

ESC Guidelines for AI driven Software

Anja Hennemuth

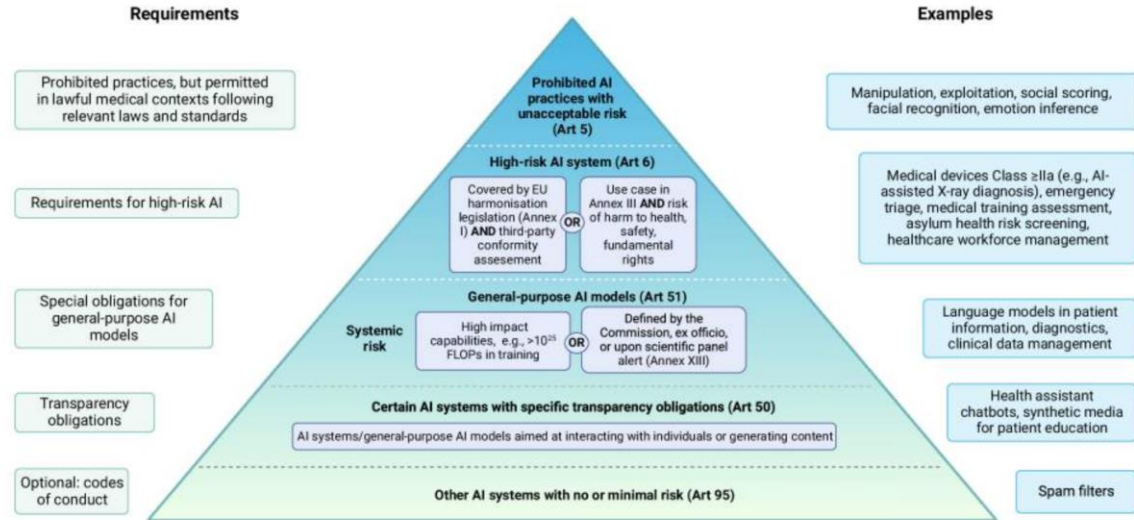
14.11.2024

Why more guidelines?

- **Existing regulations and guidelines**
 - EU AI Act
 - FUTURE-AI
 - Guideline papers
 - Reporting guidelines

AI Act – Key Requirements

- **AI Act provides**
 - Risk-based classification
 - Requirements for **High-Risk AI Systems** such as **Digital Medical Products**
- **For the application in medical context health policy experts suggest**
- *“collaboration between AI developers, healthcare professionals, patient communities, and regulators to develop specific guidelines and standards for AI in healthcare, ensuring that all health-related AI is monitored”*



Busch F, Kather JN, Johner C, Moser M, Truhn D, Adams LC, Bressen KK. Navigating the European Union Artificial Intelligence Act for Healthcare. NPJ Digit Med. 2024 Aug 12;7(1):210. doi: 10.1038/s41746-024-01213-6. PMID: 39134637; PMCID: PMC11319791.

Aboy M, Minssen T, Vayena E. Navigating the EU AI Act: implications for regulated digital medical products. NPJ Digit Med. 2024 Sep 6;7(1):237. doi: 10.1038/s41746-024-01232-3. PMID: 39242831; PMCID: PMC11379845.

van Kolschooten H, van Oirschot J. The EU Artificial Intelligence Act (2024): Implications for healthcare. Health Policy. 2024 Nov;149:105152. doi: 10.1016/j.healthpol.2024.105152. Epub 2024 Sep 7. PMID: 39244818.

Best Practice and Related Guidelines

“Governments from all countries must cooperatively lead efforts to effectively regulate the development and use of AI technologies, such as LMMs,” Dr Alain Labrique, WHO Director for Digital Health and Innovation in the Science Division

• WHO

- 2024 Update of ***Ethics and governance of artificial intelligence for health: Guidance on large multi-modal models***
 - The new WHO guidance outlines five broad applications of LMMs for health:
 - **Diagnosis and clinical care**, such as responding to patients’ written queries;
 - **Patient-guided use**, such as for investigating symptoms and treatment;
 - **Clerical and administrative tasks**, such as documenting and summarizing patient visits within electronic health records;
 - **Medical and nursing education**, including providing trainees with simulated patient encounters, and;
 - **Scientific research and drug development**, including to identify new compounds

