

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

Thomas Wiegand (Fraunhofer HHI, Germany), Naomi Lee (The Lancet, UK), Sameer Pujari (WHO), Manjula Singh (ICMR, India), Shan Xu (CAICT, China), Monique Kuglitsch (Fraunhofer HHI, Germany), Marc Lecoultre (MLLab.AI, Switzerland), Ana Riviere-Cinnamond (PAHO/WHO), Eva Weicken (Fraunhofer HHI, Germany), Markus Wenzel (Fraunhofer HHI, Germany), Alixandro Werneck Leite (University of Brasilia, Brazil), Simão Campos (ITU), and Bastiaan Quast (ITU)

E-mail: [tsbfgai4h@itu.int](mailto:tsbfgai4h@itu.int)

Web: <https://www.itu.int/go/fgai4h>

## Abstract

Artificial intelligence (AI)—the phenomenon of machines being able to solve problems that traditionally require human intelligence—has seen an enormous rise in interest due to considerable advances in effectiveness and use. The health sector, one of the most important sectors for societies and economies worldwide, is particularly interesting for AI applications, given the ongoing digitalization of health data and the promise for an improved quality of health and healthcare. However, due to the complexity of AI models, it is difficult to distinguish good from bad AI-based solutions and to understand their strengths and weaknesses. This is crucial for clarifying responsibilities and for building trust among AI developers, AI regulators, and AI users. For this reason, the International Telecommunication Union (ITU) and the World Health Organization established the Focus Group on "Artificial Intelligence for Health" (FG-AI4H). Traditionally, the governance and delivery of healthcare services are the responsibility of a government (even for private providers and health insurance systems) and, thus, WHO or ITU Member States. FG-AI4H identifies issues on AI for health—relevant data, information, algorithms, and processes, which fosters opportunities for international standardization and the application of AI for health on a global scale. With members coming from research, healthcare, regulation, telecommunications, and health ministries, and complementary fields around the globe, FG-AI4H is able to draw on a wealth of expertise to produce (a) documentation that contains guidelines on how to evaluate AI for health from various perspectives (e.g., regulatory, ethical, and data or AI solution) and (b) an online platform (and complementary tools) for the benchmarking of AI for health. This document, the whitepaper for FG-AI4H, is adapted from an earlier document prepared by Salathé et al. (2018).

## 1 Introduction

Adopted in 2015 by all United Nations Member States, the Sustainable Development Goals (SDG) are an urgent call to action for shared peace and prosperity through improving health, education, reducing inequality and economic growth. The third SDG is dedicated to "Good Health and Well-Being." Further, the WHO Constitution establishes that "the enjoyment of the highest attainable standard of health" is a basic human right (WHO, 2006). To ensure that this goal is achieved in all regions of the world, including those regions who are trailing the pack, member states have pledged to "Leave No One Behind" (UNDP, 2020). Consequently, a top priority of the WHO is attaining universal health coverage: ensuring that all people can access the health services they need without facing financial hardship. To keep itself accountable, WHO has set three strategic targets: one billion more people should benefit from universal health coverage, one billion more people should be better protected from health emergencies, and one billion more people should enjoy better health and well-being (WHO, 2018). Artificial intelligence (AI) and other digital technologies will be vital in achieving these three targets. In recognition of the growing importance of digital health

technologies—including AI—the WHO Member States unanimously adopted the resolution on Digital Health during the 71st World Health Assembly on 26 May 2018 in Geneva, Switzerland (WHO, 2018). During the opening speech of the 144th session of the WHO Executive Board on 24 January 2020 in Geneva, Switzerland, the importance of digital health (and, particularly, AI for health) was reinforced: “the future of health will be influenced by digital health significantly [and WHO must] embrace it, but at the same time, WHO should be ahead of the curve in digital health, in order to contribute to global health [...] working with the International Telecommunications Union to find new ways of using artificial intelligence to get care to remote communities” (WHO, 2020).

