An evidence pathway for trustworthy AI innovations in medicine and healthcare

bridging the gap between developer and user communities

CORE-MD webinar on AI in health | 17. October 2023

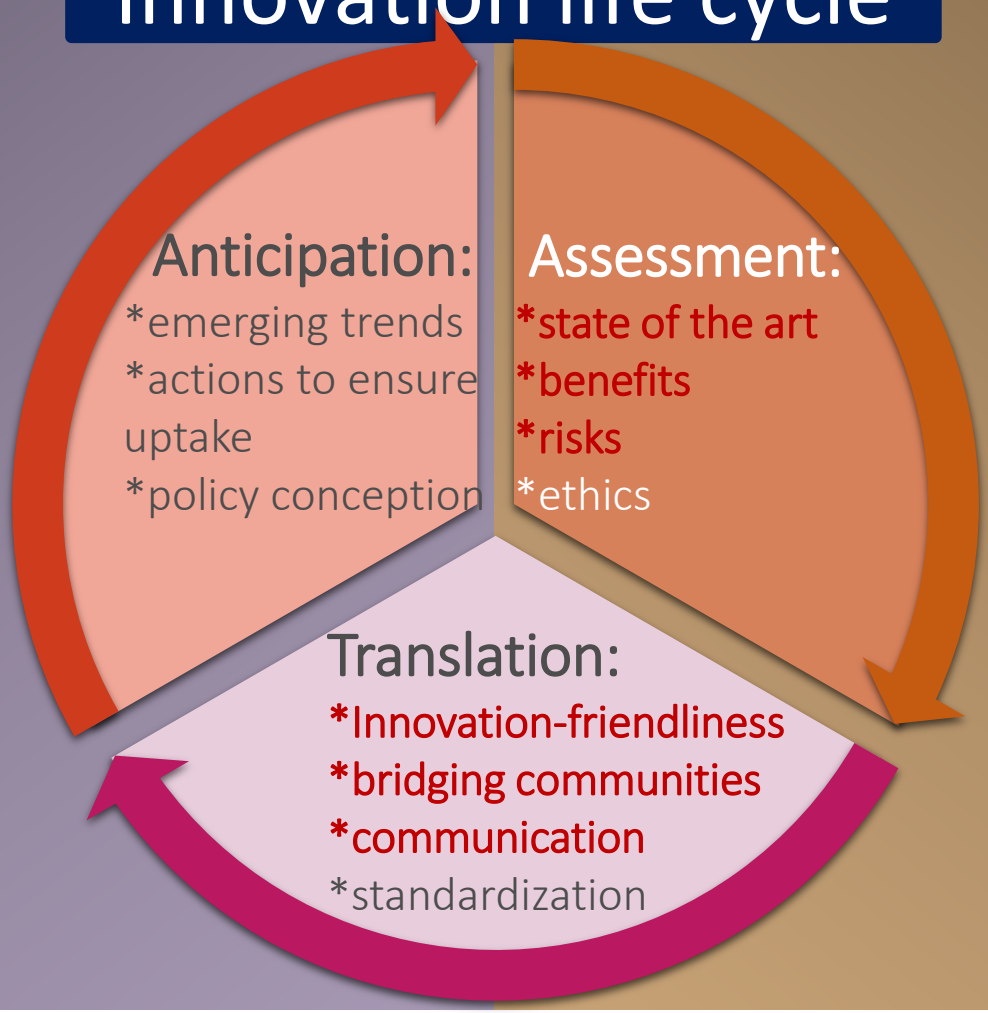
Dr. Claudius B Griesinger | European Commission | Joint Research Centre (JRC)

JRC portfolio
"Innovation in life and health sciences"

Health challenges & socioeconomic challenges

- Cancer
- Rare diseases
- AMR
- Climate change
- New infectious diseases
- Inequality

Innovation life cycle

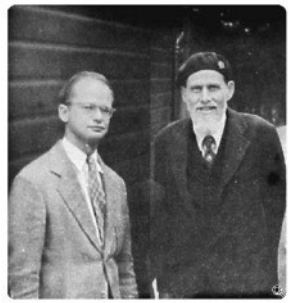


Technological & societal drivers

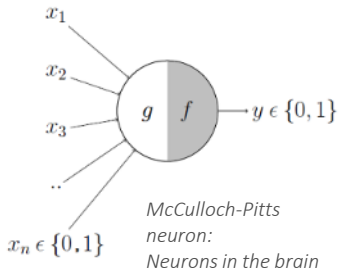
- Biorevolution
- Data, algorithms AI
- Digital twins
- Innovative therapies
- Precision medicine

