AI for Good Global Summit, May 2018: Conception of Focus Group on AI for Health

- Second AI for Good Global Summit: session on AI for health
- Need for partnership on AI for health, combining expertise in Health (WHO) and ICT (ITU)
- Idea for the Focus Group on AI for Health (FG-AI4H) is born
- ITU in corporation with WHO creates FG-AI4H in July 2018

Opening keynote by DG of WHO (Dr. Tedros)

ITU/WHO Focus Group on Artificial Intelligence for Health

AI for Health
An ITU Focus Group
In collaboration with WHO

Funding support by:
fondation BOTNAR

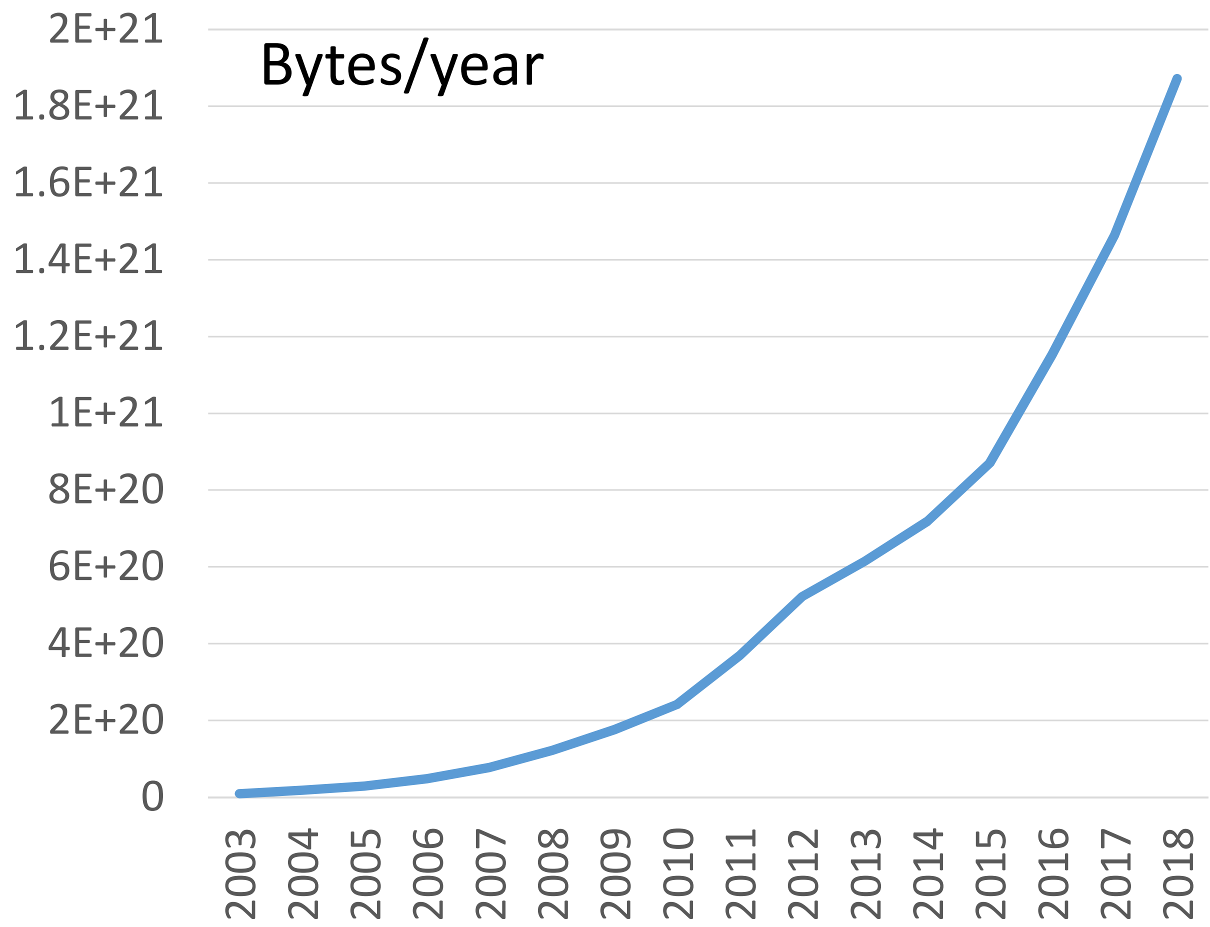
Size of the Internet

Running the Internet, ICT devices including their production, and data centers takes about 11% of the total electricity in 2018

Source: N. Jones, “How to stop data centres from gobbling up the world’s electricity,“ Nature, September 2018

„Nuclear energy now provides about 11% of the world's electricity from about 450 power reactors“

