

FGAI4H-M-036-A01

E-meeting, 28-30 September 2021

Source: Markus Wenzel¹, David Neumann¹, Sandeep Reddy², Annika Reinke³,
Manuel Wiesenfarth³, Alberto Merola⁴, Steffen Vogler⁵
¹Fraunhofer HHI, ²Deakin University, ³DKFZ, ⁴AICURA, ⁵Bayer

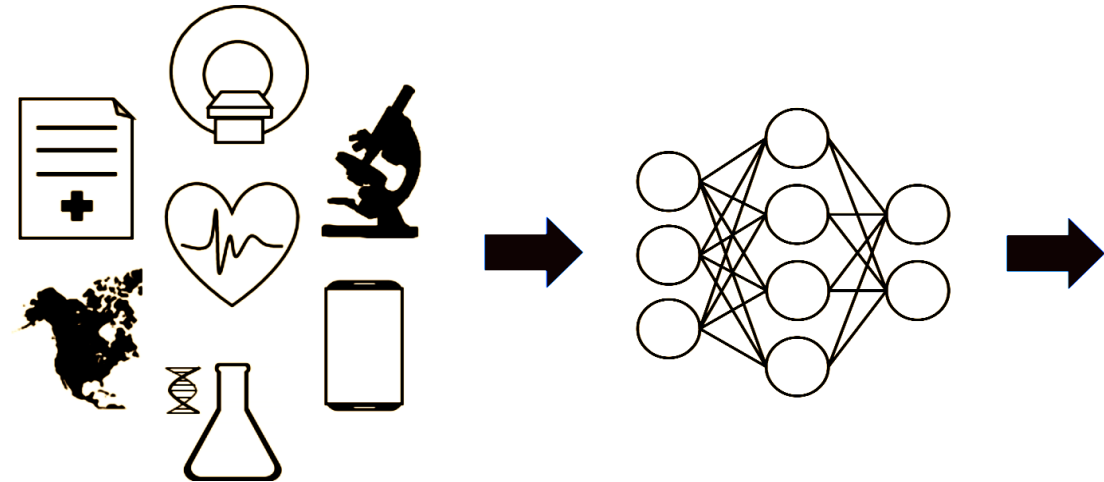
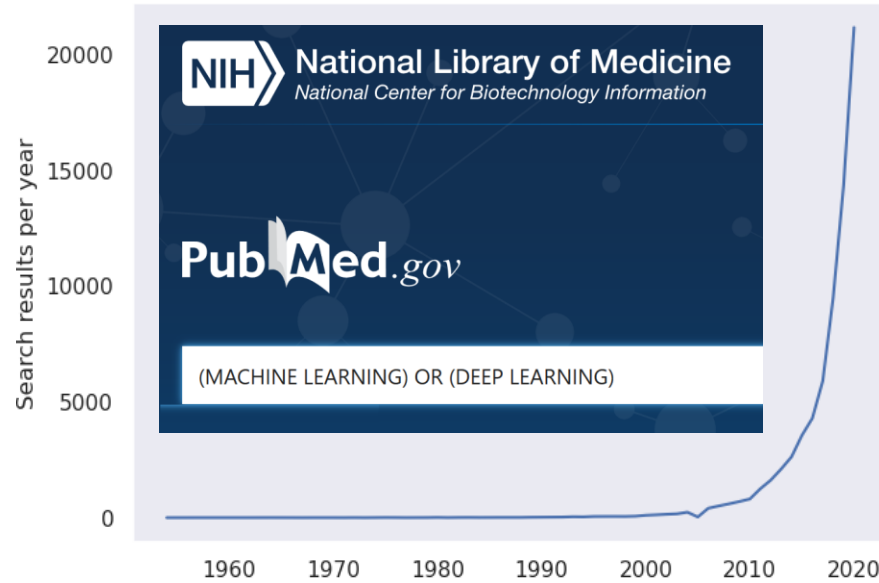
Title: Updated DEL07: AI for Health Evaluation Considerations
- Att.1 - Presentation

Purpose: Discussion

Contact: Markus Wenzel E-mail: markus.wenzel@hhi.fraunhofer.de


Abstract: This PPT summarizes the content of M-036 with the *Updated DEL07: AI for Health Evaluation Considerations*, for presentation and discussion during the meeting.

Healthcare AI Models




- Offer great potential for many applications in healthcare.
- Predict, recognize, recommend, warn early, help to decide, ...