WHO Key Recommendations for LLMs

- Governments have the primary responsibility to set standards for the development and deployment of LLMs, and their integration and use for public health and medical purposes.
 - Invest in or provide not-for-profit or public infrastructure, including computing power and **public data sets**, accessible to developers in the public, private and not-for-profit sectors, that requires users to adhere to ethical principles and values in exchange for access.
 - Use laws, policies and regulations to ensure that LLMs and applications used in health care and medicine, irrespective of the risk or benefit associated with the AI technology, **meet ethical obligations and human rights standards** that affect, for example, a person's dignity, autonomy or privacy.
 - Assign an existing or new regulatory agency to assess and approve LLMs and applications intended for use in health care or medicine – as resources permit.
 - Introduce mandatory post-release **auditing and impact assessments**, including for data protection and human rights, by independent third parties when an LLM is deployed on a large scale. The auditing and impact assessments should be published and should include outcomes and impacts disaggregated by the type of user, including for example by age, race or disability.
- Developers of LLMs should ensure that:
 - LLMs are designed not only by scientists and engineers. Potential users and **all direct and indirect stakeholders, including medical providers, scientific researchers, health care professionals and patients, should be engaged from the early stages of AI development in structured, inclusive, transparent design and given opportunities to raise ethical issues, voice concerns and provide input for the AI application under consideration.**
 - LLMs are designed to perform well-defined tasks with the necessary accuracy and reliability to improve the capacity of health systems and advance patient interests. Developers should also be able to predict and understand potential secondary outcomes.

EU Initiatives – Future AI

- **FUTURE-AI: International consensus guideline for trustworthy and deployable artificial intelligence in healthcare**
 - Focus on trustworthiness of AI
 - 117 interdisciplinary experts from 50 countries representing all continents, including AI scientists, clinical researchers, biomedical ethicists, and social scientists
 - Also addresses interaction with AI
 - Requirement of evaluation on real world data

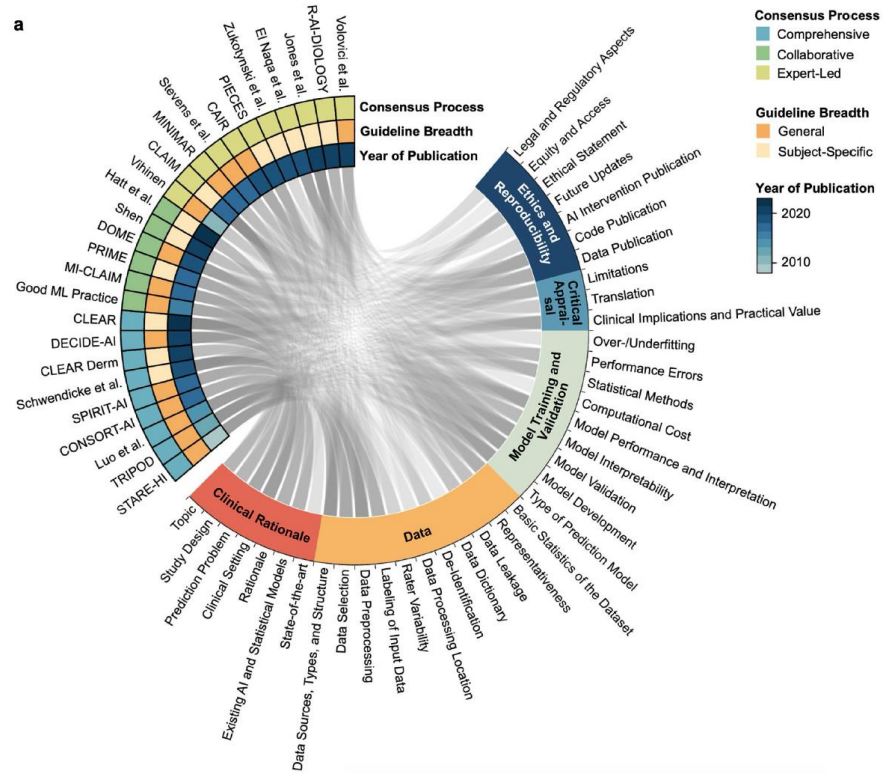
FUTURE-AI: Best practices for trustworthy AI in medicine

FUTURE-AI is an international, multi-stakeholder initiative for defining and maintaining concrete guidelines that will facilitate the design, development, validation and deployment of trustworthy AI solutions in medicine and healthcare based on six guiding principles: Fairness, Universality, Traceability, Usability, Robustness and Explainability.