Source: www.world-nuclear.org, 2019



*Sources: Cisco VNI reports

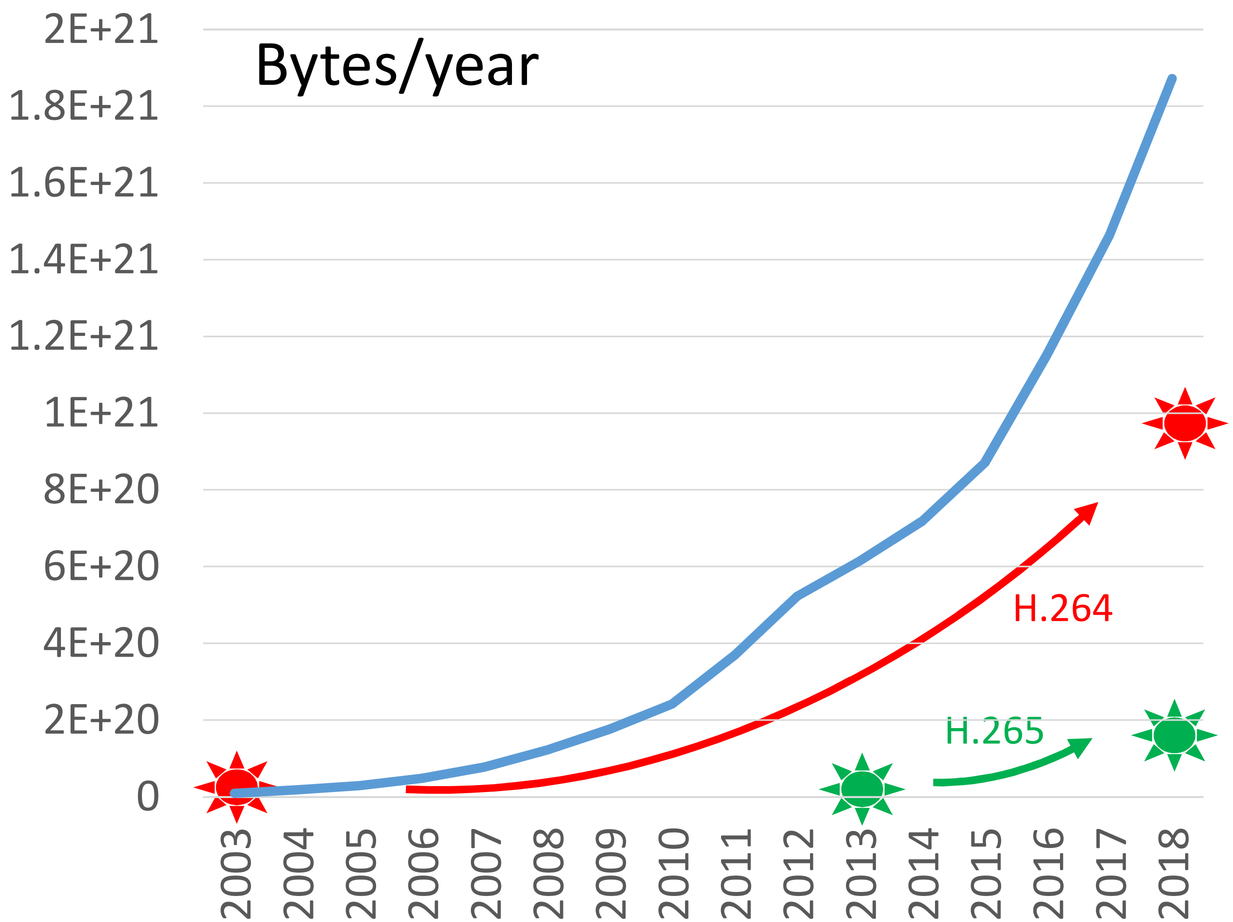
Scaling

Video compression standards

- H.264 (2003) – Now: 55%
- H.265 (2013) – Now: 7%

of all bytes on the Internet

Source: Encoding.com 2019 - Global Media Format Report

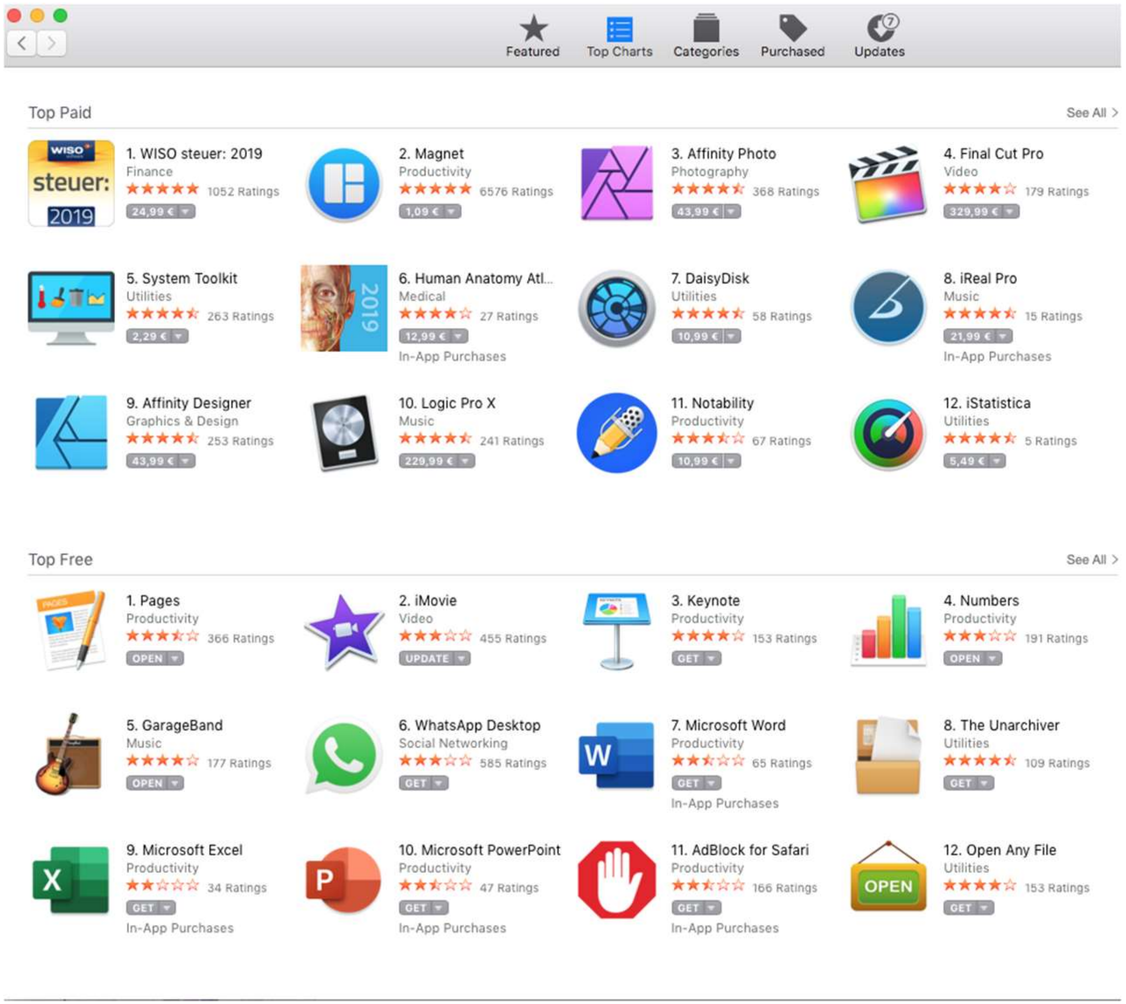


*Sources: Cisco VNI reports

What about AI for Health?

- Artificial Intelligence for Health (AI4H) offers substantial improvements for public and clinical health; e.g.,
  - early detection,
  - diagnosis,
  - risk identification,
  - treatment decision support,
  - self-management,
  - improved outcomes, …
- How can we achieve world-wide scaling of AI for Health?