For centuries, good health and affordable health care for everyone has been a key goal of most governments, and public health breakthroughs such as vaccination and antibiotics are generally credited with having saved—and continuing to save—billions of lives (Orenstein and Ahmed, 2017). Additionally, many countries see universal health coverage as a priority, and are looking for efficient ways to tackle that challenge. Thus, it is not surprising that when a new technology reaches high levels of performance, the healthcare sector is a key area of application: the potential for improving public health (compounded by economic opportunities) is enormous. AI is one of these new digital technologies that presents the opportunity to expand healthcare services, be that in remote areas to those most in need, in large urban areas by speeding up the processing of large amounts of information, or even during health emergency periods (as we are currently seeing with the COVID-19 pandemic). Therefore, it is prudent to look at the potential of AI in helping to solve health-related issues at local and global scales. This short paper describes the current applications of AI in the health domain, discusses challenges, and proposes solutions in order to unlock the full potential of the technology guided by standardized and assessed good practices that all stakeholders can apply so that the potential of AI for health is realized in a way that is effective and safe, but also that the gains are equitable.

## **2 Artificial intelligence**

The term “artificial intelligence” is not new. As an academic field, it dates back to at least the mid-20th century. Since then, it has gone through multiple cycles of substantial progress, followed by inflated expectation, and then disappointment. A combination of new machine learning algorithms, increased computational power, and an explosion in the availability of very large data sets (“big data”) (Samek et al., 2018), as a consequence of the digitalization of health information, has led to recent stunning advances, with demonstrations of machines achieving human-level competence at solving clearly defined tasks across many domains (e.g., breast cancer screening; McKinney et al., 2020). The current cycle is primarily driven by the impressive progress made by deep learning, a branch of machine learning that effectively uses artificial neural networks to address problems of unprecedented difficulty. Applications of deep learning have achieved human or superhuman performance in many fields such as image recognition and natural language processing (Esteva et al., 2019). An important feature of deep learning is that neural network parameters are tuned in an automated process of complex, multi-layered iterative training. In many cases, no expert-level knowledge is used in the training process, other than direct input and output parameters (e.g., sets of pixels and their associated labels), giving rise to the so-called “end-to-end” learning (Esteva et al., 2019). In other words, the networks learn to go directly from one end—the input—to the other end—the output—without requiring any domain-specific expertise in between. The resulting network structures are generally very large, with oftentimes billions of parameters, and of such complexity that it is impossible to describe in simple terms how they work, which has led to new challenges concerning their explainability and interpretability.

### 3 Artificial intelligence for health

The recent digitalization of all types of health data and the fact that computers are increasingly able to interpret some non-medical images and text almost as accurately as humans (He et al., 2015; Wu et al., 2016) enables countless applications of AI in health. Much of the recent work on AI for health has gone into applications that revolve around image interpretation and natural language understanding.

In the field of medical interpretation of images, one of the most publicized studies is by Esteva et al. (2017). The authors presented the accurate classification of skin lesions using a clinical image-trained deep neural network and assessed performance by comparing classifications with those made by board-certified dermatologists. This revealed that the network reached human levels of accuracy; noting, however, that there is a need for validation in a wider geographical area wherein the lesion appearance and colour may vary per the regional variation in skin colour. Litjens et al. (2017) reviewed over 300 papers using deep learning in medical image interpretation. These papers focused on detection, segmentation, or classification tasks. They covered the analysis of X-ray, CT, MRI, digital pathology, cardiac, abdominal, musculoskeletal, foetal, dermatological, and retinal images. In natural language understanding, the areas of biomedical text mining, electronic health record analysis, sentiment analysis on internet-derived data, and medical decision support systems have shown favourable results (Ching et al., 2018). Furthermore, AI methods can automatically interpret laboratory results (ranging from standard blood testing to recent advances in high-throughput genomics and proteomics; e.g., Gunčar et al., 2018) and time series (e.g., electrocardiogram, temperature, oxygen saturation, and blood pressure; e.g., Attia et al., 2019). AI can also be used beyond the specialized hospital level. For example, it can be used from primary healthcare centres to different hospital specialization levels, including national institutes of health or national reference laboratories. The role of AI differs per the requirement and feasibility at the setting.