Healthcare AI Models

- Effective, safe, accurate (on new data), robust, plausible, fair?
- Standards for trustworthy healthcare AI needed
- Careful evaluation plays crucial role 

Evaluation Deliverables Series



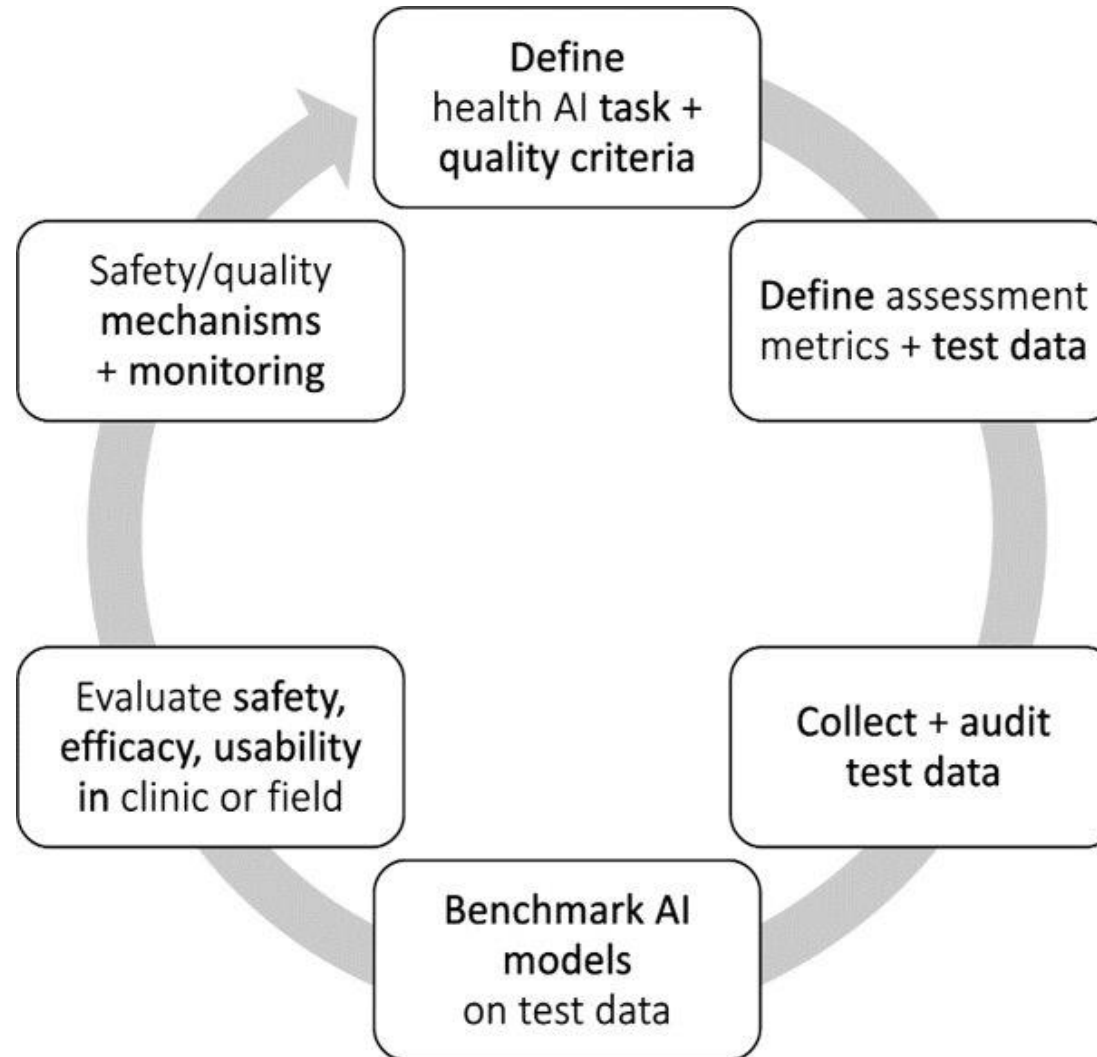
N°	Title	Editors
DEL7.0	AI4H Evaluation Considerations	Markus Wenzel (Fhg)
DEL7.1	AI4H Evaluation Process Description	Sheng Wu (WHO)
DEL7.2	AI Technical Test Specification	Auss Abood (RKI)
DEL7.3	Data+AI Assessment Methods Ref.	Luis Oala (Fhg)
DEL7.4	Clinical Evaluation	Naomi Lee (Lancet) et al.
DEL7.5	Assessment Platform	Steffen Vogler (Bayer) et al.