Reporting Guidelines

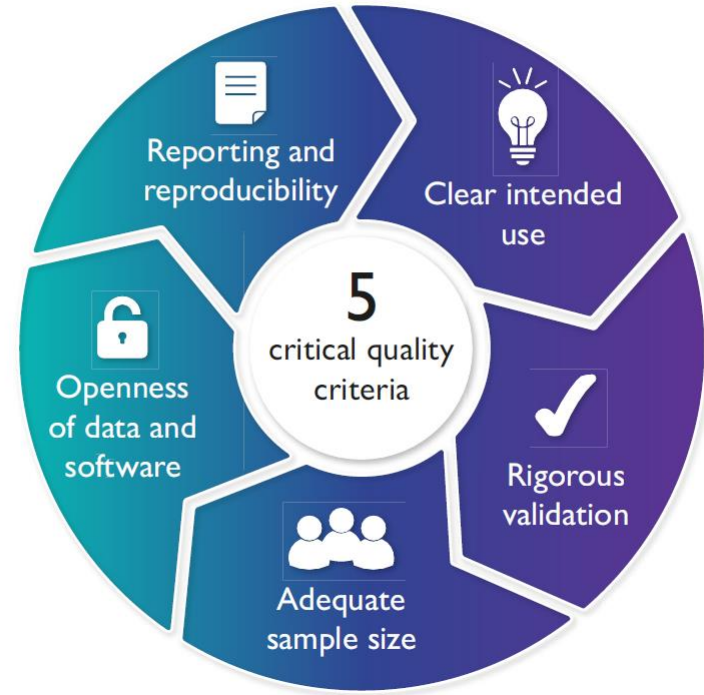
- Overview paper from 2024 found 26 guidelines for reporting of AI in medicine
 - Important suggestions for reporting of
 - Clinical Rationale
 - Data
 - Model Training and Validation
 - Ethics and Reproducibility



Quality Criteria by EHJ

Quality criteria for AI-based prediction models

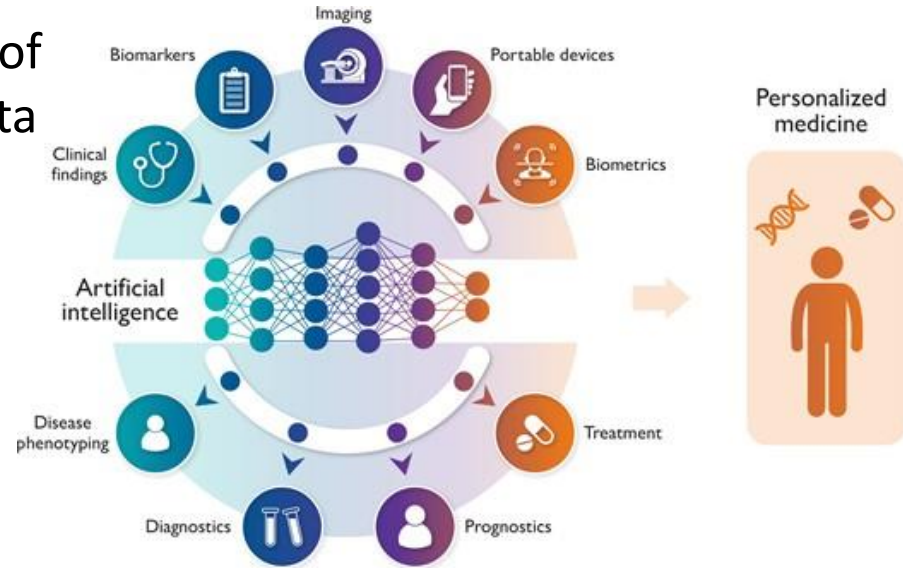
- **Editors of the European Heart Journal Digital Health have suggested 5 critical quality criteria for AI prediction models**
- **Minimal requirements**
 - Following reporting guidelines
 - Aim of the model stated clearly
 - Internal validation of the AI-based prediction models
 - A sample size for development that is substantially larger than needed for a regression-based prediction model
 - Providing contact details for data and algorithm accessibility requests



van Royen FS, Asselbergs FW, Alfonso F, Vardas P, van Smeden M. Five critical quality criteria for artificial intelligence-based prediction models. *Eur Heart J.* 2023 Dec 7;44(46):4831-4834. doi: 10.1093/eurheartj/ehad727. PMID: 37897346; PMCID: PMC10702458.

Why should ESC care?