The possible application of such technologies has global potential: a large part of the world's population has access to devices that can utilize compute-intensive AI-powered applications (i.e., computers, smartphones, and other devices storing models locally or connected via the internet to powerful computing clusters; Albertini, 2019). Given the speed at which AI-based algorithms can be developed, improved, and deployed, the technology has the potential to ensure first-class medical decision-making is accessible and affordable worldwide (Bell et al., 2018). This could allow people to be reached in a faster and easier way, conditions could be diagnosed at an earlier stage, and this may lead to better health outcomes and lower costs. However, this would also need infrastructure and internet connectivity facilities, especially at remote and peripheral settings.

Although this progress is promising, AI for health also faces a number of challenges. As previously noted, deep learning models are notoriously hard to interpret and explain, which may substantially hinder their acceptance when facing critical—even vital—decisions. Thus, interpretability, explainability, and proven robustness (e.g., to outliers and to adversarial attacks) are crucial aspects that have to be considered for trustworthiness. Despite the accuracy reported for many AI for health models, there is currently a lack of data on effectiveness (particularly comparative effectiveness), cost effectiveness, or safety in a clinical setting.

The data on which the AI model is based should also include all relevant regional, gender, and age variations to be robust enough so that it can perform well in public health settings without bias. Moreover, access to health data can be hindered by strategic issues from data owners or custodians and because data are sensitive and subject to (country and region dependent) privacy-protection laws as well as ethical considerations on their acquisition and use. Therefore, access to sufficient testing data is a major limiting factor for the predictive performance of models on previously unseen data, especially because of local legal restrictions to access health data.

This problem is further complicated because most modern AI applications are based on supervised learning and rely on data that are labelled. In the health domain, labels can typically only be given

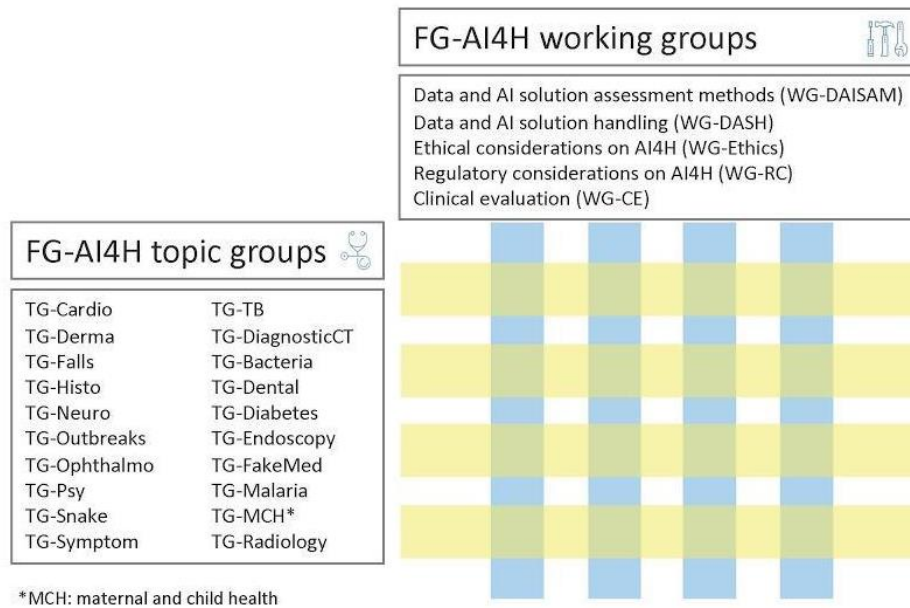
by qualified specialists. This is in contrast to, for example, simple object recognition, where photographs can be labelled by legions of laypersons. In addition, machine learning approaches must consider the biases (Caliskan et al., 2017) that data (e.g., text- and image-based medical data) likely contain (e.g., Hägele et al., 2020). In machine learning, models and training data must be considered in combination. The models cannot extrapolate. Rather, they can only learn patterns that are present in the training data. These data need to be of high quality, sufficient quantity to learn the myriad of parameters of the “data hungry” algorithms, and theoretically should cover all possible instances including outliers.

