CORE-MD

Coordinating Research and Evidence for Medical Devices

CORE-MD Webinar
October 17, 2023

The clinical evaluation of AI and standalone software: keeping the Balance between Benefit and Risk.

Housekeeping rules

- This webinar is being recorded

  Recording will be made available on https://www.core-md.eu/core-md-webinars/ within a few days
- All participants are in « lecture mode ».
- We welcome questions: please enter questions in the Q&A and the moderator and panelists will provide an answer during the discussion phase (or in written format if not enough time)
- For urgent technical questions, please use the chat function
- Interested in CORE-MD: subscribe to newsletter@ https://www.core-md.eu and follow-us on social media

CORE-MD Webinar on clinical evaluation of AI

Frank Rademakers
KU Leuven

Bernd Grimm
Luxembourg Institute of Health Department of Precision

Eva Van Steijvoort
KU Leuven

Claudius Greisinger
Administrator at European Commission

Upcoming webinar – 6th November – 17.00 CET

Systematic review with the cardiovascular and diabetic devices

17.00 - 17.05 | **Introduction by Moderator,**

Prof. Robert Byrne, Director of Cardiology and Director of the Cardiovascular Research Institute at Mater Private Network, Dublin, Ireland

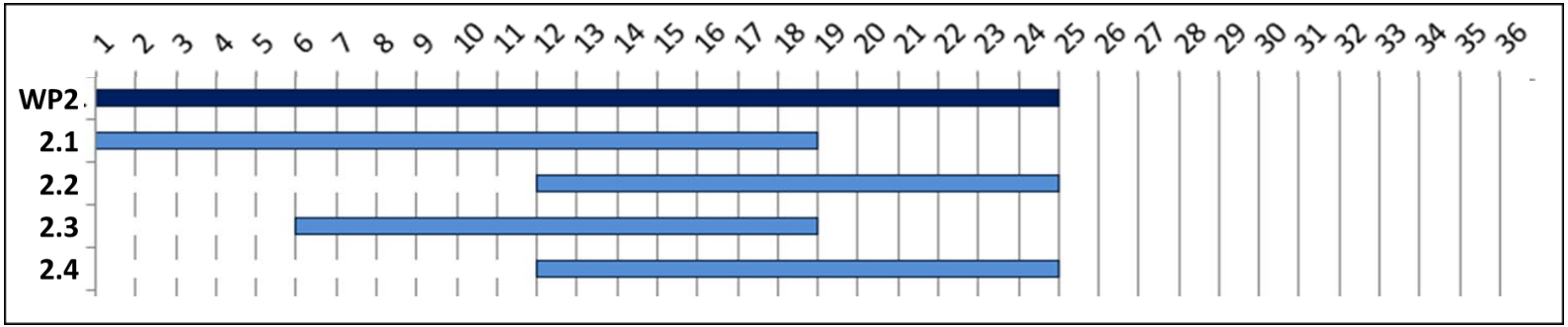
17.05 - 17.15| **Quality and transparency of clinical evidence for high-risk cardiovascular medical devices**.
PD Dr G Siontis, Inselspital, Universitätsspital Bern · Department of Cardiology

17.15 - 17.30| **Quality and transparency of clinical evidence for high-risk diabetic medical devices**,
PD Dr A Bano, Institute of Social and Preventive Medicine (ISPM)University of Bern, Switzerland

17.30 – 18.15 | Moderated discussion with the audience

Tasks

WP 2	Strengthening evidence for high-risk medical devices: New methods for generating clinical evidence	ESC/EFORT/URPL (Fraser/Kjærsgaard-Andersen/Szulc)	Consortium partners
Task 2.1	Providing evidence during the early development of high-risk medical devices	Oxford (McCulloch)	ESC, EFORT, BUH, RIVM, Team NB, URPL
Task 2.2	New designs for randomised clinical trials and studies of high-risk medical devices	UCR (James)	Oxford, ESC, EFORT, LUMC, UMIT, EPF, BUH, EAP
Task 2.3	Developing guidance for the evaluation of artificial intelligence and standalone software in medical devices	KU Leuven (Rademakers)	POLIMI, ESC, EFORT, URPL
Task 2.4	Recommendations concerning high-risk medical devices in children	EAP (Koletzko)	BioMed Alliance, ESC, EFORT



Description of the activities

- Phase I
  - Meetings with entire group and subgroup discussions
  - Background text with position of consortium on relevant topics
  - Publication with review of definitions, expert recommendations and regulatory initiatives
- Phase II
  - Practical recommendations for clinical evaluation of AI MDSW
  - Deliverable: Report 3.2023
  - Delphi Clinicians: 8.2023
  - Planned
    - consultation Regulators and NB's
    - Presentation to MDCG November 8th 2023

Artificial intelligence in medical device software and high-risk medical devices – a review of definitions, expert recommendations and regulatory initiatives

Alan G Fraser [1], Elisabetta Biasin [2], Bart Bijnens [3], Nico Bruining [4], Enrico G Caiani [5], Koen Cobbaert [6], Rhodri H Davies [7], Stephen H Gilbert [8], Leo Hovestadt [9], Erik Kamenjasevic [10], Zuzanna Kwade [11], Gearóid McGauran [12], Gearóid O'Connor [13], Baptiste Vasey [14], and Frank E Rademakers [15], for the CORE–MD consortium.

EXPERT REVIEW OF MEDICAL DEVICES 2023, VOL. 20, NO. 6, 467–491
https://doi.org/10.1080/17434440.2023.2184685

CORE–MD
Coordinating Research and Evidence for Medical Devices

EU Horizon 965246

A precise and inclusive definition is unnecessary ..

Marvin Minsky = the science of making machines do things that would required intelligence if done by men. (1968)

WHO = the ability of algorithms to learn from data so that they can perform automated tasks without every step in the process having to be programmed explicitly by a human.

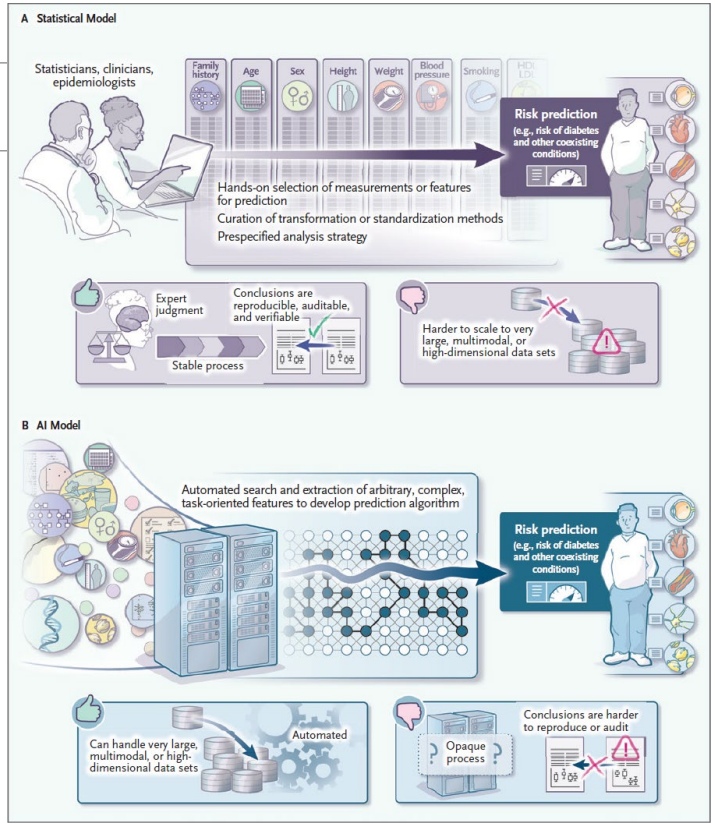
https://www.who.int/publications/i/item/9789240029200

OECD = a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments.