Interaction with other portfolios, e.g. on Cybersecurity or Trustworthy AI

Cybernetics: a natural science of perception and thought



McCulloch [right] and Pitts [left] in 1949



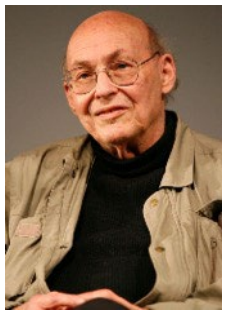
McCulloch-Pitts neuron:
Neurons in the brain can be described by propositional logic

- McCulloch* (neuroscientist) & Pitts* (philosopher/logician) (1943):
  "Logical calculus immanent in nervous activity"
- Konrad Zuse* (1941):
  first programmable computer
- John v. Neumann (1945):
  outline of 1st digital computer architecture
- Norbert Wiener (1948):
  "Cybernetics"

- *Donald Hebb (1949)*
  "The organisation of behaviour":
  Hebbian 'learning' rule:
  "neurons that fire together, wire together"
- Alan Turing (1950)
  "Can a machine think?"

Cognitivism: cognition is function of the brain calculating with symbols (-> "symbolic AI")

- **Cambridge** & DARPA-funded **Dartmouth** conferences
- **Term "artificial intelligence" coined as both a provocation & a program (manifesto character)**
- ***Brain cannot be only described through logic/mathematics, but mental states are identical to functional states based on logical processing of "symbols"***



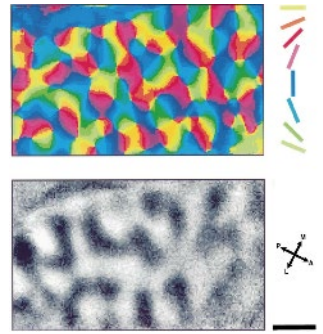
- Hilary Putnam *("Functionalism")*
- Herbert Simon
- Noam Chomsky
- Marvin Minsky*
- John McCarthy
- Claude Shannon
- Jerry Fodor
- John Searle
  ('Chinese room' argument to show that strong AI hypothesis advocated by functionalists/ cognitivists is problematic)

Connectionism: artificial neural networks based on theories of brain self-organisation & emergence of neuronal feature detectors

Brain: self-organising through Hebbian learning -> feature maps, cross-modal interaction … perception, cognition

- Frank Rosen-blatt*
- Marvin Minsky
- Seymour Papert
- Mario Bunge *("Emergentism")*
- *Donald Hebb*
- *David Hubel & Thorsten Wiesel, Richard Held…*

… today



1957: Perceptron machine
- single "neuronal" (retinal) layer trained to recognise visual patterns:
400 photo cell array connected to "neurons" (transistors) with adaptable weights of connectivity (motor-driven potentiometers)

1940-1956

1956 onwards

Huge impact of "cybernetics" phase:
* Folk psychology: computer analogy of brain * cognitivism (subsequent)
* basic computer architecture
* theories of information and complex systems …

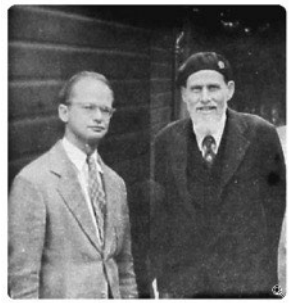
Neural Networks - deep learning

Categorisation of main periods based on F. Varela's "Kognitionswissenschaft – Kognitionstechnik", Suhrkamp Wissenschaft, Frankfurt 1988
See also: Varela, F.J. (1992). Whence Perceptual Meaning? A Cartography of Current Ideas. In: Varela, F.J., Dupuy, JP. (eds) Understanding Origins. Boston Studies in the Philosophy and History of Science, vol 130. Springer, Dordrecht.
https://doi.org/10.1007/978-94-015-8054-0_13
All names mentioned are to be understood as prominent examples; many more philosophers, computer scientists, neuroscientists, linguists etc provided essential contributions.