#### **4 Focus Group on "Artificial Intelligence for Health" (FG-AI4H)**

The ITU/WHO Focus Group on “AI for Health” (FG-AI4H) was established in July 2018 to develop international evaluation standards for AI solutions in health. The Focus Group works at the interface of multiple fields (e.g., machine learning/AI, medicine, regulation, public health, statistics, and ethics) and includes other decision-makers who value a standardized and transparent benchmarking framework. The international approach offers the opportunity to concentrate diverse national expertise in standardization frameworks on a global level.

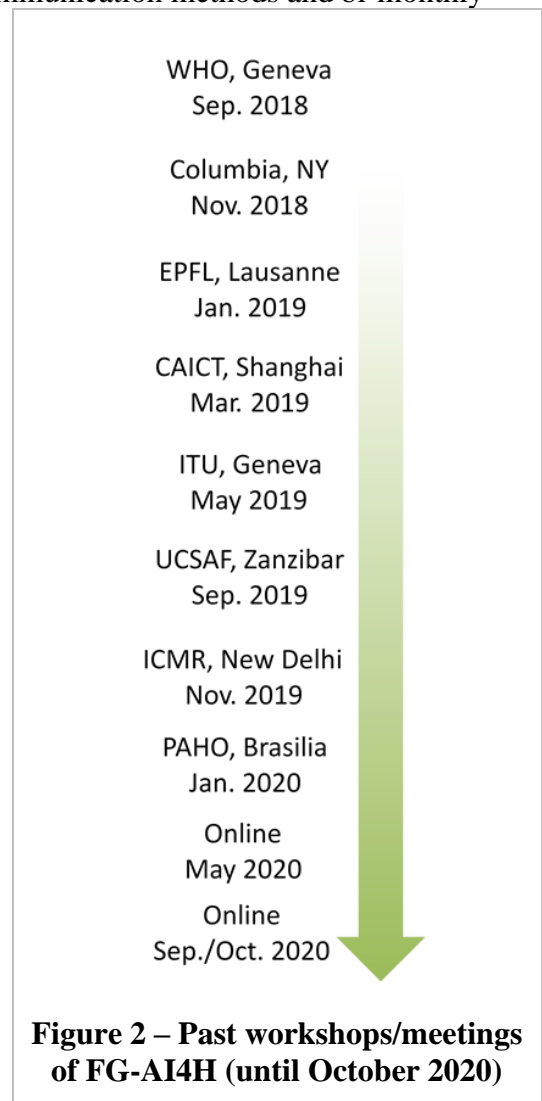
The overall objectives of the FG-AI4H are to tap this network of international expertise to create (a) guidelines (i.e., document-form “deliverables”) for the evaluation of AI for health and (b) to create an online platform and complementary tools for the benchmarking of AI for health. The Focus Group does not intend to specify the models themselves as an ITU-T Recommendation nor to standardize medical data formats or establish performance criteria of hardware running the AI algorithms. This vision was articulated in an earlier whitepaper (Salathé M et al. (2018)), has been explored in a commentary (Wiegand et al., 2019), and is updated here with the latest developments.

The success of FG-AI4H depends on the engagement of its participants. The chair and vice-chairs—with additional support from the FG-AI4H secretariat and greater management team—work in unison to ensure that FG-AI4H activities are progressing toward their objectives. Given the importance of health and machine learning and because participation is free and open to all, FG-AI4H has been able to attract and maintain a core group of participants representing stakeholders in the private and public sector from around the world: machine learning/AI researchers, healthcare practitioners and researchers, regulators, representatives of health ministries and ministries of telecommunication, international organizations, and individuals from complementary fields. With guidance from the technical experts in FG-AI4H, the various stakeholders have been able to improve their AI tools, which is evident in their subsequent progress. Collaboration within FG-AI4H is organized in two types of groups: those dedicated to specific health use cases (hereafter, topic groups) and those dedicated to overarching themes with relevance for all of the aforementioned health use cases (hereafter, working groups). Each topic group (as of July 2020, there were twenty in operation) is dedicated to a specific health use case in the context of AI with the intent of producing evidence and case studies. To do this, it aims to bring together a network of experts, to accrue or compile data, and to propose procedures to benchmark AI models developed for a special task within this health topic (e.g., the classification task of tumour tissue discrimination in haematoxylin-eosin-stained image patches could be found within the topic group on histopathology). Ultimately, however, the benchmarking is finalized through a consultation with the WHO (driven by the working groups). The working groups address those themes that affect all topic groups in a specific aspect of an AI for health application (e.g., ethical considerations or regulatory considerations). They also create definitions of best practices, establish processes and related policies, define ways to successfully benchmark AI for health algorithms, and create reference documents. This is illustrated in Figure 1. Note that the topic groups are added iteratively; therefore, they are at various stages of maturity.



