|  |  |  |
| --- | --- | --- |
| ITU Logo | INTERNATIONAL TELECOMMUNICATION UNION**TELECOMMUNICATIONSTANDARDIZATION SECTOR**STUDY PERIOD 2017-2020 | FG-AI4H-F-037 |
| **ITU-T Focus Group on AI for Health** |
| **Original: English** |
| **WG(s):** | Plenary | Zanzibar, 3-5 September 2019 |
| **DOCUMENT** |
| **Source:** | WG-O & WG-DAISAM |
| **Title:** | Draft for online collaboration tools, rules and guidelines |
| **Purpose:** | Discussion |
| **Contact:** | Markus WenzelWG-O & Fraunhofer HHIGermany | Email: markus.wenzel@hhi.fraunhofer.de  |
| **Contact:** | Luis OalaWG-DAISAM & Fraunhofer HHIGermany | Email: luis.oala@hhi.fraunhofer.de  |

|  |  |
| --- | --- |
| **Abstract:** | In this draft, we suggest a set of tools and corresponding rules for online collaboration, which is a major opportunity for carrying forward work in between meetings and for global participation in the focus group, considering time and financial constraints. Guidelines accompany these tools and rules and summarize the overall process of the focus group. |

# Motivation

The mission of the ITU/WHO focus group on “AI for Health” is to undertake crucial exploratory steps towards evaluation standards for health AI that are applicable on a global scale. This approach offers substantial potential for synergies, because many national regulatory institutions, public health institutes, physicians, patients, developers, health insurances, licensees, hospitals, and other decision-makers *around the globe* can profit from a common, standardized benchmarking framework for artificial intelligence (AI) and machine learning (ML) solutions for health.

Standards live on being sustained by a broad *community*. Therefore, the focus group is creating an ecosystem of diverse stakeholders from industry, academia, regulation, and policy with a common, substantial interest in health ML/AI benchmarking. Since its foundation in July 2018, the focus group has been organizing a series of workshops with subsequent multi-day meetings in Europe, North America, Asia, and Africa, so far, every two or three months for engaging the regional communities. Participation in the focus group is free and open to all and attending the events on-site is encouraged. The recent generous support from a charitable foundation, with travel grants for priority regions, will further foster the global participation.

However, considering time and resource constraints, it is not possible for every relevant stakeholder and expert to join every meeting. Moreover, work needs to be carried forward in between meetings. Therefore, participation via internet from remote and online collaboration are major opportunities for global inclusion of all relevant expertise.

In order to start the discussion about a global online structure of the focus group, we briefly introduce existing work in this direction at ITU. Then, we suggest a set of tools and rules for online participation possibilities. Finally, we summarize the overall process of the focus group as a guideline.

# Electronic working methods at ITU

Our objectives are backed by ITU Resolution 32 (2016) “Strengthening electronic working methods for the work of the ITU Telecommunication Standardization Sector” (<https://www.itu.int/dms_pub/itu-t/opb/res/T-RES-T.32-2016-PDF-E.pdf>). The resolution

*“considered (...)*

* that electronic working methods (EWM) enable open, rapid and easy collaboration between participants in the activities of the ITU Telecommunication Standardization Sector (ITU-T);
* that the implementation of EWM capabilities and associated arrangements will have significant benefits for the ITU-T membership, including resource-limited individuals, organizations and states, by allowing them timely and effective access to standards information and the standards-making and approval process;
* that EWM will be advantageous in improving communication among members of ITU-T and between other relevant standardization organizations and ITU, towards globally harmonized standards;
* the key role of the Telecommunication Standardization Bureau (TSB) in providing support to EWM capabilities; (...)
* the budgetary difficulty developing countries have in participating actively in face-to-face ITU-T meetings;
* (...) that ITU should further develop its facilities and capabilities for remote participation by electronic means”.

It was *noted* “that many forms of EWM have already been implemented by ITU-T, such as electronic document submission and the electronic forum service; (...) the advantage to the membership of facilitating greater electronic participation in the development and approval of Recommendations, in particular by members unable to participate in study group meetings in Geneva and elsewhere”.

It was *“resolved* that the principal EWM objectives of ITU-T” are: “that collaboration between members on development of Recommendations should be by electronic means (...); to encourage electronic participation of developing countries in ITU-T meetings (...);”

The TSB director was “*instructed* (...)

* to provide the executive authority, budget within TSB, and resources to execute the Action Plan with all possible speed;
* develop and disseminate guidelines for the use of ITU-T EWM facilities and capabilities;
* take action, in order to provide appropriate electronic participation or observation facilities (e.g. webcast, audioconference, webconference/document sharing, videoconference, etc.) in ITU-T meetings, workshops and training courses for delegates unable to attend events in person, and coordinate with BDT to assist in the provision of such facilities”.