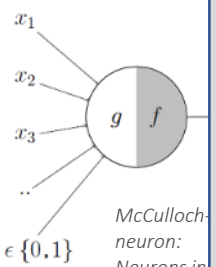
Cybernetics:
a natural science of perception and thought

Cognitivism: cognition is function of the brain calculating with symbols (-> ”symbolic AI”)

Connectionism: artificial neural networks based on theories of brain self-organisation & emergence of neuronal feature detectors



McCulloch [right] and Pitts [left] in 1949



McCulloch-neuron: Neurons in can be des… proposition

- McCulloch* (neuroscientist) & Pitts* (philosopher/logician) (1943): *“Logical calculus immanent in nervous activity”*
- Konrad Zuse* (1941): first programmable computer
- John v. Neumann (1945): outline of 1st digital computer architecture
- Norbert Wiener (1948): *“Cybernetics”*

1940-1956

Huge impact of ”cybernetics” phase:
* Folk psychology: computer analogy of brain * cognitivism (subsequent)
* basic computer architecture
* theories of information and complex systems …

Continuing innovations:

- Huge multimodal data sets
  -> *European Health Data Space*
- Data science progress
- New computing techniques
  -> *quantum computing*

Health is a data-centric discipline

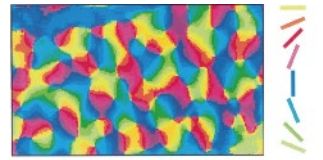
Hebbian learning -> feature maps, …ception, cognition

…ank Rosen-…att*
…arvin Minsky
…ymour Papert
…ario Bunge
…Emergentism”)
…onald Hebb
…avid Hubel &
…orsten Wiesel,
…chard Held…

… today

…trained to

… “neurons” …hts of …iometers)



…eural Networks - deep learning

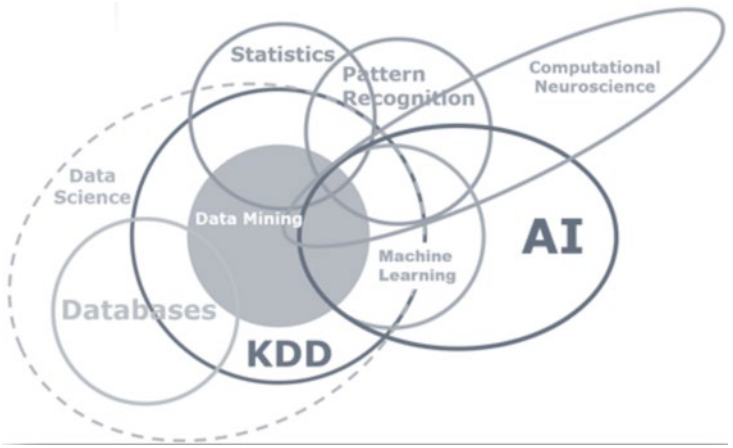
Artificial intelligence

Marvin Minsky: "The science of making machines do things that would require intelligence if done by men"

Interdisciplinary scientific field

Cognitive sciences	Philosophy	Mathematics Statistics	Computer science	…
Neuro-sciences	Linguistics	Cognitive psychology	Data science	

Technology



"Symbolic AI"

Non-adapting algorithms programmed by experts using Boolean rules applied to knowledge bases, e.g.

- *Inference engine*
- *Knowledge representation, uncertainty reasoning, expert system modelling*
- *Fuzzy logic-based approaches*

'Hybrid approaches'

blend of multiple approaches to tackle various real-world issues

- *Data mining, knowledge discovery, advanced analytics*
- *Rule-based modelling and decision-making*
- *Case-based reasoning*
- *Text mining and NLP*
- *Visual analytics, computer vision, pattern recognition*

"Machine learning"

Adapting algorithms that automate analytical model building (e.g. for classification, value prediction, clustering, dimensionality reduction…) by adapting in response to training data

- *Machine learning:*
  - *Supervised*
  - *Unsupervised*
  - *Reinforcement*
- *Neural networks & deep learning*
  - *Deep networks for supervised or discriminatory learning (e.g. CNN, RNN…)*
  - *Deep networks for unsupervised or generative learning (e.g. GAN, self-organising map…)*
  - *…*

Image credit: SAS
https://communities.sas.com/t5/Ask-the-Expert/Intro-to-Machine-Learning-Q-amp-A-Slides-and-On-Demand-Recording/ta-p/713202

Transparency -> Intelligibility, Explainability

Notions of explainability and evaluation approaches for explainable artificial intelligence

Giulia Vilone *, Luca Longo

School of Computer Science, College of Science and Health, Technological University Dublin, Dublin, Republic of Ireland