**Figure 1 – The interaction of topic groups and working groups within FG-AI4H**

Collaboration within FG-AI4H revolves around online communication methods and bi-monthly (global health conditions permitting) on-site workshops/meetings. Online communication tools are the basis for weekly teleconferences of FG-AI4H management, topic groups, and working groups. They are also used in workshops of working groups (e.g., “AI for Health Standardized Assessment Framework - Handling and assessment methods” in January 2020; and “Regulatory considerations on AI for health” in May 2020). Historically, the majority of progress within FG-AI4H has occurred through online communication. Figure 2 shows the sequence of FG-AI4H plenary meetings from its creation until May 2020 (in addition, individual working groups and topic groups have also held on-site and online events; e.g., the aforementioned workshop of [WG-DAISAM/WG-DASH, Jan. 2020](#)). Bi-monthly on-site workshops and meetings provide an outreach opportunity for FG-AI4H: at each location, a day-long workshop enables local and regional experts and stakeholders to present their activities at the interface of AI and health, and to connect with members of FG-AI4H. The subsequent meeting provides a forum for FG-AI4H members to present progress in the various topic groups and working groups, for new topic group and working group proposals to be evaluated, and for feedback to be shared. This also provides an opportunity for local and regional experts and stakeholders to become engaged and committed to FG-AI4H activities. Thus far, this approach has proven highly successful as indicated by considerable geographical diversity within the Focus Group. Financial support from the Fondation Botnar has also been key in bringing individuals from developing countries to its on-site events.



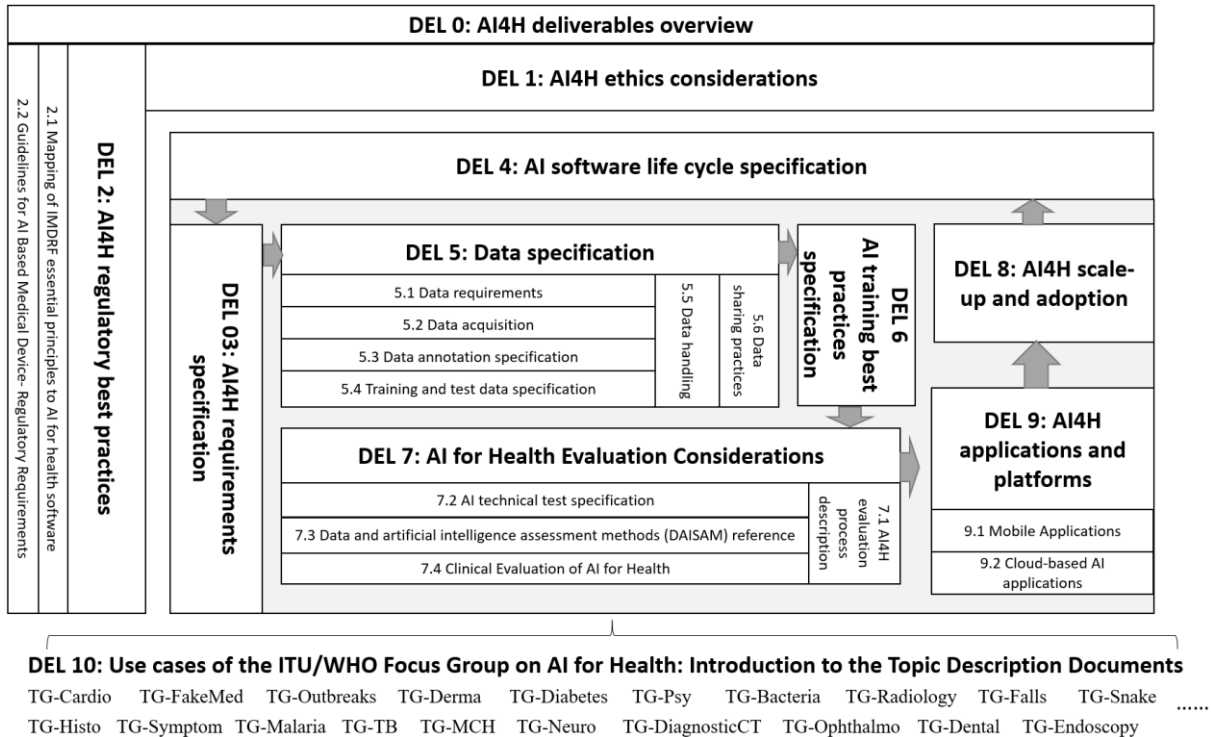
**Figure 2 – Past workshops/meetings of FG-AI4H (until October 2020)**

