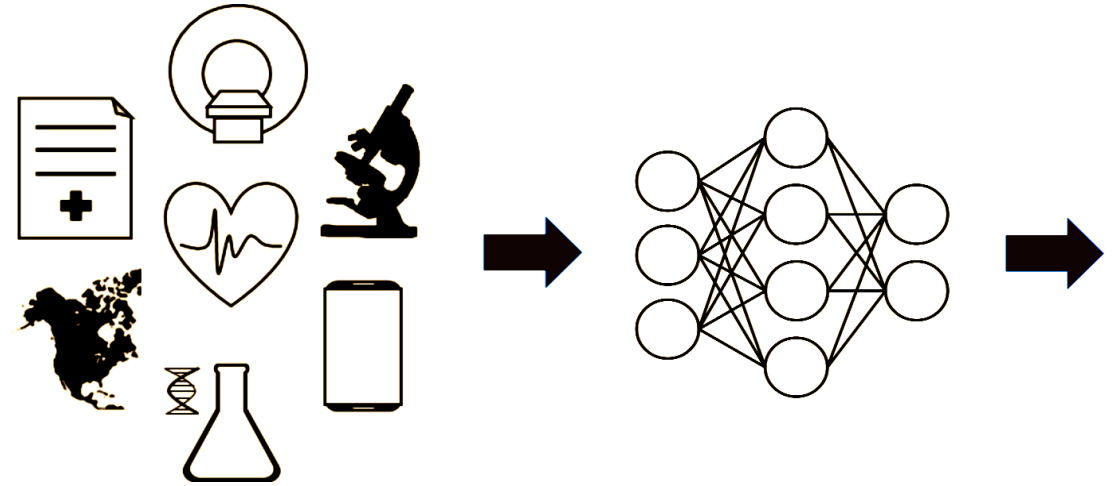
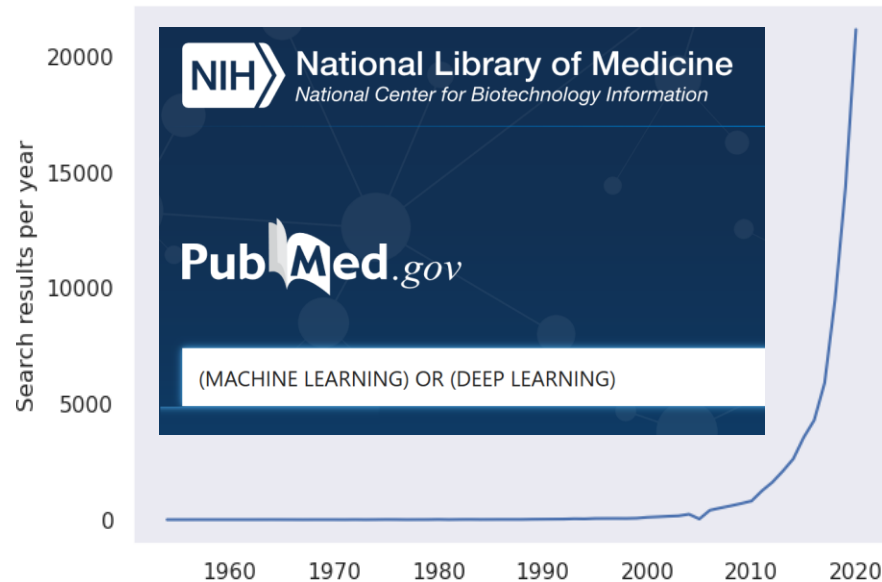


**FGAI4H-N-042-A01**

E-meeting, 15-17 February 2022


<b>Source:</b>	Markus Wenzel <sup>1</sup> , David Neumann <sup>1</sup> , Sandeep Reddy <sup>2</sup> , Annika Reinke <sup>3</sup> , Manuel Wiesenfarth <sup>3</sup> , Alberto Merola <sup>4</sup> , Steffen Vogler <sup>5</sup> <sup>1</sup> Fraunhofer HHI, <sup>2</sup> Deakin University, <sup>3</sup> DKFZ, <sup>4</sup> AICURA, <sup>5</sup> Bayer	
<b>Title:</b>	Updated DEL07: AI for Health Evaluation Considerations - Att.1 - Presentation	
<b>Purpose:</b>	Discussion	
<b>Contact:</b>	Markus Wenzel	E-mail: <a href="mailto:markus.wenzel@hhi.fraunhofer.de">markus.wenzel@hhi.fraunhofer.de</a>
<b>Abstract:</b>	This PPT summarizes the content of N-042 with the <i>Updated DEL07: AI for Health Evaluation Considerations</i> , 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, ...
- Effective, safe, accurate on new data, robust, plausible, fair?
- Standards for trustworthy healthcare AI needed
- Careful evaluation plays crucial role