World-wide Scaling: App Repository



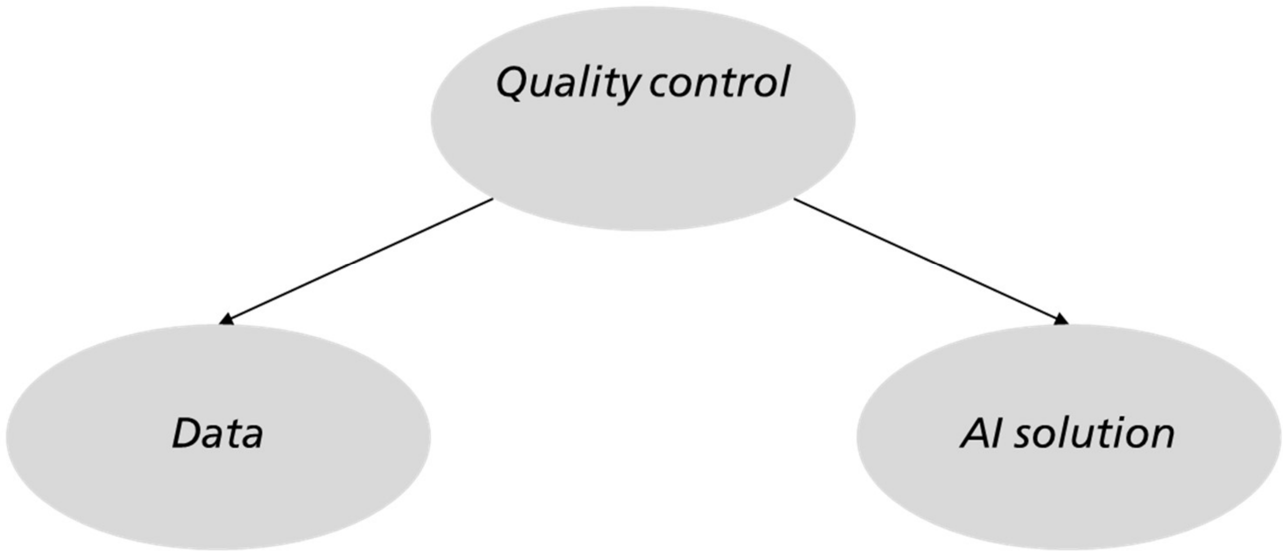
AI for Health
An ITU Focus Group
In collaboration with WHO

An App Repository for digital health?

What needs to be done to make it work?

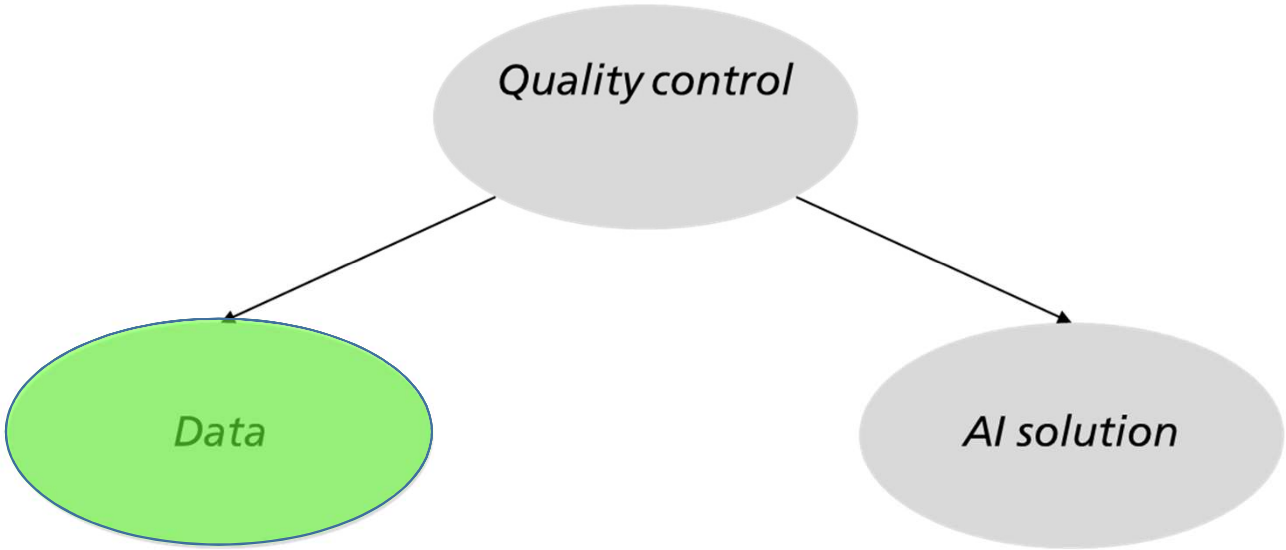
AI for Health Apps must be Quality Controlled

- For worldwide adoption, need evaluation standards
- AI heavily depends on the data used for training and testing
- Incorporation of domain (health/medicine) knowledge is essential



AI for Health Apps must be Quality Controlled

- For worldwide adoption, need evaluation standards
- AI heavily depends on the data used for training and testing
- Incorporation of domain (health/medicine) knowledge is essential



Quality control:
Reference and training data

Considerations:

1. Collection of training data
2. Reproducibility of training data
3. Statistical properties of training data
4. Generation of reference data through experts
5. Evaluation of data for machine learning systems
6. ...

1. Collection of training data

- Availability:
  Determined by the resources needed to collect data (time and finances)
- Bias:
  Circumstances can cause certain sources of data to be used preferentially
- Incompatibility:
  Determined by the comparability of data formats (interoperability, data classification system, precision of data classification, amplitudes, and ranges of values)
- Ethical and legal aspects of data collection:
  Consideration of data protection, discrimimation, etc.
- Test data must remain undisclosed
- …

2. Reproducibility of training data

- Measurement process:
  Data collection method (e.g., measurement or survey) must be reproducible
- Accompanying guidelines:
  Guidelines can vary for the same data collection process (consider stationary vs unstationary sources of data)
- Independence and consistency:
  Stationarity and ergodicity of the data to underlying processes
- …

3. Statistical properties of training data

- Homogeneity:
  Data sources should have a certain level of homogeneity (more or less equal cluster frequency in case of multiple clusters)
- Consideration of medical aspects:
  Data must be collected for all necessary categories of a given medical problem
- Completeness:
  Sufficient coverage for these categories (i.e., enough data points)
- …

4. Generation of reference data through experts

- Annotation:
  Clear guidelines for annotating outputs
- Classification:
  Approach for handling different types of annotation
- Independence:
  Flexibility to use different experts
- Active learning:
  Iteration between annotations and AI algorithms
- Reference model:
  Defining the "Gold Standard“ as a reference ("Best in Class," "average performance")
- …