**a. FG-AI4H deliverables**

A main output of FG-AI4H are guideline documents hereafter known as “deliverables” (Figure 3, Table 1). The deliverables provide the requirements needed to establish a benchmarking process of AI for health. The deliverables represent a collective effort made by members of FG-AI4H. As the collaboration is an ongoing procedure, iterative versions of the deliverables are presented at each bi-monthly meeting. Each topic group and working group produces multiple deliverables. The former include topic description documents; the latter include guidelines on ethical considerations, regulatory considerations (best practices specification), requirements specifications, software lifecycle specifications, data specifications, AI training best practices specifications, evaluation specifications, scale-up/adoption, and FG-AI4H applications and platforms.

**Table 1 – Overview of deliverables**

No.	Deliverables categories
00	Overview of the FG-AI4H deliverables
01	AI4H ethics considerations
02	AI4H regulatory best practices
03	AI4H requirements specification
04	AI software life cycle specification
05	Data specification
06	AI training best practices specification
07	AI4H evaluation considerations
08	AI4H scale-up and adoption
09	AI4H applications and platforms
10	AI4H use cases: Topic description documents



**Figure 3 – Structure of deliverables**

## b. FG-AI4H platform and tools

The second major contribution of FG-AI4H will be an open-source software package for an online benchmarking platform and associated tools (e.g., to assist with annotation or data collection). The platform will contain undisclosed testing data on which AI models can be validated. In general, the procedure (from data accrual through AI solution validation) will consist of the following steps (see Figure 4):

1. Data accrual, task identification, and selection of test/benchmarking metrics:

Each FG-AI4H topic group assembles undisclosed (i.e., not publicly available) data for the given health topic, which will be used to validate AI models. The data are evaluated to avoid the risk of bias and other relevant data quality aspects. Then, the FG-AI4H topic group identifies and proposes special tasks within the health topic for which a benchmarking procedure is defined (with corresponding test metrics; example metrics include sensitivity, specificity, and area under the receiver operating characteristics curve). Considerations about clinical evaluation are made to ensure the technical tests consider relevant, correct, and meaningful objectives for utility in a “real world” environment.



**Figure 4 – FG-AI4H benchmarking workflow**

2. AI developers train AI models to produce solutions for the benchmarking task
3. The model is submitted to the online benchmarking platform.
4. There is a regulatory assessment and then the model code is validated on the benchmarking platform using the test data and results reports are generated.

Central to the FG-AI4H evaluation workflow is the independent benchmarking platform of step 4, which uses standardized test procedures and metrics (defined in step 1) on high-quality, representative, and undisclosed test data. FG-AI4H is considering two approaches for the benchmarking platform: assessment in a closed environment or via an interface.