Information Fusion
https://doi.org/10.1016/j.inffus.2021.05.009

'Not even consensus on the concept of explainability'

The false hope of current approaches to explainable artificial intelligence in health care

Marzyeh Ghassemi, Luke Oakden-Rayner, Andrew L Beam 'Not completely achievable'

The black-box nature of current artificial intelligence (AI) has caused some to question whether AI must be explainable to be used in high-stakes scenarios such as medicine. It has been argued that explainable AI will

Lancet Digit Health 2021; 3: e745-50

Lancet Digital Health 2021

RESEARCH ARTICLE Open Access

Explainability for artificial intelligence in healthcare: a multidisciplinary perspective

Julia Amann[1*], Alessandro Blasimme[1], Effy Vayena[1], Dietmar Frey[2] and Vince I. Madai[2,3] on behalf of the Precise4Q consortium

'Don't give up on it'

BMC 2020
https://doi.org/10.1186/s12911-020-01332-6

...start a few levels lower...?

Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies

Myura Nagendran,[1] Yang Chen,[2] Christopher A Lovejoy,[3] Anthony C Gordon,[1,4] Matthieu Komorowski,[5] Hugh Harvey,[6] Eric J Topol,[7] John P A Ioannidis,[8] Gary S Collins,[9,10] Mahiben Maruthappu[3]

BMJ 2020
http://dx.doi.org/10.1136/bmj.m689

Poor trial design

Review

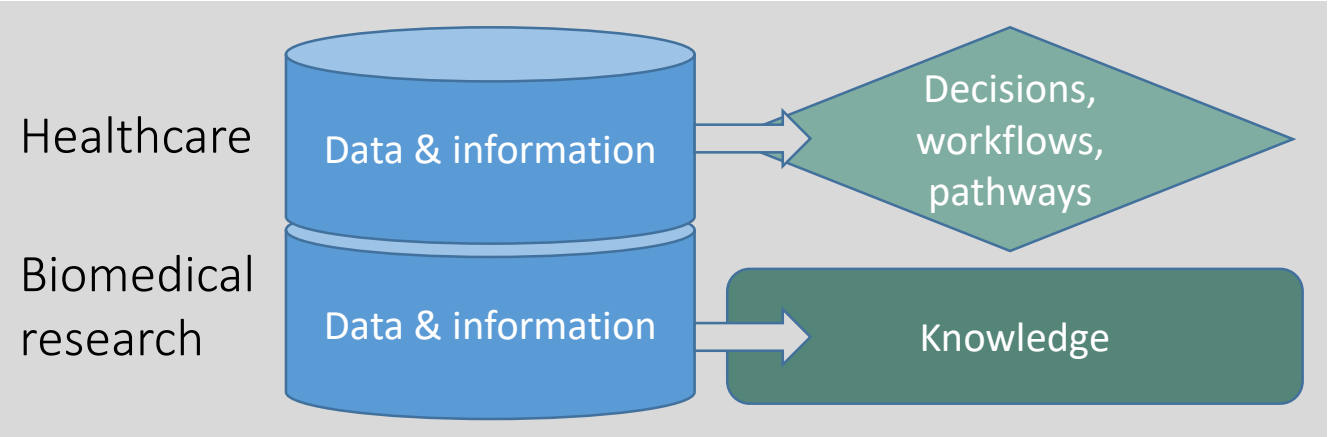
Artificial intelligence for imaging-based COVID-19 detection: Systematic review comparing added value of AI versus human readers

Christine Kriza *, Valeria Amenta, Alexandre Zenié, Dimitris Panidis, Hubert Chassaigne, Patricia Urbán, Uwe Holzwarth, Aisha Vanessa Sauer, Vittorio Reina, Claudius Benedict Griesinger

Inconsistent vocabulary, study design,
not amendable for data pooling / meta-analysis

Eur J Radiol. 10.1016/j.ejrad.2021.110028

AI in medicine and healthcare: many diverse applications



1) Healthcare

- Diagnosis & prediction-based diagnosis
- Clinical care & disease management pathways
- Risk identification, therapy optimisation…
- Active implantable devices, wearables etc.
- Robotic surgery

2) Health systems management

- Administrative workflow
- Logistics, procurement
- Chatbots & virtual nursing assistants
- Telemedicine: care at home