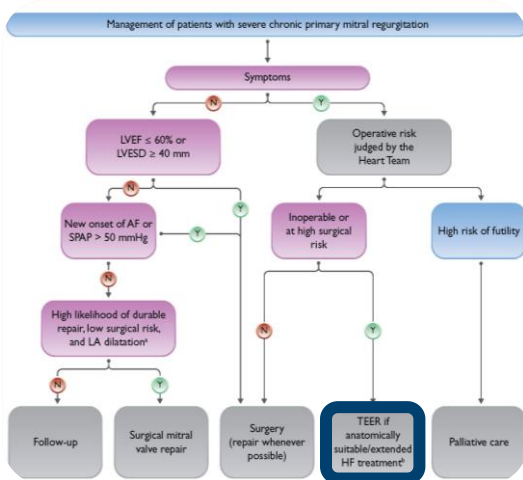
- **Specific requirements in cardiology**
 - Acquisition and interpretation of complex imaging and signal data
 - Sensitive information (photo, speech, ...)
 - High potential in usage of wearables and patient empowerment



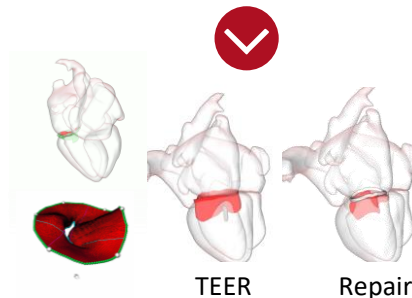
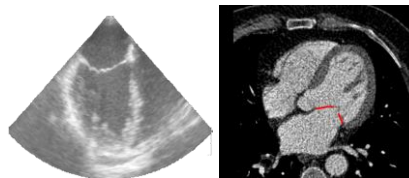
Relevance for ESC Guidelines