In response to the aforementioned Resolution 32, “TSB Electronic Working Methods (EWM) Services and FAQs” were developed: <https://www.itu.int/en/ITU-T/ewm/>

E-meeting/virtual meeting tools have been made available: <https://www.itu.int/en/ITU-T/ewm/Pages/e-meetings.aspx>

These tools include fully virtual meetings with corresponding guidelines: <https://www.itu.int/en/ITU-T/studygroups/2017-2020/Documents/SG_Guidelines_fully-virtual_meetings_082018.docx>

Instructions for e-meeting request: <https://www.itu.int/en/ITU-T/ewm/Pages/e-meeting-request.aspx>

ITU provides file repositories, so called “Informal FTP Areas” that “are used by the participants as a repository and an exchange facility for documents they are working on” (<https://www.itu.int/en/ITU-T/ewm/Pages/ifa.aspx>).

Already earlier on, in 2013, discussion boards and direct document posting were announced (cf. chapter): <https://www.itu.int/en/ITU-T/tutorials/Documents/201301/Session-14.pdf>

[Other standardization bodies provide online collaboration tools and guidelines as well, which can serve for orientation. For instance, ISO provides a portal with eCommittees, electronic balloting, web conferencing, automatic document submission and corresponding guidelines and training resources: <https://login.iso.org/portal/>; more info here: <https://isotc.iso.org/livelink/livelink/fetch/2000/2122/5156198/customview.html?func=ll&objId=5156198&objAction=browse&sort=name> ]

# Online collaboration tools and rules

We suggest the following set of online collaboration tools and corresponding rules, based on previous discussions within the focus group, according to the motivation (cf. chapter 1), and inspired by the ITU Resolution and previous work at ITU (chapter 2):

## Tools

* The future website [https://ai4health.itu.int](https://ai4health.itu.int/) of the ITU/WHO Focus Group on “AI for Health” will be the central information and collaboration hub of the focus group and a gateway to the AI benchmarking platform.
* Focus group members can contact other members via the general e-mail list (fgai4h@lists.itu.int) or via specialized lists (e.g. for the topic group on “Symptom assessment”). Before being able to send messages, members need to sign up to a mailing list as explained on the website.
* A kit of online collaboration tools will be made available to registered users. These tools will facilitate the remote collaboration among focus group members through online meeting services, discussion boards, forums, calendars, wikis, file repositories and collaborative document editing. Many of these tools are already available at ITU (e.g. Microsoft SharePoint apps [*to be unlocked*], Zoom, GoToMeeting, Pigeonhole Live, NextCloud Collabora [*to be unlocked*], GitLab [*to be unlocked*]). Further tools could be added, e.g. Microsoft Teams, Google Docs/Hangout/Drive, Slack, Monday). The website will integrate and explain these tools.
* Focus/working group members can ask the focus group secretariat at ITU (tsbfgai4h@itu.int) to set up a conference call via "Zoom" (or a comparable online meeting service).
* Potentially, a 24/7 automatic document submission system will be established.
* Visitors to the website will be able to access the full documentation of the Focus Group after registering and login. *[Comment: the usability of the registration process needs to be improved]*.

## Rules

* Relevant **communication** should be carried over the general or specialized **e-mail lists**.
* **Conference calls**
	+ should be announced one week in advance on the e-mail list; not be on major holidays; arranged at a time convenient for all interested participants.
	+ A link to the (self-starting and self-managed) online meeting ID is added to the online calendar of the collaboration platform (“SharePoint”).
	+ One of the participants of the call is appointed to take meeting minutes.
	+ A report of the conference call, which documents decisions and next steps, has to be written by the organizer of the call, supported by the minute keeper and the other attendees of the call. The report has to include the names of all participants, document any dissent if needed/requested, and has to be sent to the mailing list and to the ITU secretariat for publication on the documentation site.
* **Decision making:**
	+ Key decisions should be taken in physical meetings of the focus group.
	+ In order to allow the group to work more efficiently, an online decision-making process is considered as useful (cf. report of meeting E in FGAI4H-E-101).
	+ At e-meetings, (“delta”) decisions are taken by consensus, before they are summarized and submitted as input document for approval by the larger community.
	Note: Consensus does not imply unanimity, and is declared by the chairman or here: e-meeting organizer.
	+ The general focus group mailing list (fgai4h@lists.itu.int) is used to announce the decision being taken, providing links to relevant documents.
	+ A commenting period is specified, typically two weeks, for receiving comments with concerns. These comments should be addressed by email to the secretariat (tsbfgai4h@itu.int).
		- Absence of comments imply agreement to the proposed decision.
		- The outcome of the consultation is announced.
		- If comments are received, they are discussed and resolved by the focus group management in coordination with the commenters to create an amended decision.
		- If the amendment is minor, the chairman declares approval.
		- If the amendment is substantive, another consultation is started, or decision is postponed until the next meeting of the focus group.