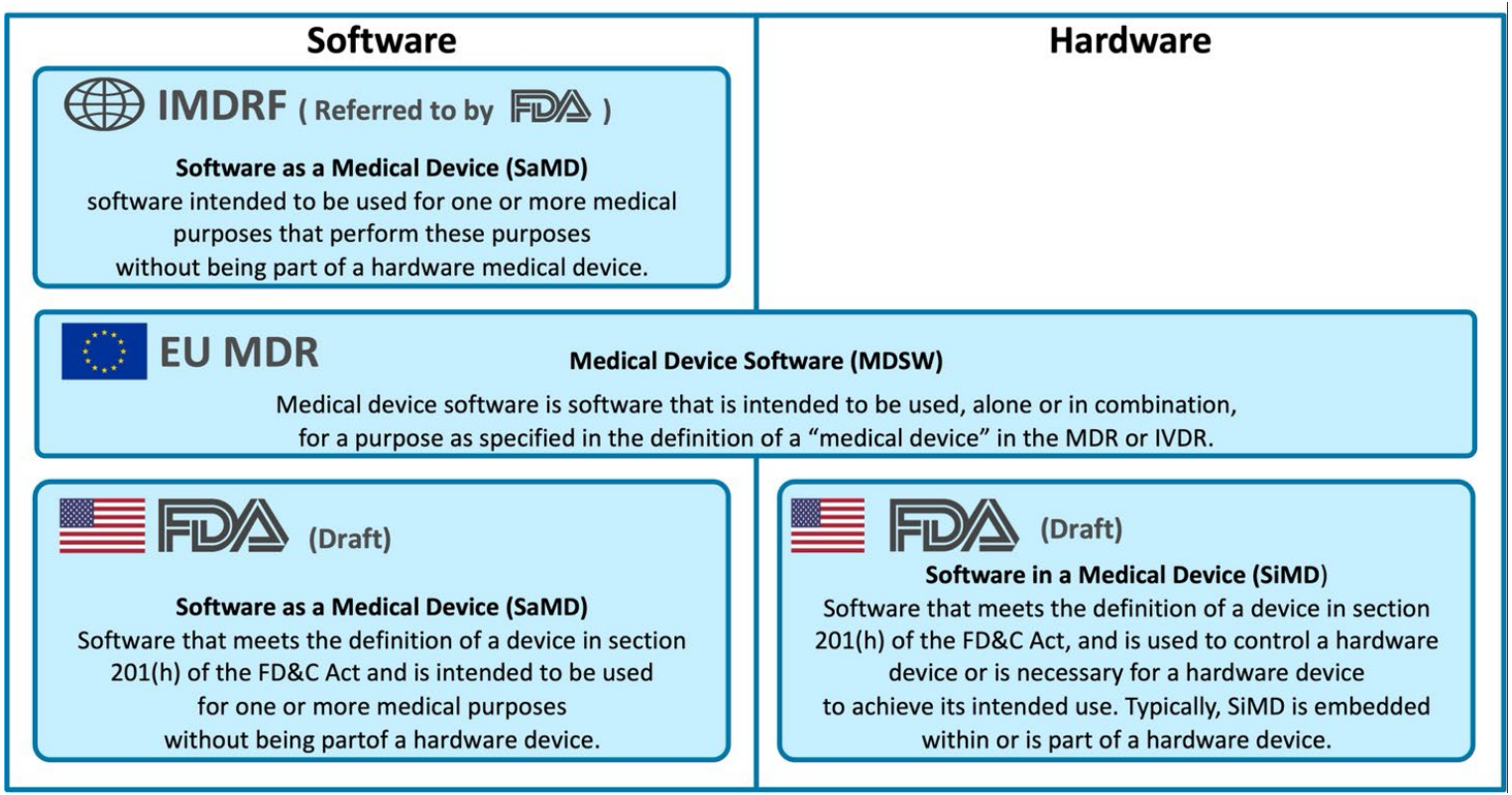
https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

Statistical modeling	Machine learning/AI
Estimating a model/Fitting	Learning
Prediction/Regression	Supervised learning
Latent variable modeling	Unsupervised learning
Case/Data point	Example/Instance
Sensitivity	Recall
Positive predictive value	Precision
Independent variable/Covariate	Feature
Dependent variable	Target
Response	Label
Parameters	Weights
Log likelihood	Loss

Faes L et al, Front Digit Health. 2022; 4: 833912



N Engl J Med 2023;389 (13):1211-9.
DOI: 10.1056/NEJMra2212850
September 28, 2023



Sources

- IMDRF SaMD WG/N10, 2013
- EU MDCG 2019-11
- FDA Premarket Guidance 4.11.2021

European and global organisations engaged in regulatory initiatives for AI

EU initiatives on governance of artificial intelligence

- 2018 / COM / **AI Watch** at the JRC in Seville
- 2018 / COM / DG CNECT / High-Level Expert Group
- 2020 / EP / STOA / **Centre for AI** (C4AI)
- 2021 / COM / Proposed **Regulation on AI** (2021/0106)

Some relevant EU legislation

- (EU) 2016/679 / **General Data Protection Regulation** (GDPR)
- (EU) 2022/868 / **Data Governance Act** (DGA)
- COM(2022) 197 / Proposal for a **European Health Data Space** (EHDS)
- COM(2022) 68 / Proposal for a **Data Act**
- EP Resolution 20.10.2020 on **IP rights for development of AI technologies**
- Directive 85/374/EEC of 25 July 1985 on liability for defective products
- COM(2022) 496 / Proposal for an **AI Liability Directive**
- Network and Information Security Directive (NIS Directive) of 2016 [**cybersecurity**]

General conclusions from the CORE–MD review of medical AI systems

- There is a real risk of over-regulation.
- **Standards should be based on scientific evidence and proportionate to the clinical risks.**
- Concordance of scope and regulatory requirements would be preferable.
- A concrete and practical initiative for global regulatory convergence is needed.
- Recommendations for medical AI (e.g. data acquisition, pre-processing, model, study population, performance, benchmarking, data availability) are relevant for all clinical studies.
- **Regulatory efforts should concentrate on gaps in advice, or challenges unique to AI devices:**
  - **specific methodologies for clinical investigations related to particular defined levels of risk**
  - how to assure use of AI system only for individuals for whom it has been validated
  - how to approve iterative changes in software that may be self-learning
  - how to conduct appropriate post-market surveillance

Inter-agency LLM task force: a progress update and beyond

Joshua Xu, Ph.D.