Integration of Decision Making

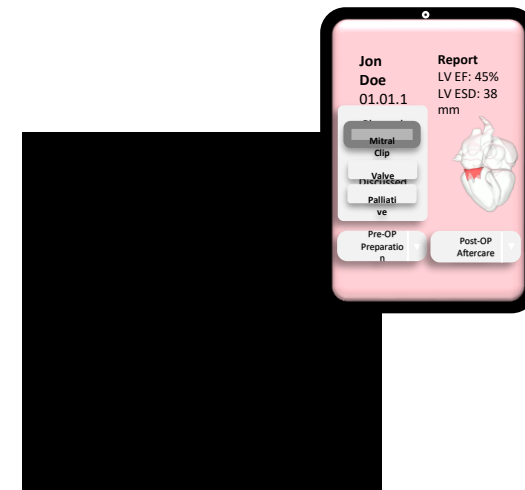
Guidelines



Individual prediction with Digital Twin

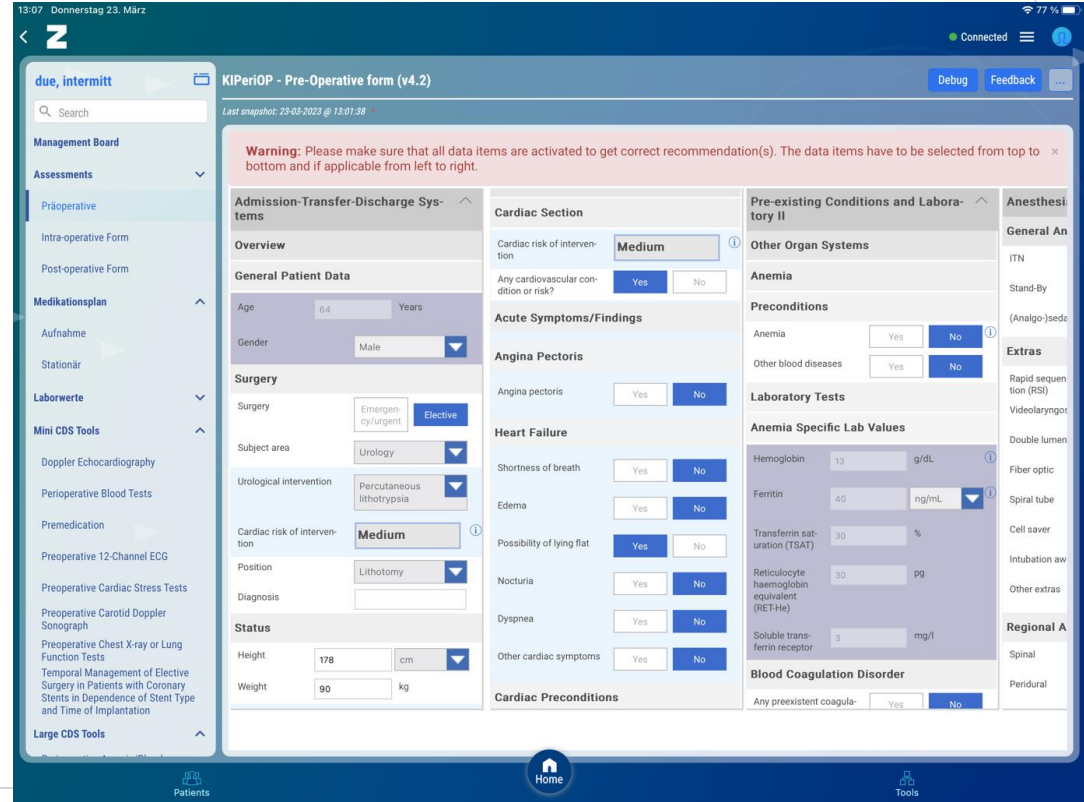


Patient participation



Relevance for ESC Guidelines

- 2022 ESC Guidelines on cardiovascular assessment and management of patients undergoing non cardiac surgery
- Guideline-based decision support tools provide structured representation of clinical information
- Combination with outcome information enables AI training



13:07 Donnerstag 23. März

due, intermitt

KIPeriOP - Pre-Operative form (v4.2)

Warning: Please make sure that all data items are activated to get correct recommendation(s). The data items have to be selected from top to bottom and if applicable from left to right.

Admission-Transfer-Discharge Systems

Overview

General Patient Data

Age: 64 Years

Gender: Male

Surgery

Surgery: Emergency/urgent, Elective

Subject area: Urology

Urological intervention: Percutaneous lithotripsy

Cardiac risk of intervention: Medium

Position: Lithotomy

Diagnosis:

Status

Height: 178 cm

Weight: 90 kg

Cardiac Section

Cardiac risk of intervention: Medium

Any cardiovascular condition or risk? Yes No

Acute Symptoms/Findings

Angina Pectoris

Angina pectoris: Yes No

Heart Failure

Shortness of breath: Yes No

Edema: Yes No

Possibility of lying flat: Yes No

Nocturia: Yes No

Dyspnea: Yes No

Other cardiac symptoms: Yes No

Cardiac Preconditions

Pre-existing Conditions and Laboratory II

Other Organ Systems

Anemia: Yes No

Other blood diseases: Yes No

Laboratory Tests

Anemia Specific Lab Values

Hemoglobin: 13 g/dL

Ferritin: 40 ng/mL

Transferrin saturation (TSAT): 30 %

Reticulocyte haemoglobin equivalent (RETHe): 30 pg

Soluble transferrin receptor: 3 mg/l

Blood Coagulation Disorder

Any preexistent coagulation disorder: Yes No

Patients

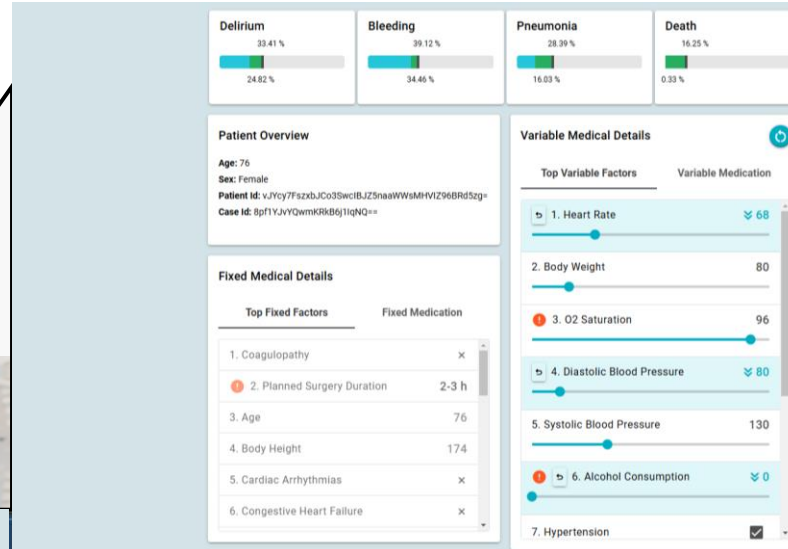
Home

Tools

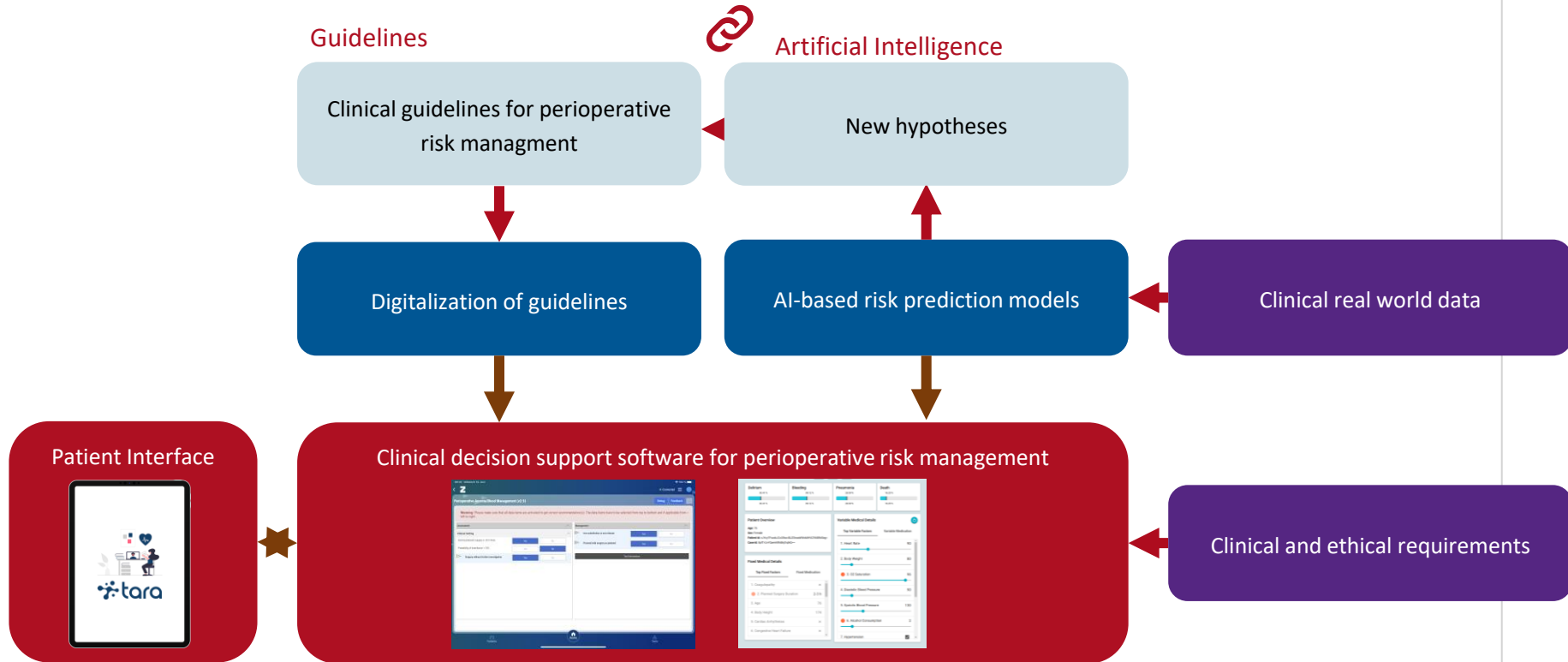
Interfaces | Peri-operative Risk minimization

User interface for AI prediction
Visualization of interface of modifiable risk factors (weight, medication, ...)