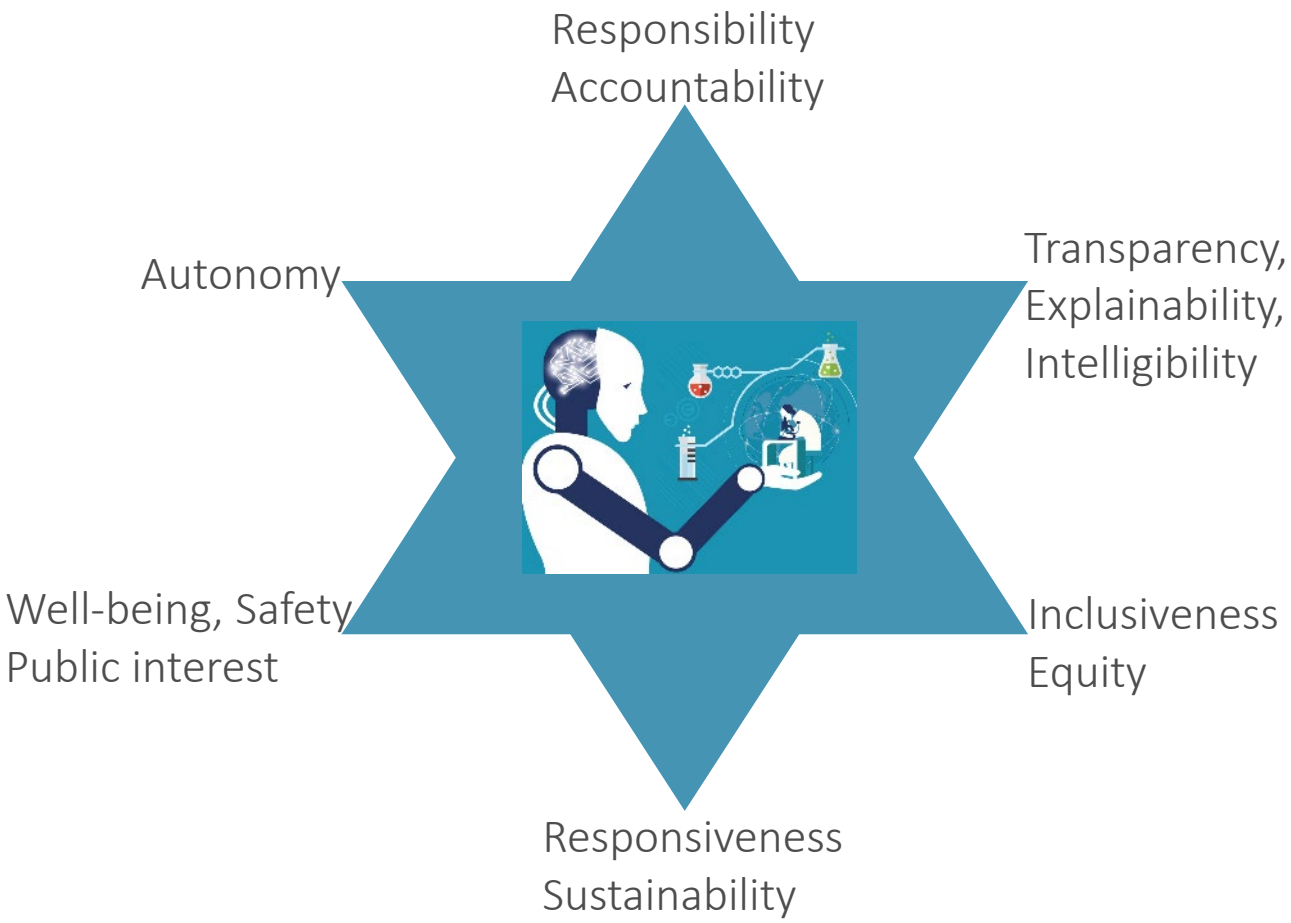
3) Public health & surveillance

- Disease outbreaks monitoring
- Pandemic preparedness
- Health promotion & disease prevention

4) Health research

- Health data for research & development (including AI)
- Electronic health records: optimisation of clinical care
- Drug / Vaccine development & repurposing
- Genomic medicine & personalised medicine
- …