Branch Chief, Research-to-Review (R2R)
Division of Bioinformatics and Biostatics
National Center for Toxicological Research
U.S. Food and Drug Administration
Email: Joshua.xu@fda.hhs.gov

13TH GLOBAL SUMMIT ON REGULATORY SCIENCE (GSRS23) (September 27-28, 2023, EFSA, Parma, Italy)

Disclaimer: *The information in this presentation represents the opinions of the speaker and does not necessarily represent NCTR's or FDA's position or policy.*

CORE-MD
Coordinating Research and Evidence for Medical Devices

An Explosion of LLMs Presents Opportunities

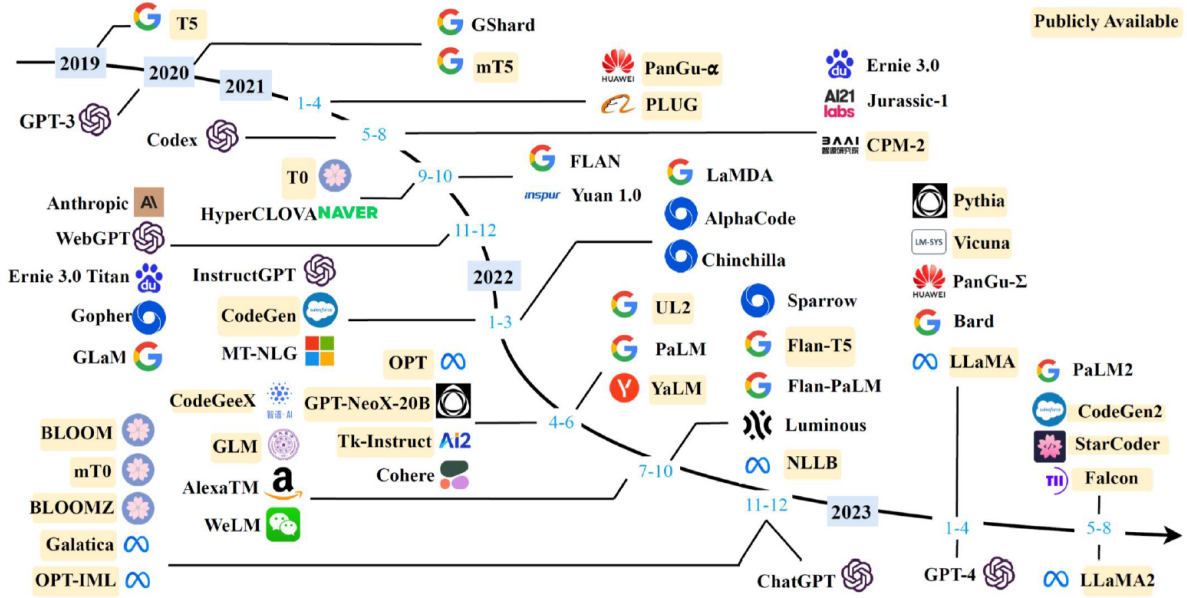


Fig. 2: A timeline of existing large language models (having a size larger than 10B) in recent years. The timeline was

Zhao, et al. A Survey of Large Language Models (http://arxiv.org/abs/2303.18223)

Source Activity:	Work package 2, Task 2.3
Title:	Expert advice on criteria for the regulatory evaluation of ML and AI
Lead Beneficiary:	KU Leuven
Nature:	Report
Dissemination level:	Public
Editor:	Frank E. Rademakers (KU Leuven)
Authors:	Frank E Rademakers (KU Leuven), Elisabetta Biasin (KU Leuven), Bart Bijnens (KU Leuven), Nico Bruining (Erasmus MC)*, Enrico G. Caiani (POLIMI), Koen Cobbaert (Philips)*, Rhodri H. Davies (University College London)*, Job N. Doornberg (University Medical Center Groningen)*, Stephen Gilbert (Technische Universität Dresden), Leo Hovestadt (Elektra), Erik Kamenjasevic (KU Leuven), Zuzanna Kwade (Dedalus)*, Gearoid McGauran (HPRA), Gearoid O'Connor (HPRA), Baptiste Vasey (UOXF) and Alan G Fraser (ESC) **External experts involved in CORE-MD activities*
Status:	Final

Principles

- Combine Scientific <-> Regulatory
- Balance between too prescriptive and too generic
- Risk – based
- Explainability
- Usability and Acceptability by individuals, patients and caregivers
- Moving from
  - Waterfall system

    to
  - Agile approach

The in silico spectrum

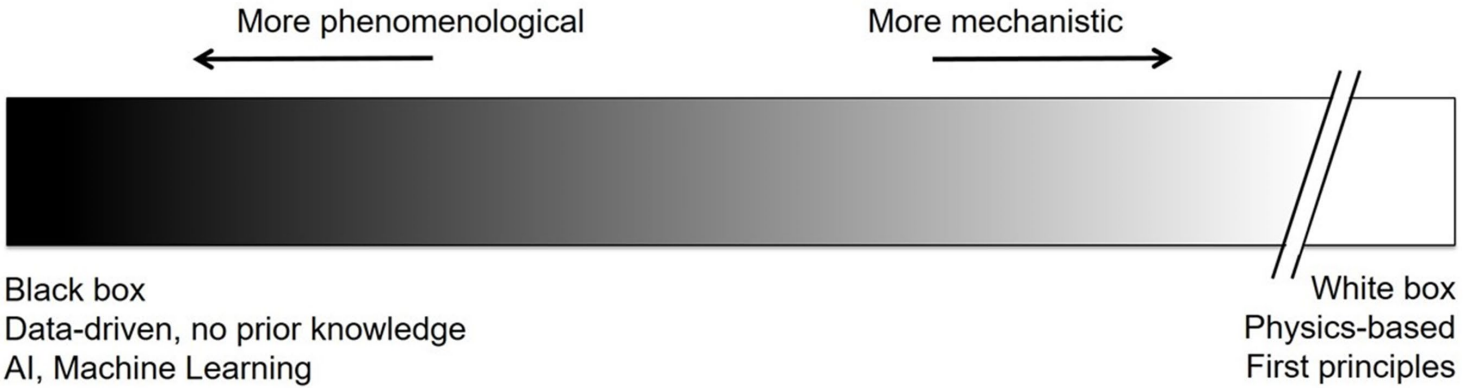


Table 1. Evaluation for AI software compared to the approval processes of drug and devices for healthcare

Study phases	Drug	Device	AI in healthcare	Examples of study methods
Phase 0 Discovery and invention	Compound development In vitro/animal tests	User needs and workflow assessment Prototype design and development	User needs and workflow assessment Data quality check Algorithm development and performance evaluation Prototype design	Ethnographic studies to identify user needs, laboratory studies on limited data sets to measure algorithm prediction accuracy
Phase 1 Safety and dosage	Determine optimal dose Identify potential toxicities	Quality control Design updates	In silico algorithm performance optimization Usability tests	Determination of thresholds to balance sensitivity and specificity for a particular clinical use case, scenario-based testing to assess cognitive overload
Phase 2 Efficacy and side effects	Early efficacy tests Adverse event identification	Proof-of-concept tests Potential harm identification Design and quality improvement	Controlled algorithm performance/efficacy evaluation by intended users in medical setting Interface design Quality improvement	Retraining and reassessing model performance with larger real-world data sets, measurement of the efficiency of information delivery and workflow integration with representative users, pilot study of predictive algorithm in a clinical setting
Phase 3 Therapeutic efficacy	Clinical trial Adverse event identification	Clinical trial Adverse event identification	Clinical trial Adverse events identification	Randomized controlled trial to test whether delivery AI-based decision support affects clinical outcomes and/or results in user overtrust
Phase 4 Safety and effectiveness	Postmarketing surveillance	Postapproval studies	Postdeployment surveillance	Measurement of algorithmic performance drift