# Evaluation Deliverables Series



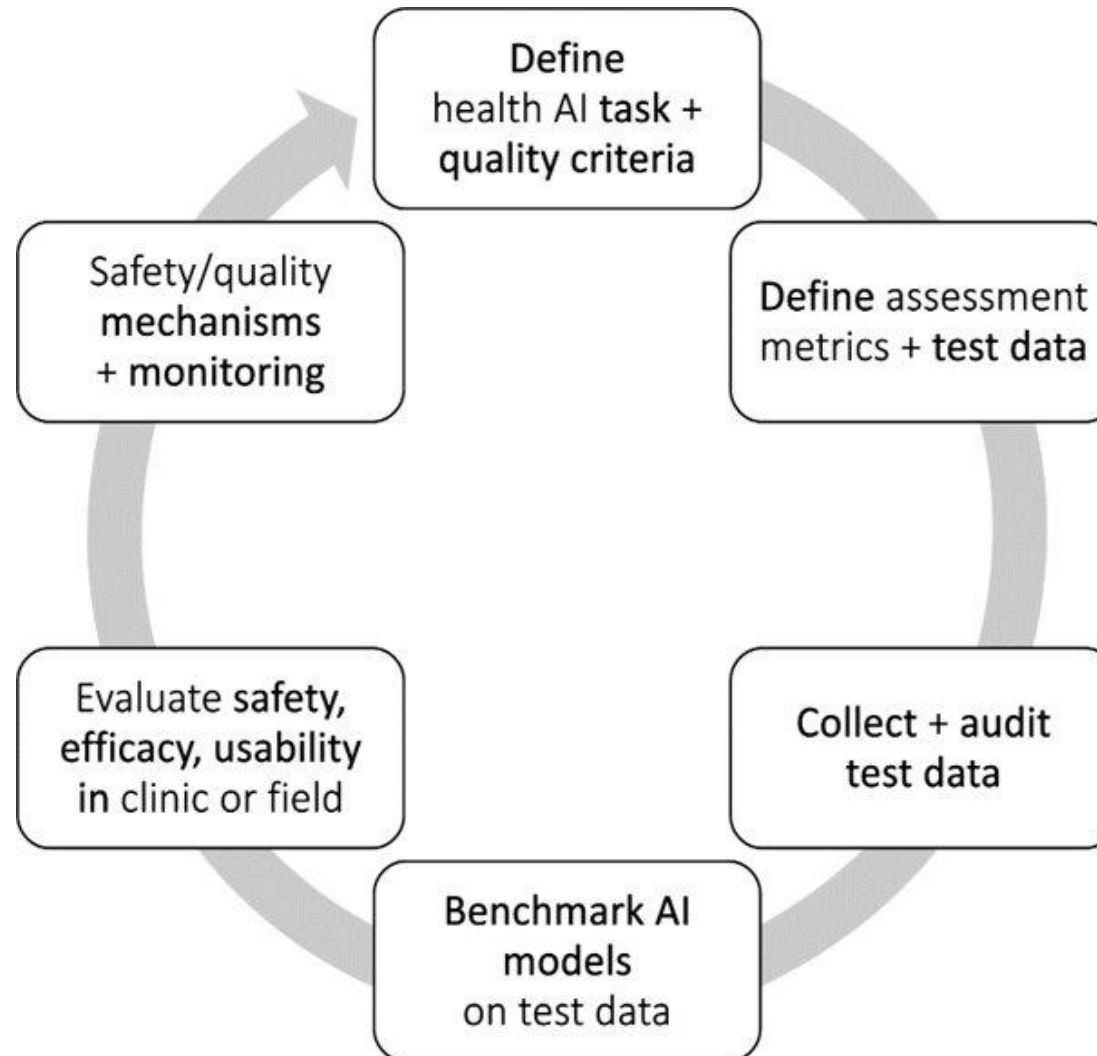
N°	Title	Editors
DEL7.0	AI4H Evaluation Considerations	Markus Wenzel (Fhg)
DEL7.1	AI4H Evaluation Process Description	Sheng Wu (WHO)
DEL7.2	AI Technical Test Specification	Auss Abood (RKI)
DEL7.3	Data + AI Assessment Methods Ref.	Luis Oala (Fhg)
DEL7.4	Clinical Evaluation	Naomi Lee (Lancet) et al.
DEL7.5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