- a. Closed environment: the AI developer submits the trained model to the platform. Within the closed environment, the model is given undisclosed test data. The test data are processed by the model and the resulting output is then compared with ground truth using the predefined evaluation metrics.
- b. Interface: The platform sends undisclosed test data to the model. The test data are processed by the model and the resulting output is then returned to the benchmarking platform. The platform compares the received output with the ground truth and computes the benchmarking metrics. The distinguishing feature of this approach is that the AI models are never uploaded to the benchmarking platform.

In addition to developing the benchmarking platform, FG-AI4H is providing tools (i.e., a data acquisition package, a data storage package, an annotation package, a prediction package, an

evaluation package, and a reporting package) to assist AI developers. One of these tools is dedicated to standardizing annotation. All topic groups contribute examples that, once validated, could help with establishing a standardized benchmarking for their task. This always includes a test dataset with labels/annotations (“*gold standard datasets*”) that an AI is expected to provide. If the annotation is biased in some way, the validation loses credibility. A data annotation tool that could be applied to various use cases would be invaluable for AI developers and evaluators. FG-AI4H is currently developing such a tool, which will have the following features:

1. **Common features:** An analysis was conducted of available annotation tools to identify common features. For example, two-dimensional image-based tasks (the most common input data modality in AI for health) always offer zoom, pan, rotate, contrast, brightness, drawing of outline, boxes, and circles.
2. **Specific features:** In order to cover all cases in data modality and task classification of various topic groups, a survey<sup>1</sup> collected topic group requirements for data annotation procedure and tools. These specific requirements will also be considered for the data annotation tool development.
3. **Performance requirements:** Some systematic performance requirements are also being considered. These include interoperability, transparency, compliance by design, safety, and security.

## 5 Future directions

Artificial intelligence offers a novel approach to address the shortage of healthcare professionals. This shortage is exacerbated by demographic changes and population growth, which implies that the demand for AI solutions for health will continue to grow in the future. The COVID-19 pandemic is a learning experience, which has revealed that the development of AI tools for health is in high demand. Currently, AI is rarely deployed in medical practice at a global scale due to technical, regulatory, ethical, and other concerns. Through establishing standardized assessment of AI-based solutions for health, FG-AI4H is assuring quality (and safety), fostering the adoption of AI in healthcare practice, and supporting global health. Since its establishment in 2018, FG-AI4H has made considerable progress toward its objectives. However, there is still much to be accomplished. The future activities of FG-AI4H will include:

- *Documentation:* Iterative revisions of the deliverables will reflect advances in this dynamic field. When deemed sound, the deliverables will be reviewed for approval from FG-AI4H members as output documents that could be adopted as international recommendations.
- *Evaluation of criteria for an online platform:* As indicated in the previous section, the online benchmarking platform is an undertaking that requires consideration of many variables. In addition to the overall approach, we need to consider requirement identification, program discussion, technical route selection, development and implementation of the online benchmarking platform, and development and implementation of associated tools (e.g., data annotation tool). Because this is an international effort, it may involve political discussion and require consensus.
- *Assembly of expertise and data:* From assembling undisclosed data to drafting deliverables, FG-AI4H requires a collective effort. Maintaining an engaged core community of FG-AI4H members and attracting new experts is key to ongoing success.
- *Identification of new themes:* As FG-AI4H moves full steam ahead toward its goals, it is important to maintain a level of pragmatism. As technology develops and healthcare priorities emerge, new themes for topic groups and working groups will be proposed, vetted,

---

<sup>1</sup> See <https://forms.gle/51iuHG5SrP6E8Hfr7>.



and incorporated into the overall agenda. A prime example is an ad-hoc group on Digital Technologies for COVID Health Emergency that was recently established in response to the ongoing COVID-19 crisis.