JAMIA Open, 3(3), 2020, 326–331

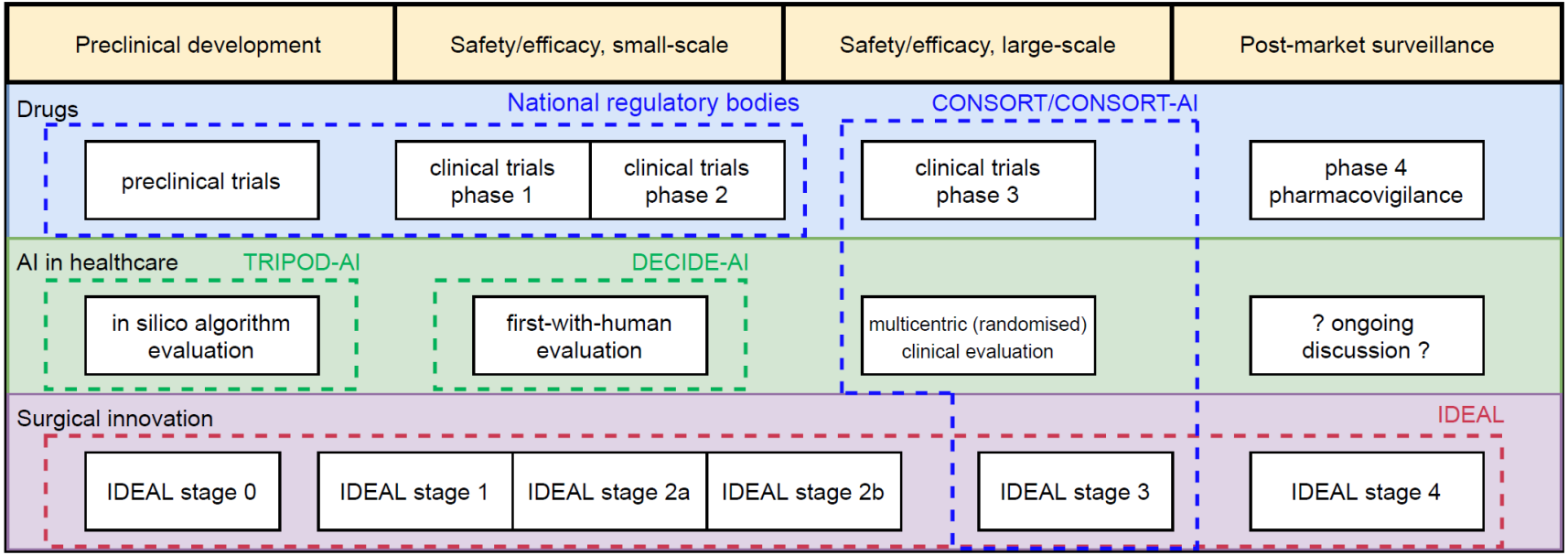


Figure 1: Comparison of the development pathways for drugs, AI-based algorithms and surgical innovation. The dotted lines indicate reporting guidelines.

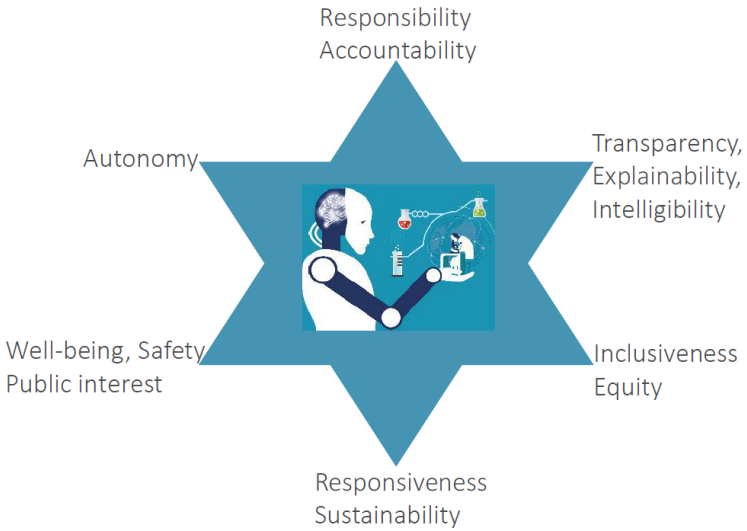
Proposed solutions

- AI is not totally different from other devices, regular stats and software
  - It has specific features, risks and challenges
- Risk–based approach with scoring system
  - Type of disease, condition, healthcare situation
  - Significance of information
  - Human interpretability & usability in clinical workflow
  - Quality and transparency of data used for training, validation, testing
- Depending on risk score
  - Use MDCG 2020-1 doc on Guidance on Clinical Evaluation
  - Matrix of requirements for clinical evaluation: the ability of the AI tool to yield clinically meaningful output, in accordance with the intended purpose
    - Pre-market
    - Post-market

Partial but significant shift to Post-Market surveillance

- Specific challenge of AI tool to be evaluated in limited testing
  - Shift in user perspectives and capabilities
  - Drift in target population
  - Adaptive learning
    - Stepwise
    - Continuous
  - Personalized use
- “Conditional” release
  - Exclude Higher risk AI categories from such release
- In comparison to FDA, less emphasis on manufacturer characteristics
  - Risk exclusion academic, SME’s and startups

WHO: 6 principles for AI in health



1	Protecting human autonomy: humans remain in control, confidentiality, privacy, consent through legal frameworks
2	Promoting **human well-being and safety and the public interest**: safety, accuracy, efficacy for well-defined use cases/indications. Measures of quality control/improvement in practice
3	Ensuring **transparency, explainability and intelligibility**: sufficient information available before deployment, for public consultation and debate on how AI should / should not be used
4	Fostering **responsibility and accountability**: use under appropriate conditions by appropriately trained people. Mechanisms for questioning and redress in case of adverse effects
5	Ensuring **inclusiveness and equity**: widest possible equitable use & access, irrespective of age, sex, gender, income, race, ethnicity, sexual orientation, ability or other characteristics protected under human rights
6	Promoting AI that is responsive and sustainable: designers, developers, users assess AI applications during use. Minimize environmental impacts, enhance energy efficiency; governments and companies should address disruptions, e.g. training & adaptation to AI use, potential job losses

Illustration: mmustafabozdemir/iStock

European Commission

CORE-MD

Coordinating Research and Evidence for Medical Devices

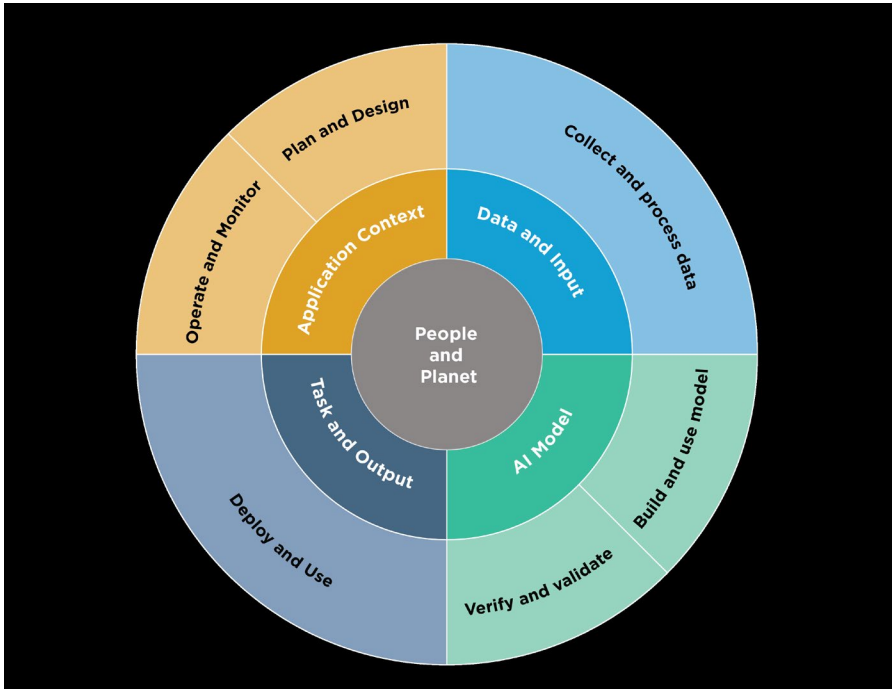
EU Ethics Guidelines for Trustworthy Artificial Intelligence

