Switzerland).

***Overall process of the focus group:*** (1) Invite proposals for use cases and data from the area of artificial intelligence for health. Select relevant use cases (e.g., automatic early detection of diabetic retinopathy from photos to prevent vision loss). Collect undisclosed test data per use case. Description of the prediction problem, including the format of the test data x and the possible output y. Agree on benchmarking metrics for comparisons. (2) Invite submissions of AI technology. (3) Benchmark AI with undisclosed test data: AI generates output variables y from undisclosed test data x. Compare output y with “ground truth” (labels/annotations).

***Logistics:*** Participation in the FG-AI4H is free of charge and open to all. For workshop/meeting attendance, please register on the website (<https://www.itu.int/go/fgai4h>), where you can also read the Whitepaper and key output documents, and sign up to the mailing list (with the free ITU account, which gives you access to the Collaboration site with all documents). The FG-AI4H secretariat has the following email address: [tsbfgai4h@itu.int](mailto:tsbfgai4h@itu.int)

***Current state:*** Workshop A at WHO-HQ · Geneva · Sep. 2018 (Talks from medical, AI, public health, and industry experts, initial working group structure established, white paper published, call for proposals (CfP) for use cases). Workshop B at Columbia University · NYC · USA · Nov. 2018 (Talks from medical, AI, regulatory, IT security, public health, and industry experts, 15 proposals for use cases of "AI for Health" presented, 8 use cases considered for entering the stage of a feasibility study ranging from risk assessment of breast cancer based on histopathological images to the early detection of diabetic retinopathy for the prevention of vision loss; Documents: data acceptance criteria, thematic classification, and data handling policy). Workshop C at EPFL · Lausanne · Switzerland · 22-25 January 2019 (workshop with talks about many important aspects of the endeavour; ongoing work in the meeting towards step 2 - cf. above - of the overall process).

***Upcoming workshops/meetings:***2–5 April 2019 in Shanghai, China; ​29-31 May 2019 in Geneva, Switzerland; September 2019, t.b.d.; November 2019 in New Delhi, India