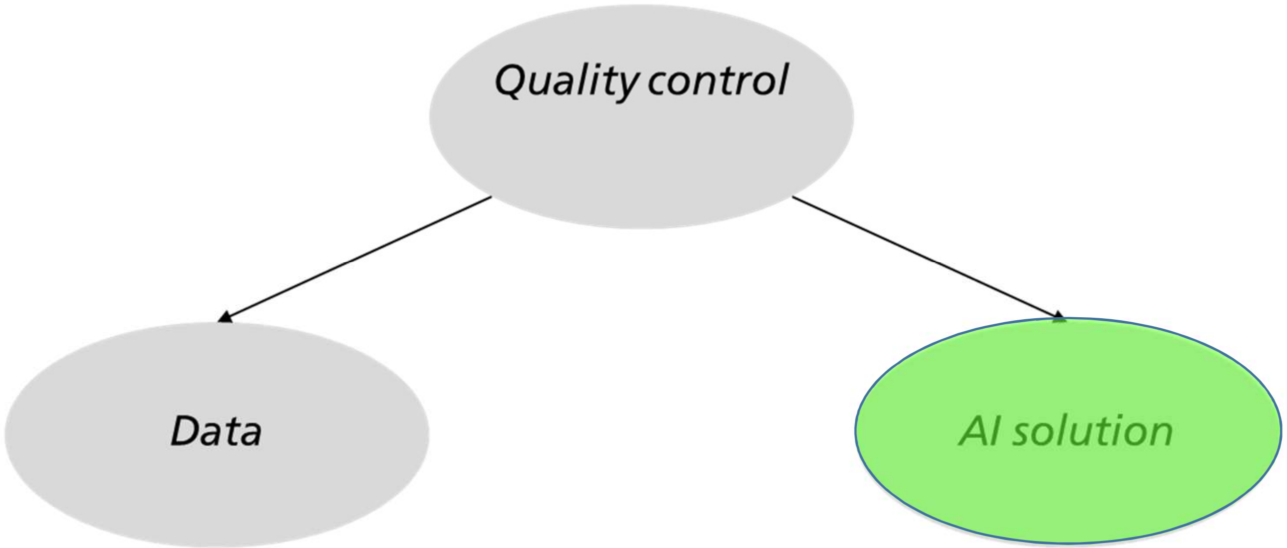
5. Evaluation of data for machine learning systems

- Dynamics:
  AI algorithm is evaluated for a given reference dataset and improved over time
- Incremental data collection:
  Feedback is acquired through implementation of evaluated AI algorithm, experts provide evaluation, and reference dataset is enhanced
- Guidelines:
  Guidelines for updating data come through experience and considering the properties relevant for machine learning
- Knowledge:
  Distinguishing among "forgetting“/"learning from," "specialization," and "problematic bias" …

AI for Health Apps must be Quality Controlled

AI for Health
An ITU Focus Group
In collaboration with WHO

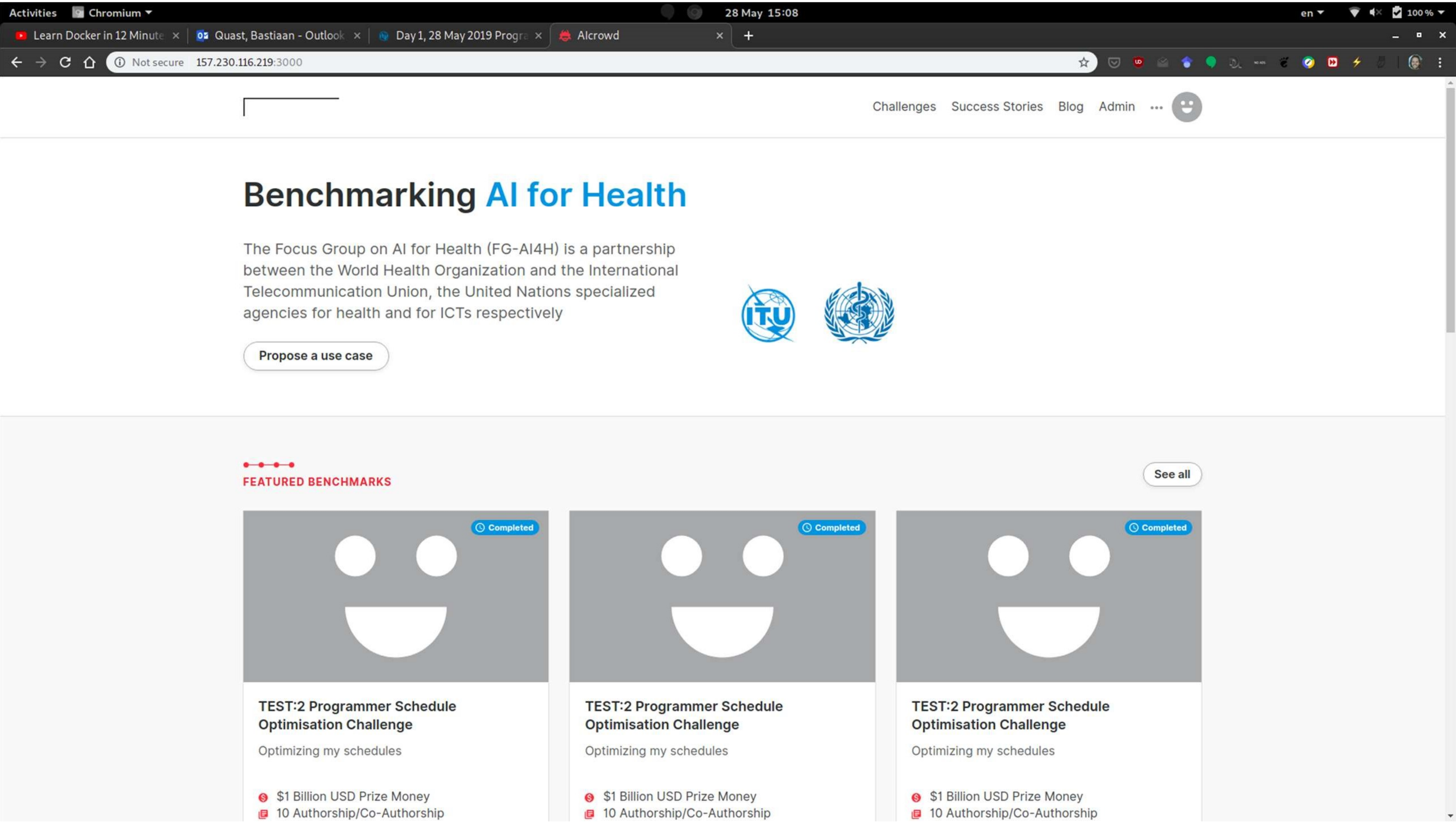
- For worldwide adoption, need evaluation standards
- AI heavily depends on the data used for training and testing
- Incorporation of domain (health/medicine) knowledge is essential