1. Human agency and oversight
2. Technical robustness and safety
3. Privacy and Data governance
4. Transparency
5. Diversity, non-discrimination and fairness
6. Societal and environmental well-being
7. Accountability

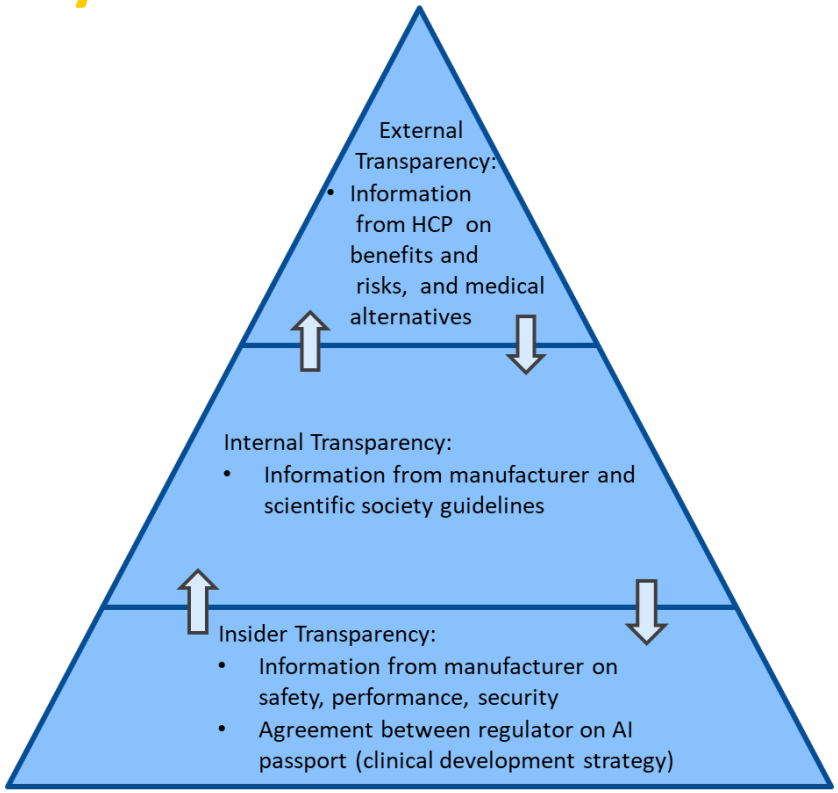
"Doctors can potentially perform a more accurate and detailed analysis of a patient's complex health data, even before people get sick .. leading to earlier detection of diseases, more efficient development of medicines, more targeted treatments and ultimately more lives saved"

https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

Lifecycle and Key Dimensions of an AI System. Modified from OECD (2022) OECD Framework for the Classification of AI systems—OECD Digital Economy Papers. The two inner circles show AI systems' key dimensions and the outer circle shows AI lifecycle stages. Ideally, risk management efforts start with the Plan and Design function in the application context and are performed throughout the AI system lifecycle. See Figure 3 for representative AI actors.



Transparency

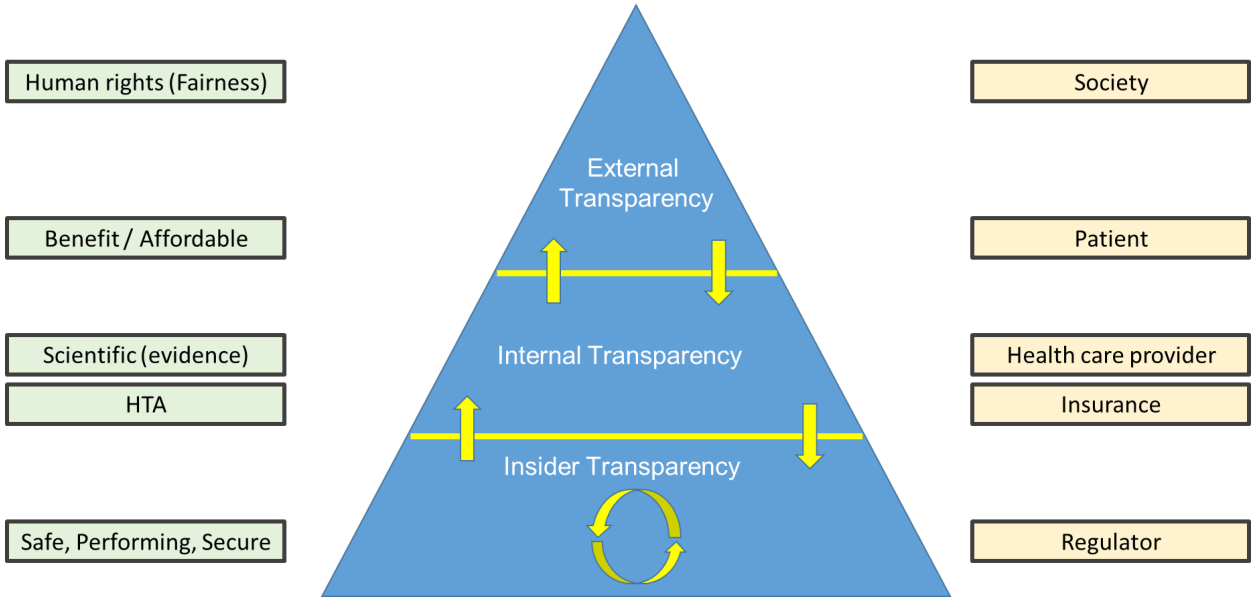


Goals

- Society assessment on human rights and sustainability impact
- Patient assessment on benefits and risks, and alternatives
- Informed consent from patient to HCP
- Shared decision making with patient (explicit or implicit)
- Regulator assessment safety, performance, (cyber) security

Kiseleva A, Kotzinos D, De Hert P. Transparency of AI in healthcare as a multilayered system of accountabilities: between legal requirements and technical limitations. Front Artif Intell. 2022;5. doi:10.3389/frai.2022.879603