Model cards for transparency on model training and performance



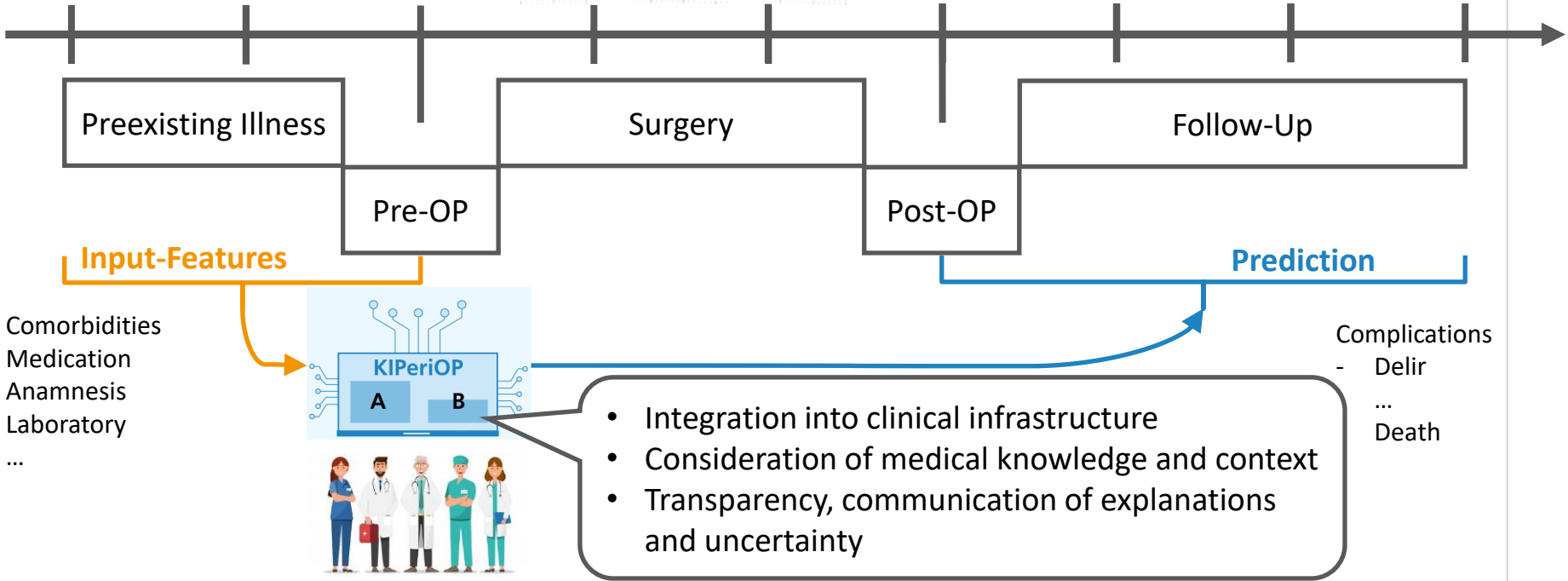
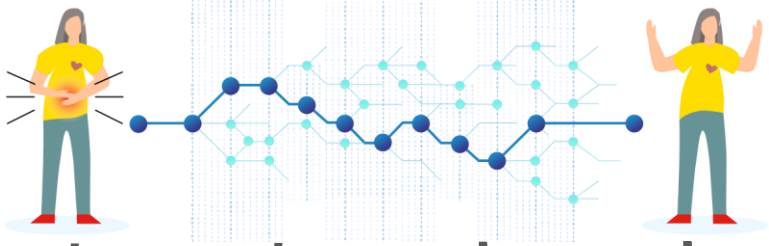
Integration with Clinical Interfaces



What to do?

- **Task force**
 - Clinical experts from different disciplines
 - EU initiatives
 - Software providers
 - AI specialists
- **Collaboration on training and benchmarking**
 - Define requirements and support benchmarking initiatives
- **Collaboration on interface requirements**
 - Integration of user perspective
 - Clinicians
 - Patients
 - Care Experts
- **Define interpretation of terminology**
 - E.g. explainability vs interpretability
 - Data processing levels, ...
- **Emphasis on monitoring and critical diagnostic tasks**
 - E.g. emergency echo, ECG, ...

Questions?



AI Based Approach to Peri-operative Risk Minimization

Model training with routine data of 30.000 patients

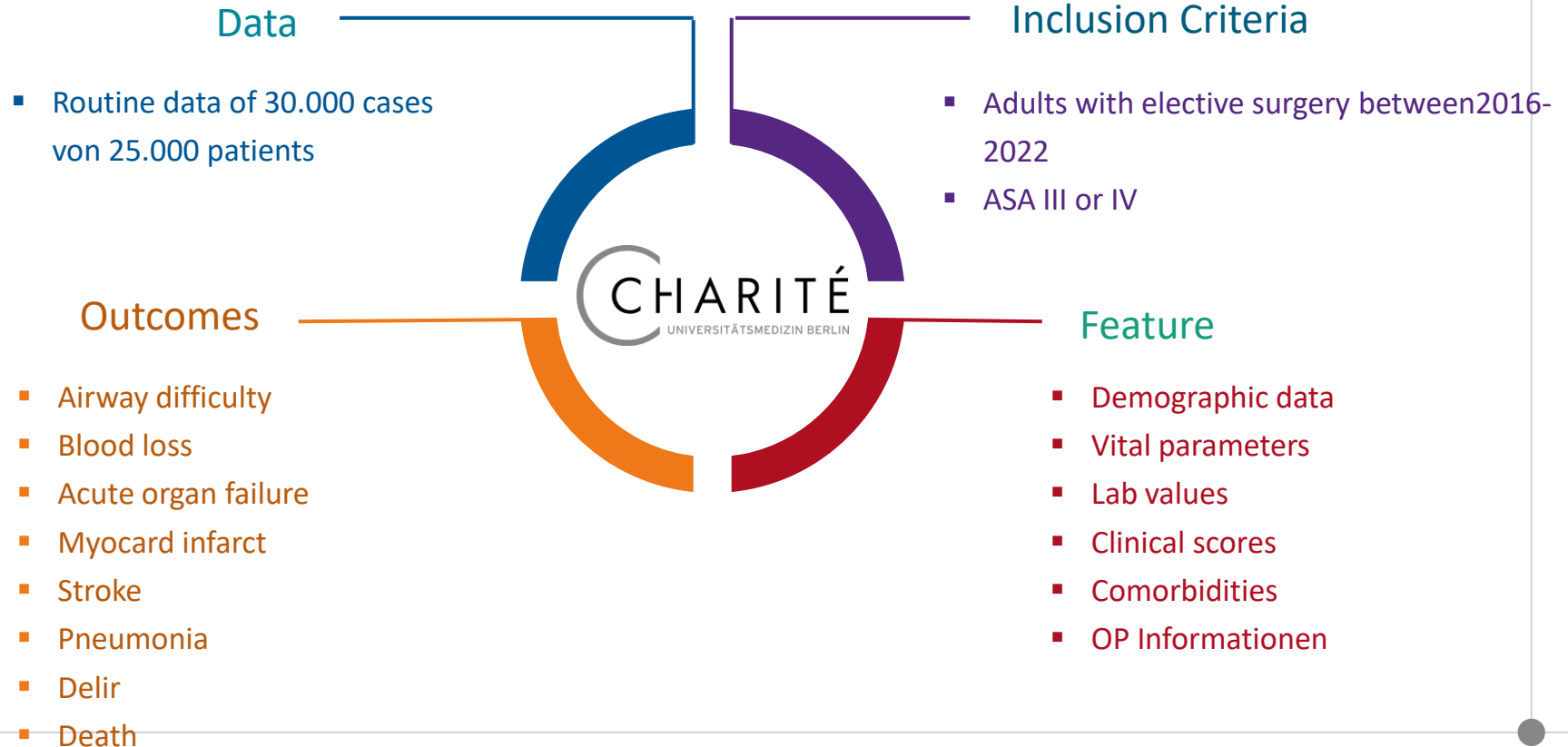
1. assessment of patient-specific risks
2. identify the main risk factors for post-operative complications
3. assess the impact of changes in variable risk factors that can be influenced by patients and their physicians (e.g. body weight, blood pressure, medication)