Quality control: AI solution

Quality indicators:

1. Performance measurement
2. Robustness
3. Uncertainty
4. Explainability
5. Generalizability
6. ...


Activities
Chromium
28 May 15:08
en
100 %
Learn Docker in 12 Minute
Quast, Bastiaan - Outlook
Day 1, 28 May 2019 Progra
AIcrowd
Not secure 157.230.116.219:3000
Challenges
Success Stories
Blog
Admin
Benchmarking AI for Health
The Focus Group on AI for Health (FG-AI4H) is a partnership between the World Health Organization and the International Telecommunication Union, the United Nations specialized agencies for health and for ICTs respectively
Propose a use case
FEATURED BENCHMARKS
See all
Completed
TEST:2 Programmer Schedule Optimisation Challenge
Optimizing my schedules
$1 Billion USD Prize Money
10 Authorship/Co-Authorship
Completed
TEST:2 Programmer Schedule Optimisation Challenge
Optimizing my schedules
$1 Billion USD Prize Money
10 Authorship/Co-Authorship
Completed
TEST:2 Programmer Schedule Optimisation Challenge
Optimizing my schedules
$1 Billion USD Prize Money
10 Authorship/Co-Authorship

1. Performance measurement

- Metrics:
  Selection of measures and implied consequences (consider how a data output can impact practice); metrics can be discrete (y/n), Euclidian metrics, Hamming distance, distance for segmentation tasks, Pearson correlation, etc., but needs to be specific for the problem. Multiple/combined metrics are concievable.
- Downstream consequences:
  Importance of metric accuracy for subsequent processes and risk of error propagation in the case of inaccuracy
- Qualities of AI assessment:
  Ease of comparison, validity, classification, and reproducibility
- …

2. Robustness

- Input data:
  Influence of data quality issues (e.g., noise, manipulation, and errors)
- Expectations:
  Can the AI algorithm complete a classification if the input signal has been disturbed (and experts are also able to make a classification)?
- Integration und guarantees:
  Robustness in training and testing? What limitations must be made to the input signal?
- Evaluation:
  How can robustness be tested?
- Unknown:
  How to handle data that we know exists but on which the AI has not been trained? How to handle data that we do not know exist?
- …

3. Uncertainty

- Measurement:
  How do we measure the safety of an AI? How do we measure the safety of the output of an AI algorithm?
- Quantification:
  How will experts assess the uncertainty of the reference data?
- Application:
  How will users evaluate the uncertainty of the AI algorithm?
- …

4. Explainability

- Identification:

  What type of explanation is necessary? Should it provide plausibility/verification, information about origins, and improvement of models?
- Reproducibility:
  Can the output of the AI algorithm be ascribed to an input signal?
- Explanation of relationships:
  Can a causality vs correlation be explained?
- Impact of data:
  Can patterns be detected in the input signal that explain the output of the AI algorithm?
- …

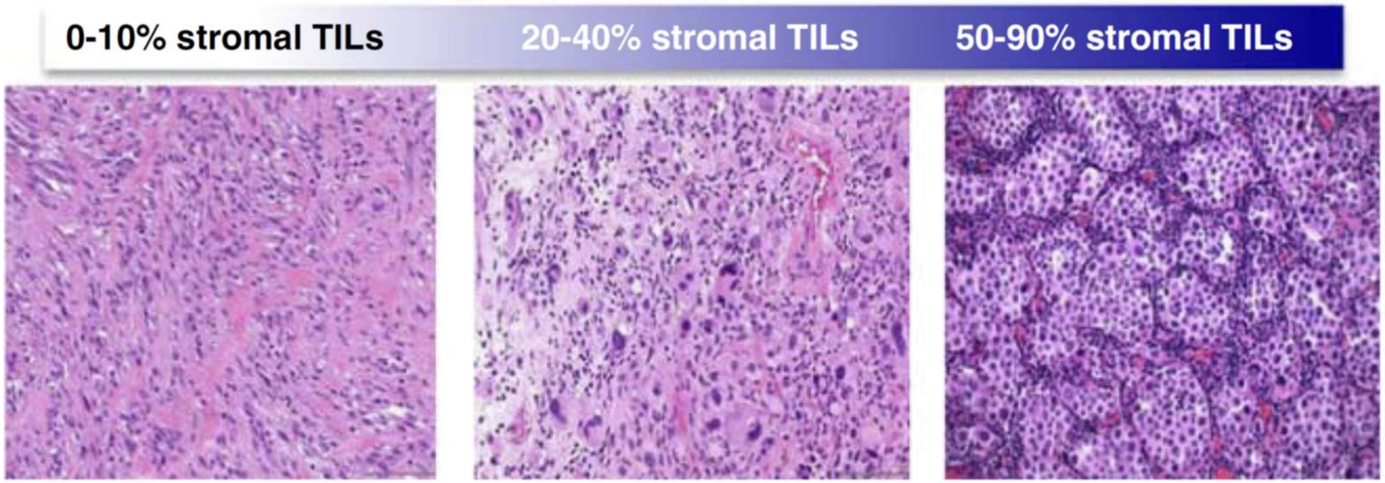
5. Generalizability

- Nature of the data:
  How can we test generalizability? What properties should the test set data have?
- Guarantees:
  Can the AI algorithm be used for the next datapoint of a test set?
- …

Our first proof-of-concept benchmark: Diagnostic Support for Breast Cancer

AI for Health
An ITU Focus Group
In collaboration with WHO

- Tumor infiltrating lymphocytes (TILs) are implicated in eliminating tumor cells
- Quantification of TILs relevant for patient prognosis estimation and therapy
- Replace "eye-balling" by pathologist with Machine Learning
- Focus Group: specify process on data generation and evaluate accuracy of Machine Learning method



Source: Hendry, S., Salgado, R., Gevaert, T., Russell, P. A., John, T., Thapa, B., ... & Sanders, M. (2017). Assessing Tumor-Infiltrating Lymphocytes in Solid Tumors: A Practical Review for Pathologists and Proposal for a Standardized Method from the International Immuno-Oncology Biomarkers Working Group Part 2 (...). Advances in anatomic pathology, 24(6), 311-335.