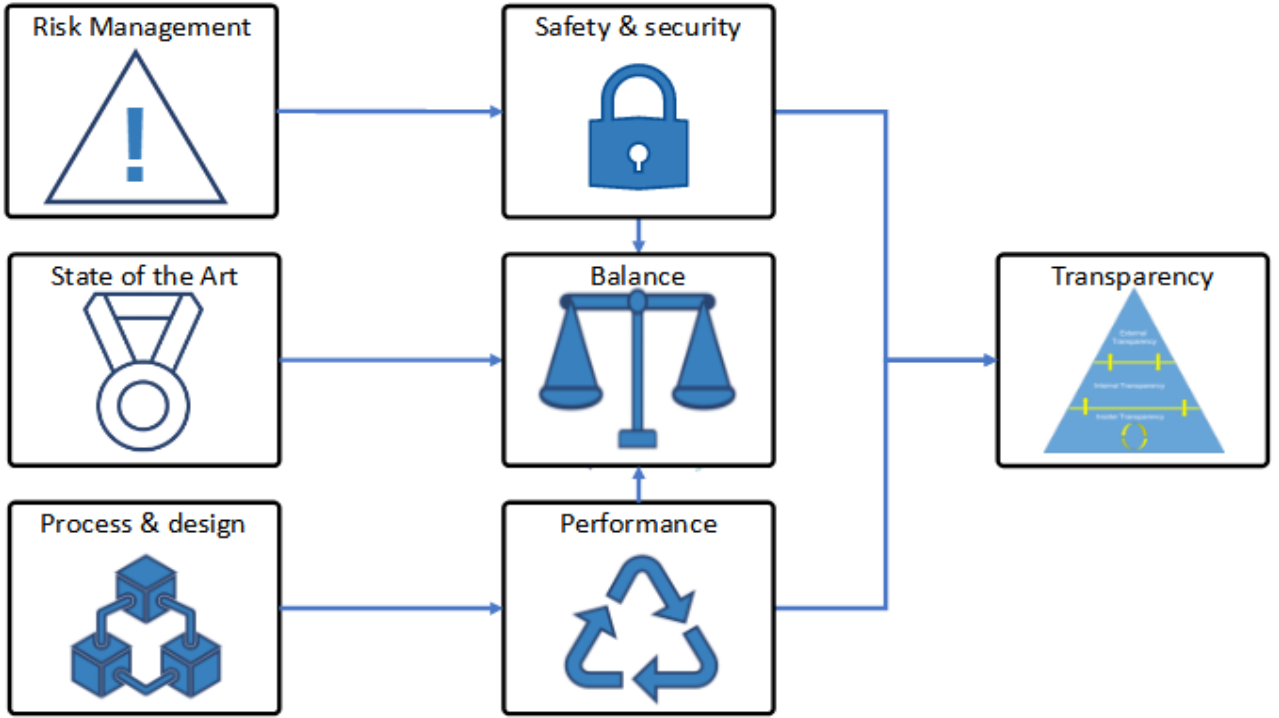
Transparency



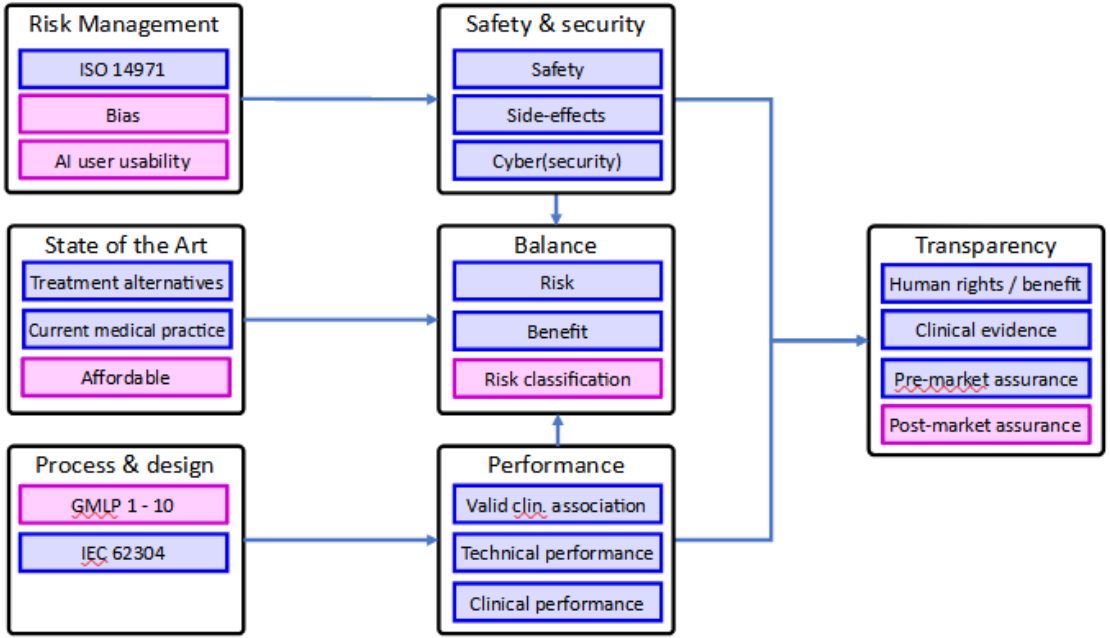
CORE-MD

Coordinating Research and Evidence for Medical Devices

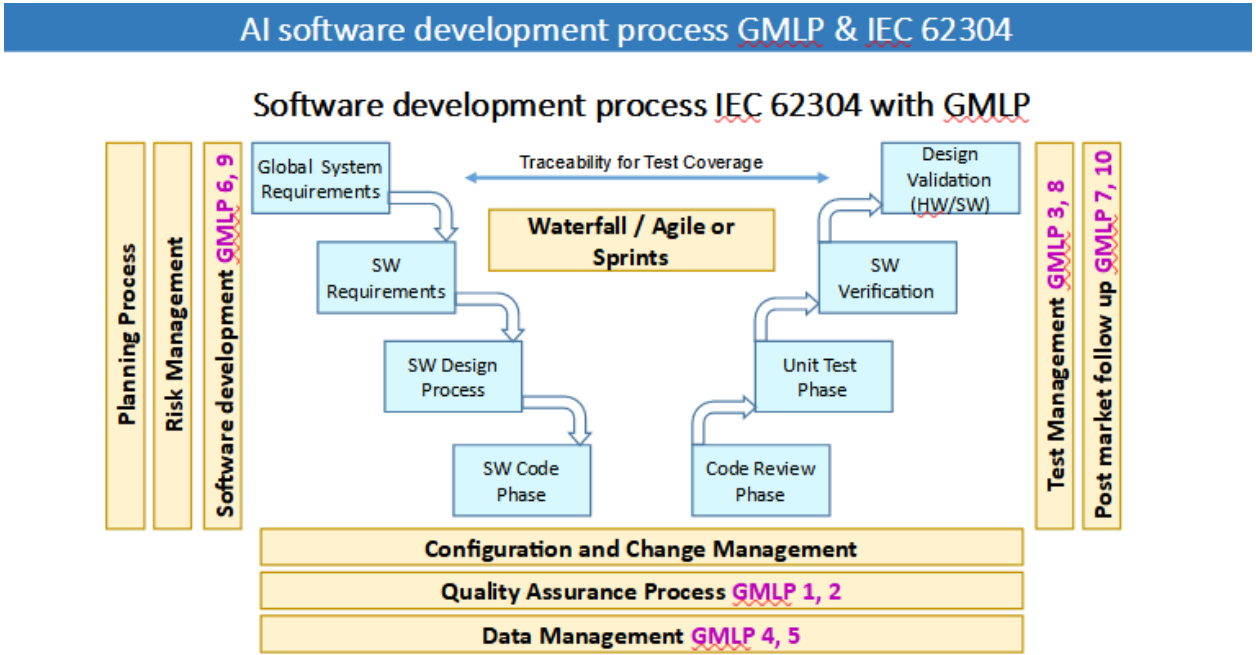
AI medical device – regulatory requirements