# Overview of the focus group process as guidance for online collaboration

The focus group works as follows:

Specialized *topic groups* take charge of specific health domains with corresponding ML/AI tasks. At present, the topic groups address AI-based cardiovascular disease risk prediction, dermatology, histopathology, outbreak detection, ophthalmology, radiotherapy, symptom assessment, tuberculosis prognostics/diagnostics, and several further domains. In each topic group, different stakeholders, including competing companies, with a common interest in the topic are working together. “Calls for topic group participation” are published on the website (<https://www.itu.int/go/fgai4h>), introduce the respective topic group and invite participation. The creation of many other topic groups in response to the open “call for proposals: use cases, benchmarking, and data” is expected. Selection criteria include the prospect for a widespread and, ideally, global impact, a clear concept described in sufficient detail, and preliminary evidence for feasibility.

Every topic group defines its scope, the specific ML/AI tasks and the evaluation procedures with corresponding test data and metrics in a *topic description document* in full detail. Statistical metrics for assessing the model performance are, e.g., precision, specificity, F1 score, and area under curve, but can be multiple or combined metrics, too. In particular, it should be assured that the (e.g. clinical) endpoints are meaningful in practice. Further criteria should be taken into account, e.g. robustness to noise and other variations in the input data, or to manipulations. Humans prefer transparent decision-making: Can the model adequately quantify the uncertainty and plausibly explain the decision? These criteria beyond mere performance should be taken into account, too.

The topic description document must capture a range of aspects related to the test data, because they determine largely if the evaluation procedure is appropriate and meaningful. The procedure can return conclusive results if and only if the test data are realistic, i.e. close to the actual application, of representative coverage, and of traceable provenance from different sources. Data acquisition must be transparently documented in full detail, including potential annotation guidelines, for reproducibility, replicability, and scalability. All ethical and legal questions related to the acquisition, storage and processing of health data must be taken into careful consideration. Bias must be avoided, if possible, or at least documented clearly. The document must specify quality and quantity criteria for the test data, including corresponding references. Annotations or labels, must be given by experts with a defined level of expertise, with potentially several independent annotations per sample (if applicable). Technical matters, e.g. data formats and data management, need to be specified. A reference model can potentially be defined (e.g. “average human performance for this task”, “best in class”). Limiting factors for data availability should be referred to, such as finances or time.

The plan detailed in the topic description document must be implemented in practice. The test data must be provided or acquired, and measures for quality assurance taken. The evaluation routine must be implemented and the code published together with at least a few example data with references (e.g. annotated images) to enable the developers carrying out a trial run of their code.

For a clean and fair evaluation, a trusted third party should receive the trained model or expert system, as independent arbiter, and conduct the tests on data that have never been published before. This cautious procedure prevents unfair conduct, e.g. tuning the model for optimal performance on this particular test set (“overfitting”), without actually being able to generalize well to real-world data, which can be expected in practice. Therefore, widely available, public data sets cannot be used for the evaluation and the entire test data set must remain secret, i.e. neither labeled nor unlabeled test data should be made available. The model performance should be evaluated in a closed computing environment without internet access. Otherwise, test data could be leaked, against the rules, and the model be tweaked on the test data. Besides, leaderboard probing and other potential pitfalls known from ML challenges must be kept in mind. The trusted third party is responsible to protect both test data and ML/AI model, because the test data have to remain secret for subsequent meaningful testing, and because the AI models may contain business-relevant intellectual property/trade secrets of the developer.

The focus group has created dedicated *working groups* that tackle cross-sectional tasks, such as data and AI solution handling, assessment methods, health requirements, operations, and regulatory considerations.

The working group *data and AI solution assessment methods* reviews the topic description documents, in collaboration with independent *experts* with substantial record of accomplishment in the respective health topic, with proficient knowledge in ML/AI, and with transversal competences from areas such as ethics and statistics. During a repeated review cycle, the working group and the experts check that the topic description documents are accurate, complete, sound, understandable, and objective, and give according feedback for improvement to the respective topic group and the entire focus group.

The working group *data and AI solution handling* takes charge for a range of tasks related to conducting the tests, which requires bringing the test data and the to-be-tested AI solutions together. Relevant aspects include, e.g., transfer agreements, secure data and solution transfer, data checks, IT infrastructure, access rights, traceability, IT security, test implementation, and report generation.

The working group for *regulatory considerations* is involved in the entire process, with representatives of FDA (USA), CMDE/ NMPA (China), EMA (Europe), and BfArM (Germany) so far. In close collaboration with the WHO, the working group facilitates subsequent steps (e.g. clinical trials, certification etc.) towards deployment of the health AI solution in practice.

Further information about the scope and general process of the focus group can be found in a commentary in *The Lancet* and a white paper on the website, where also the full documentation of all previous meetings is published. *[Parts of section 1 and 4 of this draft originate from a manuscript that will be submitted to the ITU Kaleidoscope conference 2019.]*

The figure on the next page illustrates the overall working process.



*Figure 1.* Illustration of the overall working process with deliverables and responsibilities of working groups (WG) interacting with one of several topic groups (TG).

TG: Topic Group
WG-DAISAM: Working Group Data and AI Solution Assessment Methods
WG-DASH: Working Group Data and AI Solution Handling
WG-RC: Working Group Regulatory Considerations
Experts: Health/ethics/statistics experts
Algo.: Algorithm for data/model evaluation, i.e. test implementation (automatic/with humans).

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