DEL7.0) AI4H Evaluation Considerations

1. Overview of deliverables DEL7.1-5
2. Preliminary evaluation process
3. Novelty
4. Standardized model benchmarking
5. Benchmarking platform requirements
6. Best practices from literature

Cyclical Evaluation Process

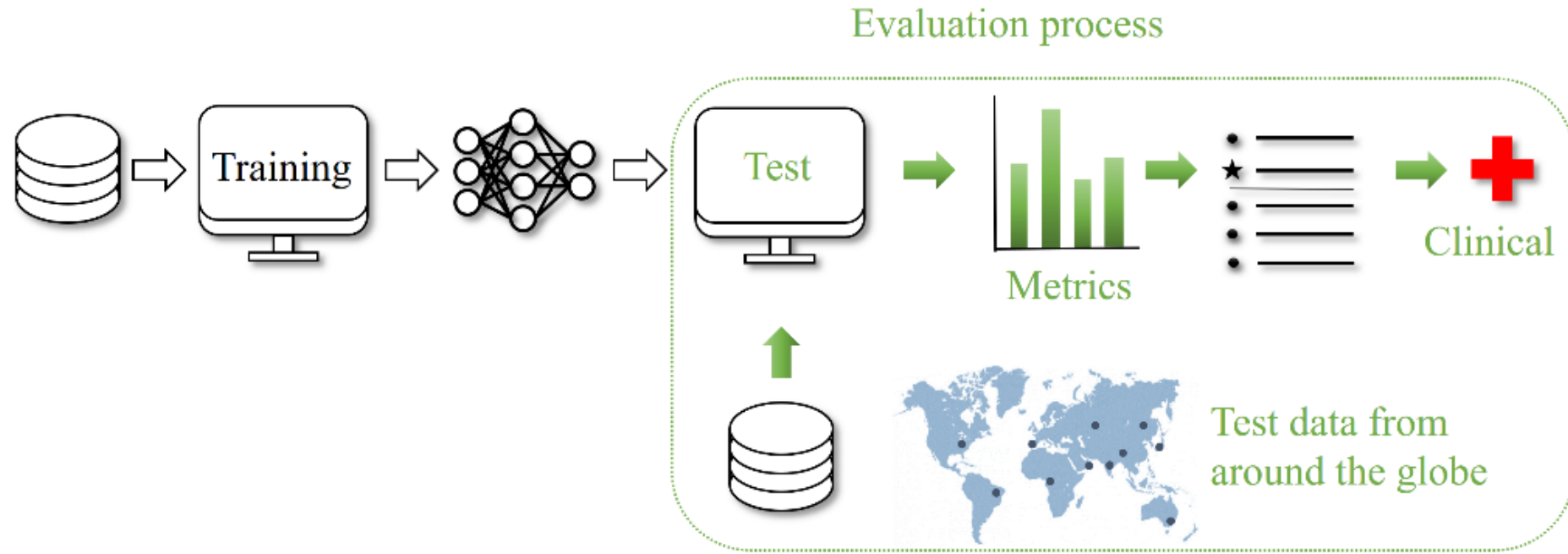
Preliminary considerations



Novel Aspects of Health AI Model Evaluation

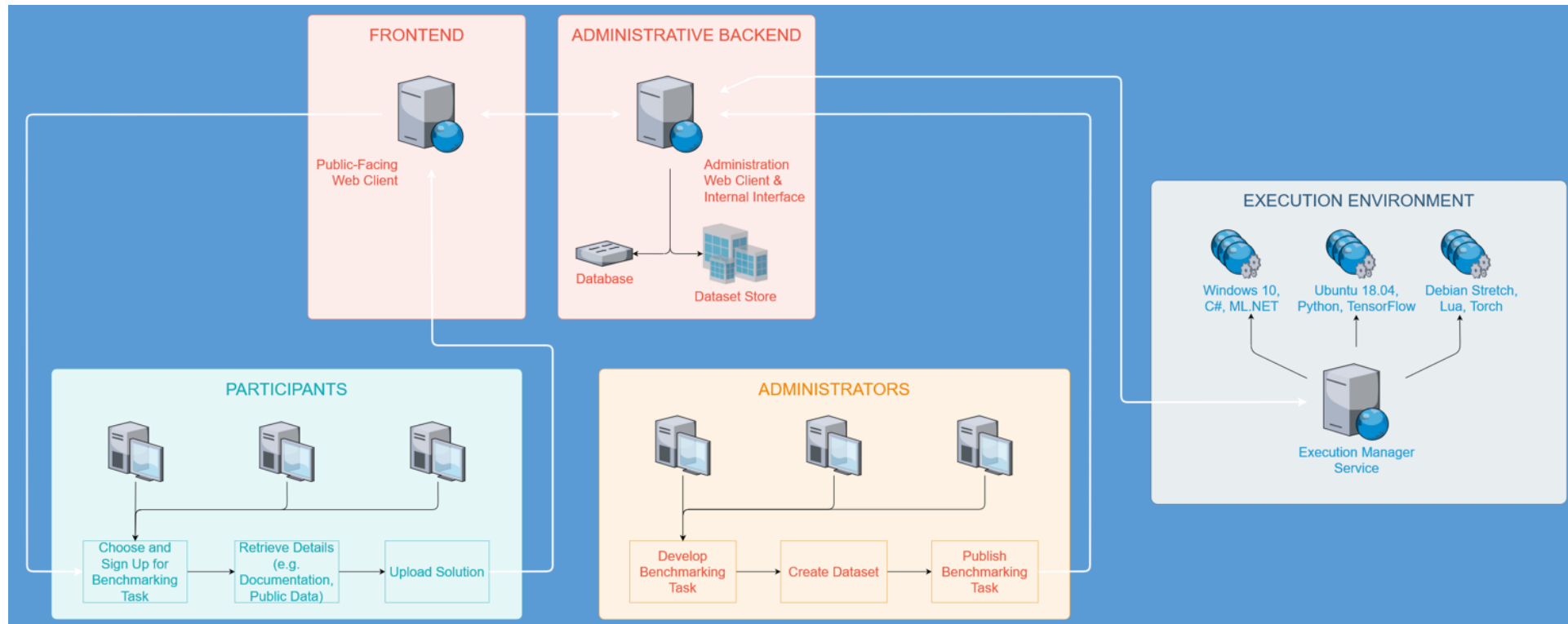
N°	Aspect (key word; full description in DEL7)	Phase of Evaluation Cycle					
		Define task	Define task metrics	Collect/ audit test data	Benchmark	Evaluate	Monitor
1	Appropriate test data sets	●	●	●	●		●
2	Internal vs. external validation			●	●		
3	Technical vs. clinical criteria				●	●	
4	Proper benchmarking				●		
5	Complex models and settings	●	●				
6	Self-learning algorithms				●		
7	Human factors					●	●
8	Misuse accurate tools					●	●
9	Design of clinical trials with AI					●	
10	Agreement on/interoperable encoding of benchmarking cases	●	●		●		
11	Testing existing AI products vs. ML-challenges	●			●		

Standardized Benchmarking



- Independent model validation with standard procedures on separate test data and subsequent clinical evaluation
- (1) Closed environment, (2) Via interface , (3) Federated

Benchmarking Platform – Requirements



- System overview (administrative backend · frontend · execution environment)