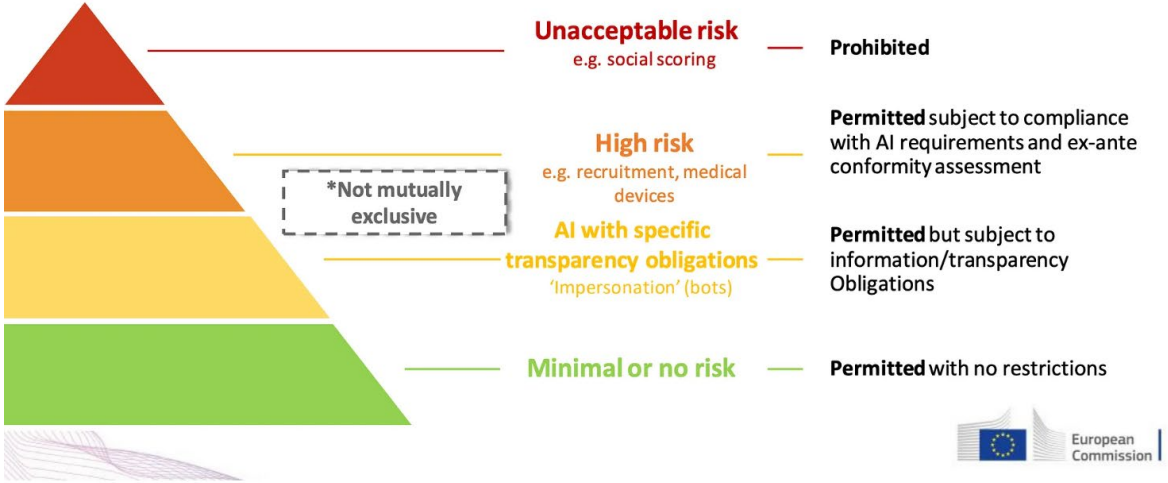
AI medical device – high level regulatory requirements



Process & Design



Risk Benefit Balance

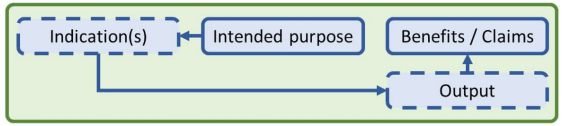


A Risk-Based Approach to Regulation. Source: A European Strategy for Artificial Intelligence April 21, 2023.

State of Healthcare situation or patient condition	Significance of Information provided by the MDSW to a healthcare situation related to diagnosis/therapy: **High** Treat or diagnose *~ IMDRF 5.1.1*	**Medium** Drives clinical management *~ IMDRF 5.1.2*	**Low** Informs clinical management *(everything else)*
Critical situation or patient condition *~ IMDRF 5.2.1*	**Class III** *Category IV.i*	**Class IIb** *Category III.i*	**Class IIa** *Category II.i*
Serious situation or patient condition *~ IMDRF 5.2.2*	**Class IIb** *Category III.ii*	**Class IIa** *Category II.ii*	**Class IIa** *Category I.ii*
Non-serious situation or patient condition *(everything else)*	**Class IIa** *Category II.iii*	**Class IIa** *Category I.iii*	**Class IIa** *Category I.i*

Table 1: Classification Guidance on Rule 11

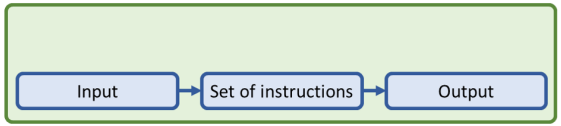
Risk Score: MDCG guidance



Clinical Performance:
The MDSW should generate **clinically relevant output** or **benefits** when **used as intended**.

Schematic view clinical performance score

Criterion	Associated Levels	Partial score	Clinical Performance Score (CPS)
Type of disease, condition, disability, healthcare situation: risk for patient	Non serious	1	
	Serious	2	
	Critical	3	
Significance of information: use in clinical flow	Inform	1	
	Drive	2	
	Diagnose or treat	3	
TOTAL CPS			Sum of the two



Technical Performance:
The MDSW **output** should be **accurate** and **reliable** for the input

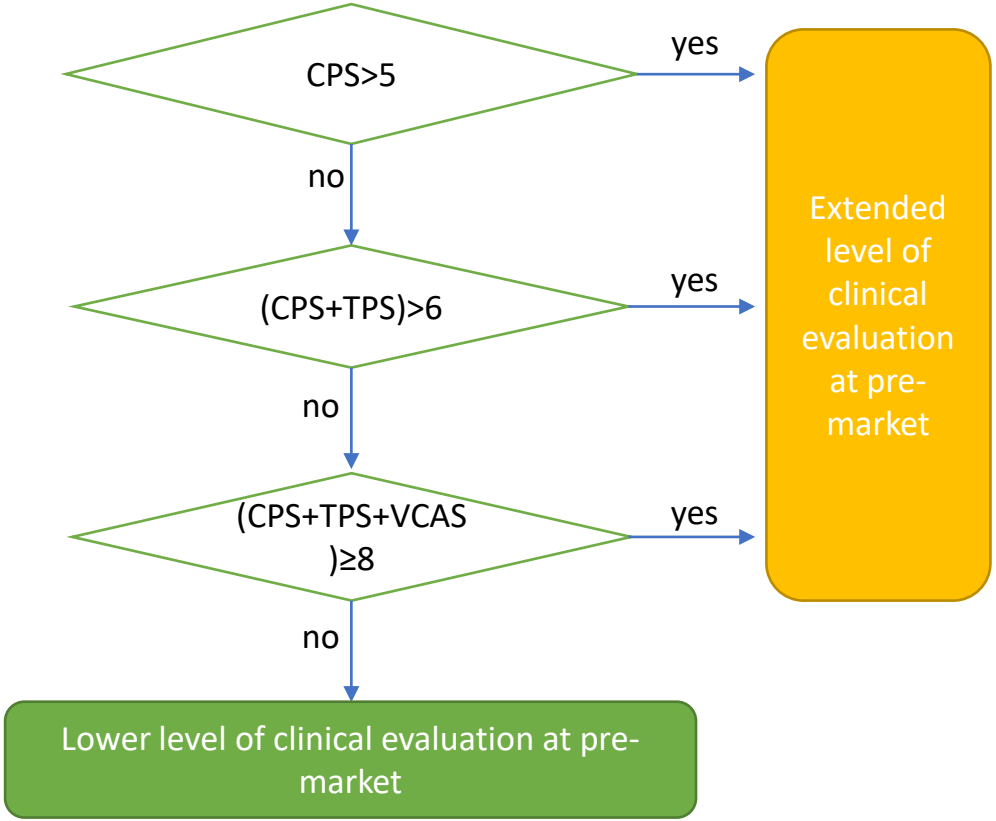
Schematic view technical performance score

Extension of validation/testing	Level of validation/testing	Technical performance score (TPS)
Broad external validation	strong	1
Narrow external validation	moderate	2
Internal validation	weak	3

Valid clinical association:
The MDSW **output** should **associate** with an **indication** (clinical condition or physiological state).

Schematic view of valid clinical association

Transparency and Oversight	Valid clinical association score (VCAS)
Easy	1
Difficult	2
Impossible	3


CPS>5
yes
no
(CPS+TPS)>6
yes
no
(CPS+TPS+VCAS)≥8
yes
no
Extended level of clinical evaluation at pre-market
Lower level of clinical evaluation at pre-market

Requirements through AI life-cycle

Plan and design: audit and impact assessment: articulate and document

- System's concept and objectives
- Underlying assumptions and context

Data and Input: collect and process data

- Gather, validate and clean data
- Document the metadata and characteristics of the dataset

AI model build and use

- create or select algorithm
- train model

AI model verify and validate

- calibrate
- interpret model output

Deploy and integrate

- Check compatibility with legacy systems
- Verify regulatory compliance
- Manage organizational changes (including pathway analysis)
- Evaluate training requirements

Pilot evaluation

- Clinical utility
- System safety (including analysis of errors and harms)
- User experience/human factors/usability
- Iterative improvement and documentation of changes

Comparative evaluation

- Effectiveness/impact assessment (all affected persons)
- Safety at scale

Long term operation and monitoring

- Performance monitoring
- Safety monitoring
- Drift monitoring
- Update versioning and documentation
- Decommissioning