SW has to consider requirements

- Regulatory and conventions
- Ethics
- Usability
- Data Formats



AI Model

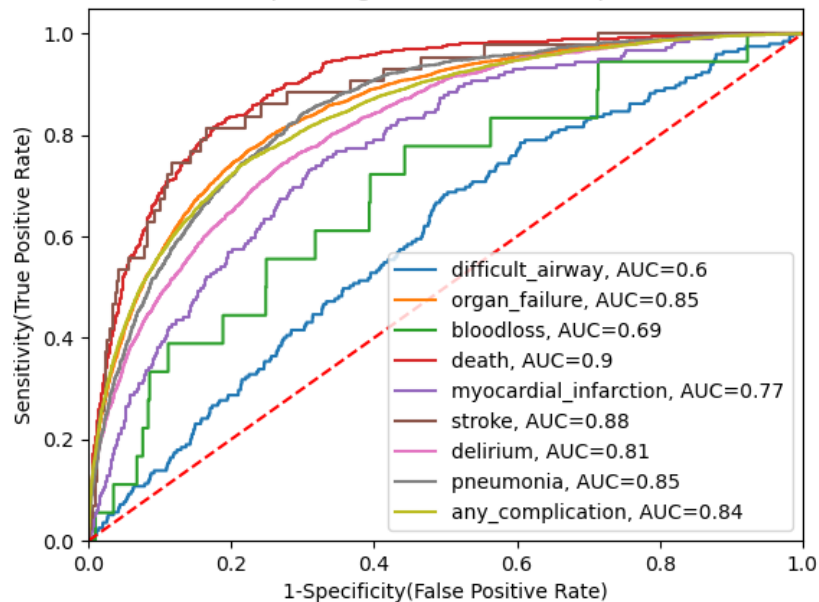


	Charité cohort		KIPeriOP cohort
	training set (N=20663)	temporal validation set (N=7701)	external validation set (N=231)
Complications n(%)			
difficult airway	0.9%	0.6%	-
acute organ failure	22.9%	19.9%	1.3%
bloodloss	0.1%	0.1%	8.7%
death	1.9%	1.9%	2.2%
myocardial infarction	0.9%	0.7%	1.3%
stroke	0.2%	0.3%	0.4%
delirium	12.0%	30.6%	1.7%
pneumonia	7.1%	6.2%	1.3%
any of the above	30.2%	38.6%	13.4%
Demographics			
age	67.3 (14.3)	67.3 (14.9)	68.7 (12.4)
sex			
female	43%	44%	33%
male	57%	56%	67%

Prediction Model

Preliminary results

Receiver Operating Characteristic, cv performance



Receiver Operating Characteristic, temporal performance