- General considerations (security · hosting · computing resources · availability)

Collection of Publications With Best Practices



- Scientific literature and other documents
- From 29+ important sources with relevant quotes & brief discussion
- Please point us to literature you find relevant!

Status Update ↻

- Added references & characteristics based on meeting L feedback, improved text
- Turning vision for AI4H benchmarking platform into practice: <http://github.com/fg-ai4h> & <https://ai4h-audit.org/>
- See talk yesterday: evaluation package of open code initiative

Home All Benchmarks Documentation Sign Up Log In

AI:+ **AI for Health**
ITU-WHO Focus Group

Evaluating AI models in health

The ITU/WHO Focus Group on artificial intelligence for health (FG-AI4H) works 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. The group was established by ITU-T Study Group 16 at its meeting in Ljubljana, Slovenia, 9-20 July 2018.

Host Benchmark Participate

Next →

- Looking for contributors! You? DEL editors, WGs, TGs, OCI?
- markus.wenzel@hhi.fraunhofer.de
- <https://itu.int/go/fgai4h/collab> → Deliverables → DEL07.docx
- Progress of DEL7.1-5
- Many thanks to my co-authors:
 - David Neumann (Fraunhofer HHI)
 - Sandeep Reddy (Deakin University)
 - Annika Reinke, Manuel Wiesenfarth (DKFZ)
 - Alberto Merola (AICURA medical)
 - Steffen Vogler (Bayer)