## References

- Albertini M (2019). New ITU data reveal growing Internet uptake but a widening digital gender divide. <https://www.itu.int/en/mediacentre/Pages/2019-PR19.aspx>
- Attia Z, Noseworthy P, Lopez-Jimenez F, Asirvatham S, Deshmukh A, Gersh B, *et al.* (2019). An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. *Lancet* 394(10201):861–7. ([https://doi.org/10.1016/S0140-6736\(19\)31721-0](https://doi.org/10.1016/S0140-6736(19)31721-0))
- Bell D, Gachuhi N, Assefi N (2018). Dynamic Clinical Algorithms: Digital Technology Can Transform Health Care Decision-Making. *Am. J. Trop. Med. Hyg.* 98(1), 9–14. (doi:10.4269/ajtmh.17-0477)
- Caliskan A, Bryson J, Narayanan A (2017). Semantics derived automatically from language corpora contain human-like biases. *Science* 6334:183–186.
- Ching T, Himmelstein D, Beaulieu-Jones B, Kalinin A, Do B, *et al.* (2018). Opportunities and obstacles for deep learning in biology and medicine. *J. R. Soc. Interface.* (doi: 10.1098/rsif.2017.0387)
- He K, Zhang X, Ren S, Sun J (2015). Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, 1026–1034. (doi:10.1109/ICCV.2015.123)
- Esteva A, Kuprel B, Novoa R, Kuo J, Swetter S, Blau H, Thrun S (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542, 115–118. (<https://doi.org/10.1038/nature21056>)
- Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, Cui C, Corrado G, Thrun S, Dean J (2019). A guide to deep learning in healthcare. *Nature Medicine* 25: 24–29
- Gunčar G, Kukar M, Notar M, *et al.* (2018). An application of machine learning to haematological diagnosis. *Sci Rep* 8, 411. (<https://doi.org/10.1038/s41598-017-18564-8>)
- Hägele M, Seegerer P, Lapuschkin S, Bockmayr M, Samek W, Müller K-R, Binder A (2020). Resolving Challenges in Deep Learning-based Analyses of Histopathological Images using Explanation Methods. *Scientific Reports* 10:6423.
- Litjens G, Kooi T, Bejnordi B, Setio A, Ciompi F, *et al.* (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis* 42:60–88.
- McKinney S, Sieniek M, Godbole V, *et al.* (2020). International evaluation of an AI system for breast cancer screening. *Nature* 577, 89–94. (<https://doi.org/10.1038/s41586-019-1799-6>)
- Orenstein W, Ahmed R (2017). Simply put: Vaccination saves lives. *PNAS* 114(16): 4031–4033.
- Salathé M, Wiegand T, Wenzel M, Kishnamurthy R. (2018). Focus Group on Artificial Intelligence for Health. [https://www.itu.int/en/ITU-T/focusgroups/ai4h/Documents/FG-AI4H\\_Whitepaper.pdf](https://www.itu.int/en/ITU-T/focusgroups/ai4h/Documents/FG-AI4H_Whitepaper.pdf)
- Samek W, Wiegand T, Müller K-R (2018). Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *ITU Journal: ICT Discoveries* 1(1), 39–48. (<https://arxiv.org/abs/1708.08296>)
- UNDP (2020). Sustainable Development Goals. <https://www.undp.org/content/undp/en/home/sustainable-development-goals.html>

WHO (2006), Constitution of the World Health Organization, 45th edition, as amended in October 2006. [https://www.who.int/governance/eb/who\\_constitution\\_en.pdf](https://www.who.int/governance/eb/who_constitution_en.pdf).

WHO (2018). WHO Director-General's closing speech for the 71st World Health Assembly. <http://www.euro.who.int/en/health-topics/Health-systems/e-health/news/news/2018/5/seventy-first-world-health-assembly-comes-to-a-close-after-charting-a-new-course-for-who>

WHO (2020). WHO Opening Speech at the 144th Session of the Executive Board. <https://www.who.int/dg/speeches/detail/opening-speech-at-the-144th-session-of-the-executive-board>

Wiegand et al., WHO and ITU establish benchmarking process for artificial intelligence in health. *The Lancet*, 394(10192), 9-11. [https://doi.org/10.1016/S0140-6736\(19\)30762-7](https://doi.org/10.1016/S0140-6736(19)30762-7)

Wu Y, Schuster M, Chen Z, Le QV, Norouzi M, *et al.* (2016). Google's neural machine translation system: Bridging the gap between human and machine translation. arXiv:1609.08144.

---