CORE-MD

Coordinating Research and Evidence for Medical Devices

Requirement Matrix

Phase	AI life-cycle stages	Sub-stages	Requirement categories		
					comment
Pre release	Plan and design: audit and impact assessment	System's concept and objectives	+	+	
		Underlying assumptions and context	+	+	
	Data and Input: collect and process data	Gather, validate and clean data	+	+	
		Document the metadata and characteristics of the datasets	+	+	
	AI model build and use	Create or select algorithm	+	+	
		Train model	+	+	
	AI model verify and validate	Calibrate	+	+	
		Interpret model output	+	+	
	Deploy and integrate	Check compatibility with legacy systems	+	+	
		Verify regulatory compliance	+	+	
		Manage organizational changes (including pathway analysis)	-	+	
		Evaluate training requirements	-	+	
	Pilot evaluation	Clinical utility	+	+	
		System safety (including analysis of errors and harms)	+	+	
		User experience/human factors/usability	-	+	
		Iterative improvement and documentation of changes	-	+	
	Comparative evaluation	Effectiveness/impact assessment (all affected persons)	-	+	
		Safety at scale	-	+	
	Long term operation and monitoring	Performance monitoring	-	-	
		Safety monitoring	-	-	
		Drift monitoring	-	-	
		Update versioning and documentation	-	-	
		Decommissioning	-	+	

Requirement Matrix

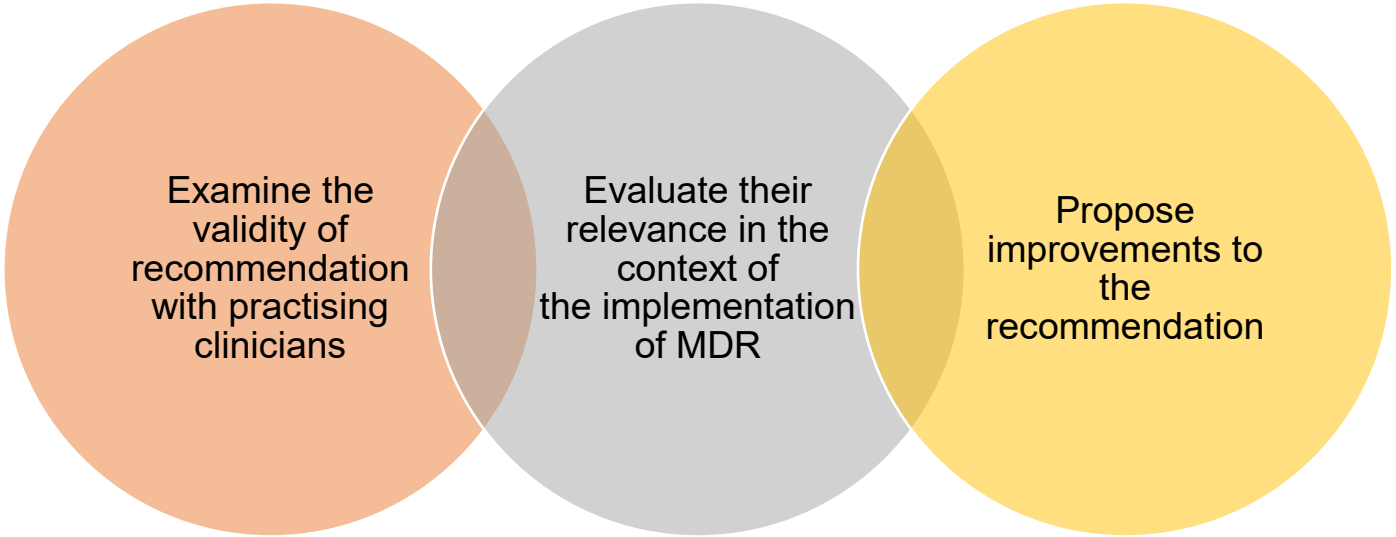
Phase	AI life-cycle stages	Sub-stages	Requirement categories		comment
	Plan and design: audit and impact assessment	System's concept and objectives	+	+	Drift
		Underlying assumptions and context	+	+	Drift
	ata and Input: collect and process data	Gather, validate and clean data	+	+	Depending on change
		Document the metadata and characteristics of the datasets	-	-	Unless changed
	AI model build and use	create or select algorithm	-	-	Unless changed
		train model	-	-	
	AI model verify and validate	calibrate	-	-	
		interpret model output	-	-	
		Check compatibility with legacy systems	+	+	
		Clinical utility	+	+	
	Pilot evaluation	System safety (including analysis of errors and harms)	+	+	
		User experience/human factors/usability	+	+	
		Iterative improvement and documentation of changes	+	+	
	Comparative evaluation	Effectiveness/impact assessment (all affected persons)	+	+	
		Safety at scale	+	+	
	Long term operation and monitoring	Performance monitoring	+	+	
		Safety monitoring	+	+	
		Drift monitoring	+	+	
		Update versioning and documentation	+	+	
		Decommissioning	+	+	

Practical Questionnaire List

CORE-MD

Coordinating Research and Evi
for Medical Devices

Objective of the Delphi process

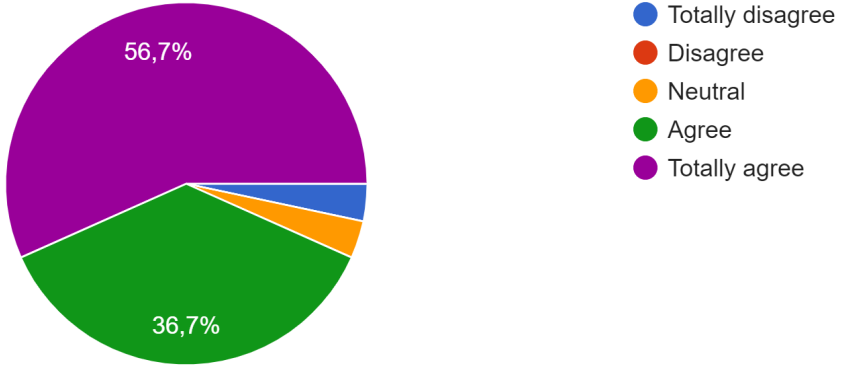


Preliminary Results

1. Do you agree or disagree that the AI manufacturer should provide/ensure the information required for external transparency for any AI medical device?

30 réponses

QUESTION 1	
Total answers	30
Threshold 70 %	21
Passed (Agree + totally Agree)	93,33 %



CORE-MD, Coordinating Research and Evidence for Medical Devices, aims to translate expert scientific and clinical evidence on study designs for evaluating high-risk medical devices into advice for EU regulators.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 945260

For more information, visit: www.core-md.eu

ESC
EFORT – JOINT EFFORTS
Biomedical Alliance in Europe
European Academy of Paediatrics
Paediatric Section of U.E.M.S
EPF European Patients Forum
Leiden University Medical Center
University of Oxford
Region Uppsala
RCSI University of Medicine and Health Sciences
INSELSPITAL
UNIVERSITÄTSSPITAL BERN
HÔPITAL UNIVERSITAIRE DE BERNE
KU LEUVEN
UMIT TIROL
THE TYROLEAN PRIVATE UNIVERSITY
UNIVERSITY OF GOTHENBURG
POLITECNICO MILANO 1863
HPRA
An tÚdarás Rialála Táirgí Sláinte
Health Products Regulatory Authority
LÆGEMIDDELSTYRELSEN
DANISH MEDICINES AGENCY
National Institute for Public Health and the Environment
Ministry of Health, Welfare and Sport
ISTITUTO SUPERIORE DI SANITÀ
aihta
HTA Austria
Austrian Institute for Health Technology Assessment GmbH
AETSA
TEAM NB

CORE-MD
Coordinating Research and Evidence for Medical Devices