ITU/WHO Focus Group on Artificial Intelligence for Health (FG-AI4H)

Chair
- Thomas Wiegand, Fraunhofer HHI / TU Berlin, Germany

Vice-Chairs:
- Stephen Ibaraki, ACM, Canada
- Ramesh Krishnamurthy, World Health Organization
- Naomi Lee, The Lancet, United Kingdom
- Sameer Pujari, World Health Organization
- Shan Xu, CAICT, China

WG: “Regulatory Considerations”

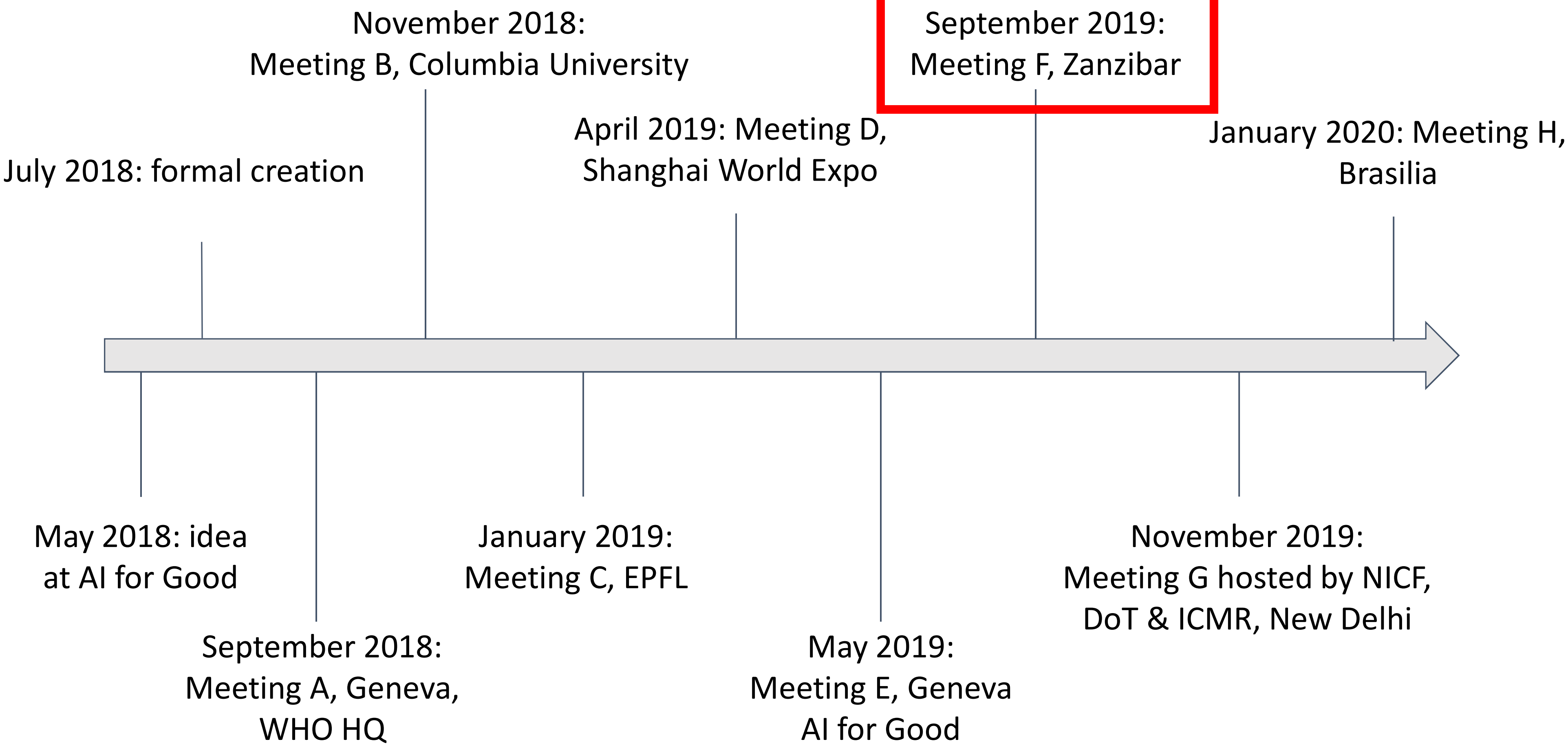
Chair

- Naomi Lee, The Lancet, United Kingdom

Vice-Chairs are representatives of:

- Khair ElZarrad, FDA, USA
- Paolo Alcini, EMA, Europe
- Peng Liang, HPMA, China
- Wolfgang Lauer, BfArM, Germany

Timeline


May 2018: idea at AI for Good
July 2018: formal creation
September 2018: Meeting A, Geneva, WHO HQ
November 2018: Meeting B, Columbia University
January 2019: Meeting C, EPFL
April 2019: Meeting D, Shanghai World Expo
May 2019: Meeting E, Geneva AI for Good
September 2019: Meeting F, Zanzibar
November 2019: Meeting G hosted by NICF, DoT & ICMR, New Delhi
January 2020: Meeting H, Brasilia

COMMENT | ONLINE FIRST

WHO and ITU establish benchmarking process for artificial intelligence in health

Thomas Wiegand • Ramesh Krishnamurthy • Monique Kuglitsch • Naomi Lee • Sameer Pujari • Marcel Salathé
et al. Show all authors