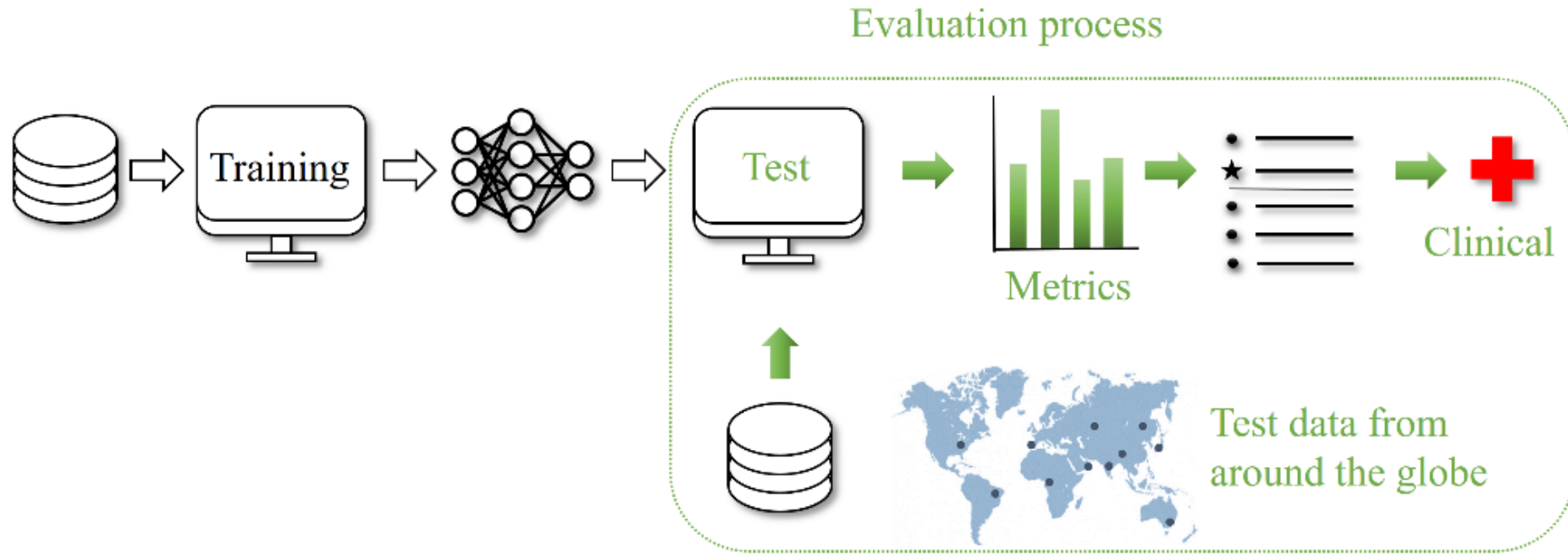
Preliminary considerations



# Novel Aspects of Health AI Model Evaluation

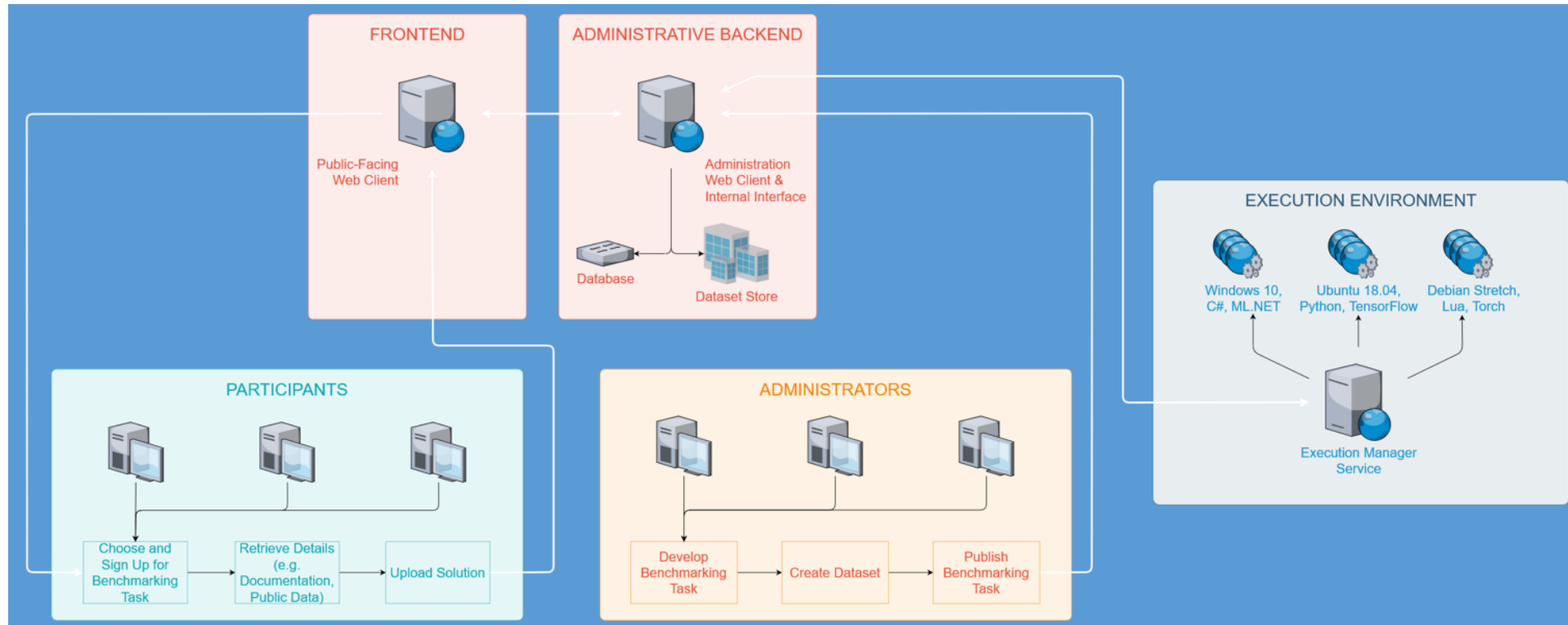
N°	Aspect (key word; full description in DEL7)	Phase of Evaluation Cycle					
		Define task	Define task metrics	Collect/ audit test data	Benchmark	Evaluate	Monitor
1	Appropriate test data sets	●	●	●	●		●
2	Internal vs. external validation			●	●		
3	Technical vs. clinical criteria				●	●	
4	Proper benchmarking				●		
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 31+ important sources with relevant quotes & brief discussion
- Please point us to literature you find relevant!

# Status Update

- Added references & improved text
- Turning vision for AI4H benchmarking platform into practice: <http://github.com/fg-ai4h> & <https://health.aiaudit.org/>
- See talk on day 1: evaluation package of open code initiative



# Outline next steps to publish deliverable DEL7

- In good shape, quite happy about current version
- Review by editors of DEL7.1-5, WGs, TGs, OCI, AI/ML4H-Audit would be valuable
- Depends as intro to series on the progress of DEL7.1-5
  - DEL7.1: AI4H evaluation process description?
  - Fluid, dynamic DEL7.3. Now: "DEL7.3: ML4H trial audits–Iteration 2.0 Playbook", ex: "DEL7.3: Data and artificial intelligence assessment methods (DAISAM) reference"?
  - DEL7.5: Assessment platform?
- Detail: Colorful title page would be nice
- Many thanks to my co-authors: David Neumann (Fraunhofer), Sandeep Reddy (Deakin Univ.), Annika Reinke, Manuel Wiesenfarth (DKFZ), Alberto Merola (AICURA medical), Steffen Vogler (Bayer)
- [markus.wenzel@hhi.fraunhofer.de](mailto:markus.wenzel@hhi.fraunhofer.de)