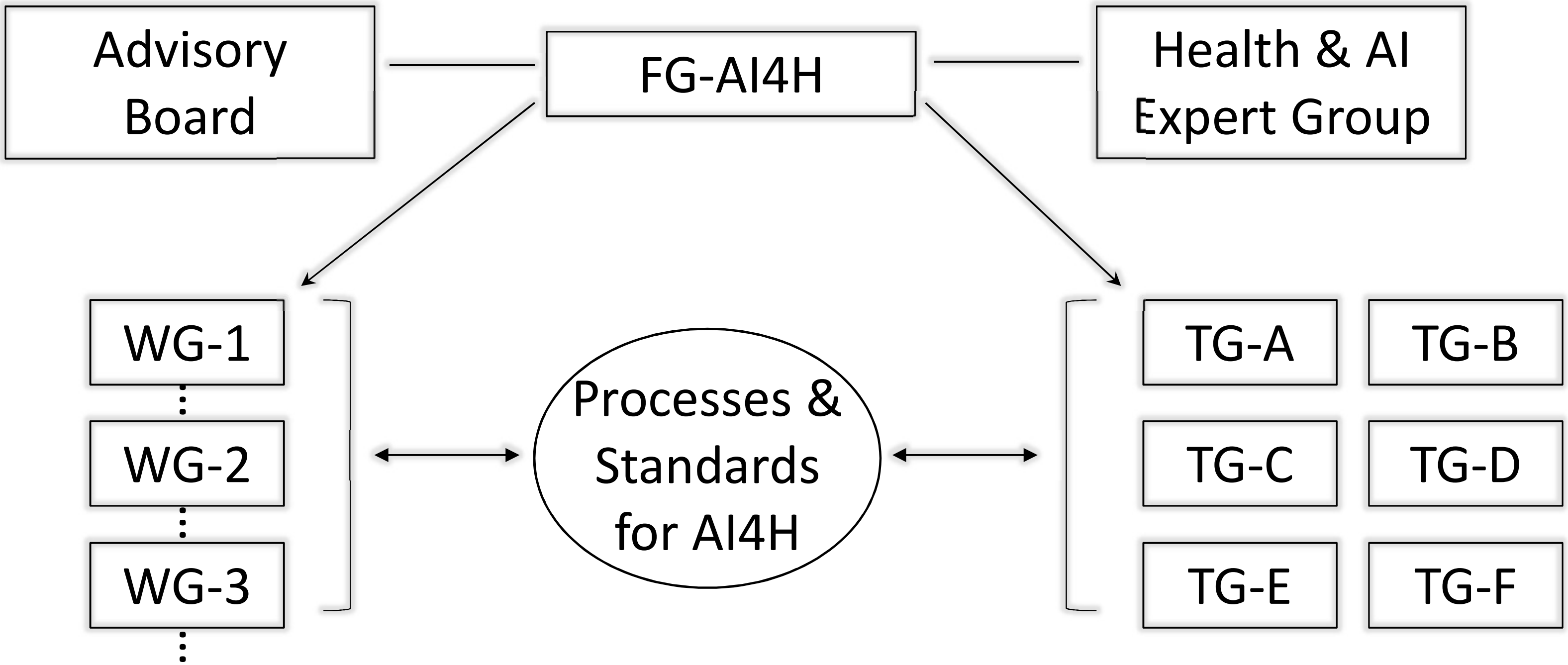
Published: March 29, 2019 • DOI: https://doi.org/10.1016/S0140-6736(19)30762-7 • Check for updates

Growing populations, demographic changes, and a shortage of health practitioners have placed pressures on the health-care sector. In parallel, increasing amounts of digital health data and information have become available. Artificial intelligence (AI) models that learn from these large datasets are in development and have the potential to assist with pattern recognition and classification problems in medicine—for example, early detection, diagnosis, and medical decision making.[1, 2] These advances promise to improve health care for patients and provide much-needed support for medical practitioners.

Stakeholders & *Cooperations*

- **WHO – World Health Organization**
- **ITU – International Telecommunication Union**
- *IANPHI – International Association of National Public Health Institutes*
- *Regulators (per country or via WHO)*
- *IAP – InterAcademy Partnership*
- *AI4Good – AI for Good Global Summit*
- *WHS – World Health Summit*
- *Philanthropic Foundations*

Structure of FG-AI4H



Topic Group Group Operation

A) Community: Creating and extending a community around a health topic

B) Proposals: Solicitation of specific AI4H proposals

C) Evaluation: Setting up evaluation criteria including reference data sets and metrics

D) Report: Publishing reports about the evaluation and the results

E) Dissemination: Deployment of AI for health solution in practice

Discussion of Process Steps

- The number of health topic communities will be very large
- Health topic discussions need to be moderated and chaired by impartial health/AI experts for fair and transparent process
- Monitoring and documenting by ITU or WHO official/staff
- Online-cooperation and virtual meetings
- Once an AI4H solution is deployed, data and results brought back and process repeated
- For AI4H solutions that learn/change during deployment, a specific benchmarking process should be developed
- The process (A-E) will also be evaluated every cycle for continuous improvement

Current Example Health Topic Groups

1. Cardiovascular disease risk prediction (TG-Cardio)
2. Dermatology (TG-Derma)
3. Falls among the elderly (TG-Falls)
4. Histopathology (TG-Histo)
5. Neuro-cognitive diseases (TG-Cogni)
6. Outbreak detection (TG-outbreaks)
7. Ophthalmology (TG-Ophthalmo)
8. Psychiatry (TG-Psy)
9. Radithereapy (TG-Radiotherapy)
10. Snakebite and snake identification (TG-Snake)
11. Symptom assessment (TG-Symptom)
12. Tuberculosis (TG-TB)
13. Volumetric chest computed tomography (TG-DiagnosticCT)

More Information:
ITU/WHO Focus Group on AI for Health

- Search: use “AI4H” as string
- Website: https://itu.int/go/fgai4h
- Next meetings:
  - 11-15 November 2019
    New Delhi, India
  - January 2020
    Brasilia, Brazil

AI for Health
An ITU Focus Group
In collaboration with WHO

Focus Group on "Artificial Intelligence for Health"

Focus Group on Environmental Efficiency for Artificial Intelligence and other Emerging Technologies

Focus Group on Artificial Intelligence for Health

Focus Group on Vehicular Multimedia

Focus Group on Technologies for Network 2030

Focus Group on Machine Learning for Future Networks including 5G

Focus Group on Application of Distributed Ledger Technology

Focus Group on Digital Currency including Digital Fiat Currency

Focus Group on Data Processing and Management

Concluded Focus Groups

ITU World Health Organization

FG-AI4H

The ITU-T Focus Group on artificial intelligence for health (AI4H) was established by ITU-T Study Group 16 at its meeting in Ljubljana, Slovenia, 9-20 July 2018. The Focus Group will work in partnership with the World Health Organization (WHO) to establish a standardized assessment framework for the evaluation of AI-based methods for health, diagnosis, triage or treatment decisions. Participation in the FG-AI4H is free of charge and open to all.

The scope and general process of the focus group are described in a commentary in The Lancet and a white paper. The documentation of all previous meetings can be found on the collaboration site (free ITU account needed).

Terms of reference > **Parent group >** ITU-T Study Group 16

Topic areas:

- Dermatology (TG-Derma)
- Falls among the elderly (TG-Falls)
- Histopathology (TG-Histo)

Meetings and Related Events | Focus Group News | Focus Group Videos

Geneva, Switzerland, 29 May - 1 June 2019

Breakthrough on artificial intelligence for health @ "AI for Good" Global Summit (29 May) and **5th meeting of FG-AI4H** (30 May - 1 June) (Announcement | Logistics)

Please register for both events below.

Breakthrough on AI4H (29 May)

- The workshop will be *part of* the "A.I. for Good" Global Summit 2019.
- Please register here Registration is *separate* from the FG meeting itself)

FG Meeting (30 May - 1 June)

- Register here (see instructions for help)
- Documents for this meeting
- Submit written proposals by e-mail to tsbfgai4h@itu.int before the deadline (22 May 2019 @ 23:59 CEST). [Use this template - Please do NOT submit as PDF]
- Remote participation via Zoom

Is that it? No, not even close…

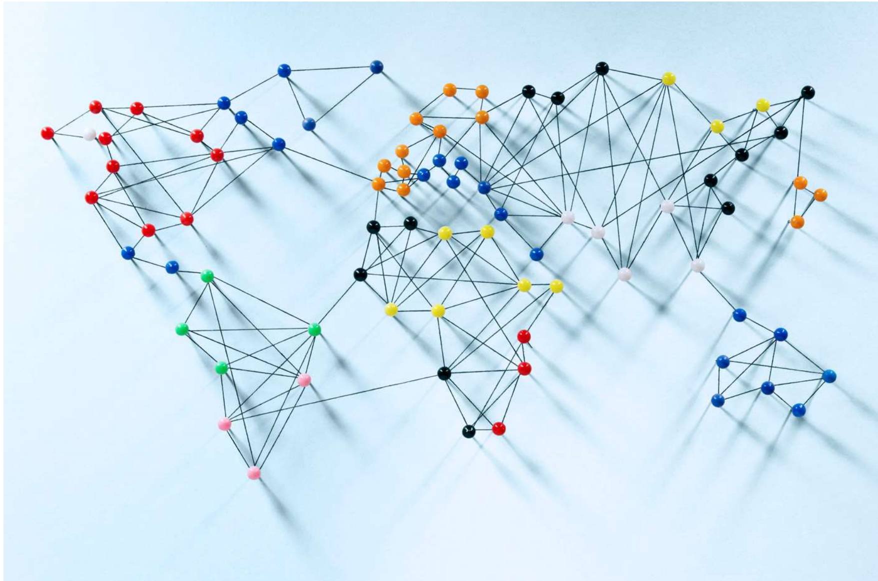
- What are the conditions of use for research health data?
  - Complicated, but solvable (UN-based servers, distributed locations, …)
  - When health app is free of charge, then research health data usage conditions maybe free
  - When health app bears license fee > 0, use of research health data should be charged
  - …

- What are the conditions of use for the health apps?
  - WHO/ITU process via FG-AI4H provides a component for quality control
  - Further quality control steps depend on health topic and country
  - Complicated, but solvable (royalty-free or other licensing models, updates, …)
  - …

- Who will provide the input to the work of FG-AI4H?

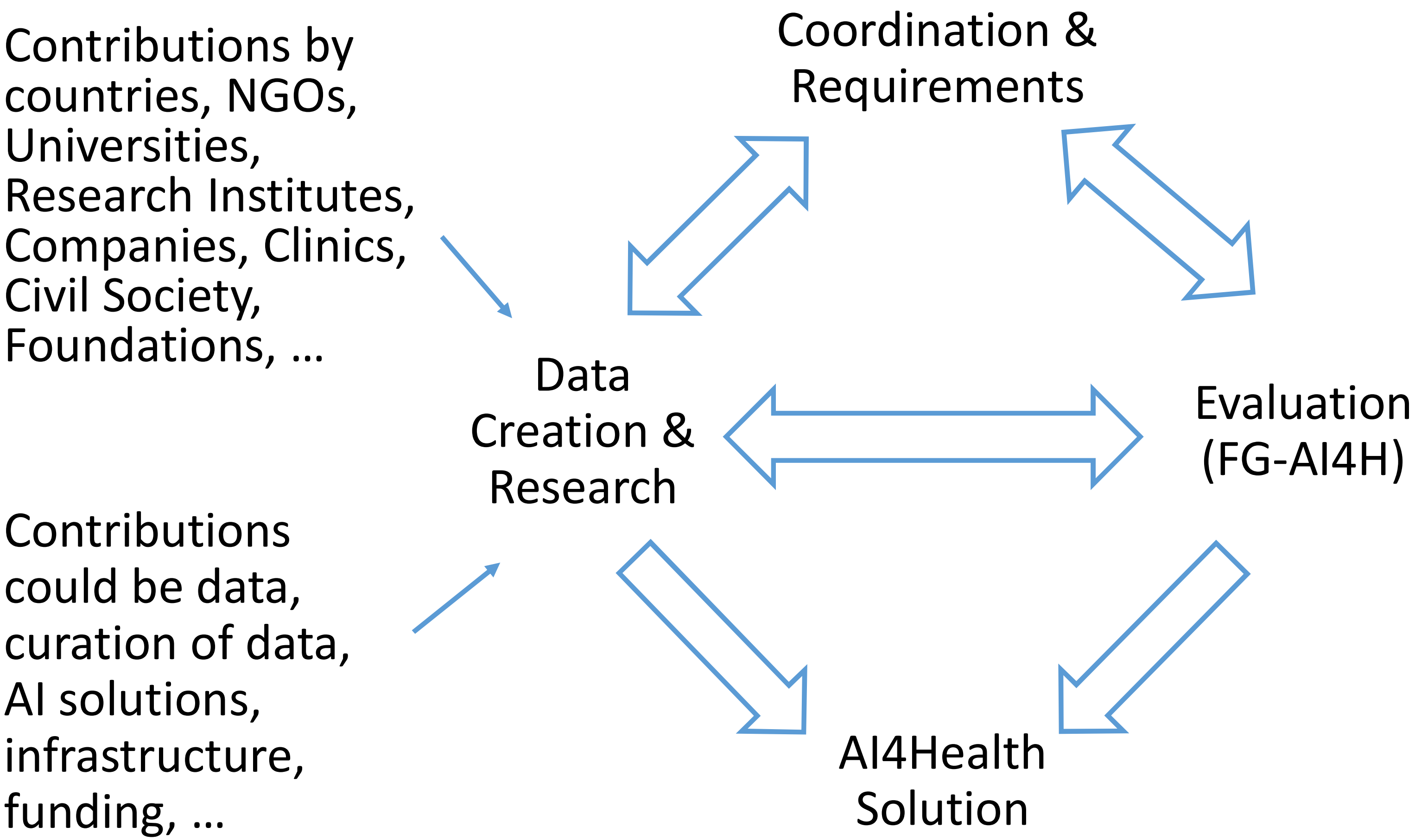
World-wide Network for Collaborative Research on AI for Health

- FG-AI4H creates the “pipeline” to bring AI4H research towards global adoption
- Linking and coordinating research together with the pipeline
- Encouraging national programs to support AI4H work: data collection and annotation, algorithm design and testing, …
- Discussions about funding have started



Source: http://www.baurome.com/global-network

Research and Evaluation Pipeline on AI for Health



Thank you very much for listening and participating!

Questions?