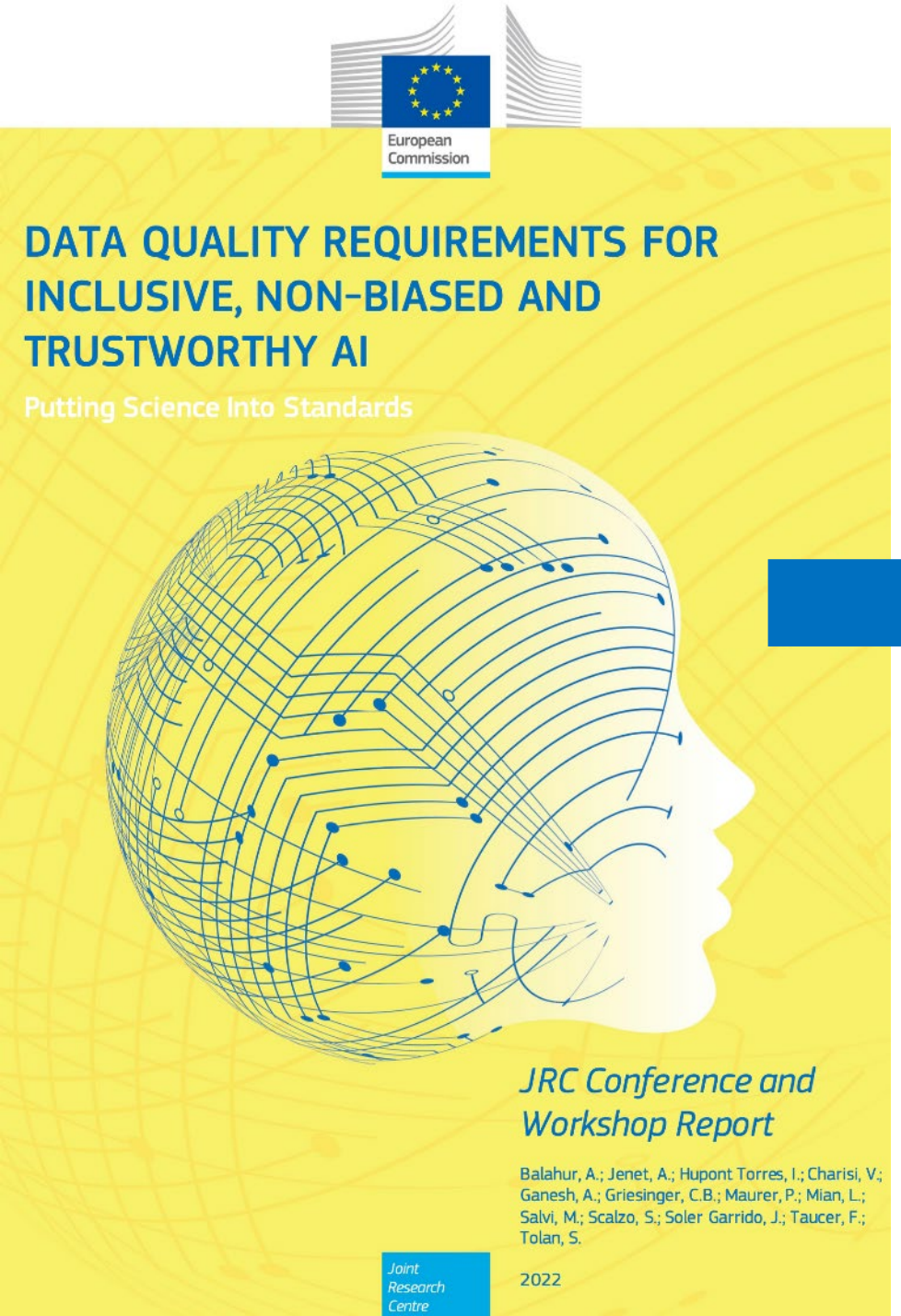
WHO: 6 principles for AI in health



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

European Commission

Artistic illustration: mmustafabozdemir/iStock



Science for standards: data quality for AI

JRC – CEN.CENELEC workshop 2022

Chapter 4.5
Medicine and Healthcare

EU publications

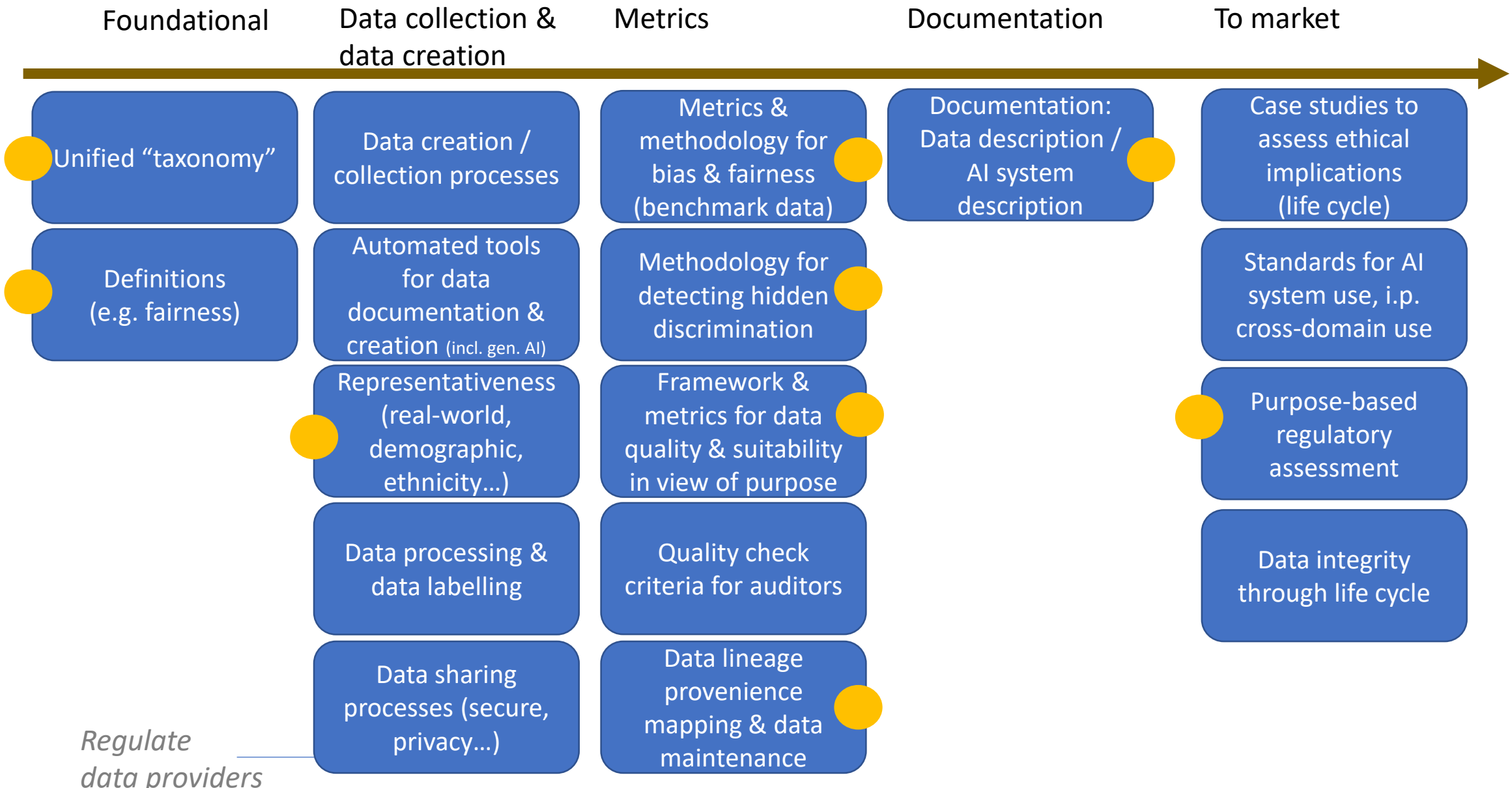
- DOI: https://data.europa.eu/doi/10.2760/365479
- Or publications office website: https://op.europa.eu/en/publication-detail/-/publication/b11a0504-75eb-11ed-9887-01aa75ed71a1

CEN.CENELEC website

- https://www.cencenelec.eu/media/CEN-CENELEC/Events/Events/2022/2022-06-08_PSISevent/workshopreport_psis.pdf

European Commission

Main outcomes of JRC-CEN.CENELEC workshop: *AI in healthcare & medicine*



Study commissioned by DG SANTE (2021)

-> AI uptake in healthcare (not biomedical research!) is slow in the EU

- Absence of harmonized regulatory framework that addresses specificities of AI systems in health
- Lack of appropriate enabling environment for the flourishing of AI
- Lack of trust and transparency

Author affiliations:
- Open Evidence: research & consulting firm (spin-off of Universitat Oberta de Catalunya)
- Ernst & Young consultancy
- Universidad Politécnica de Madrid
  Life Supporting Technologies Group (LifeSTech)

European Commission

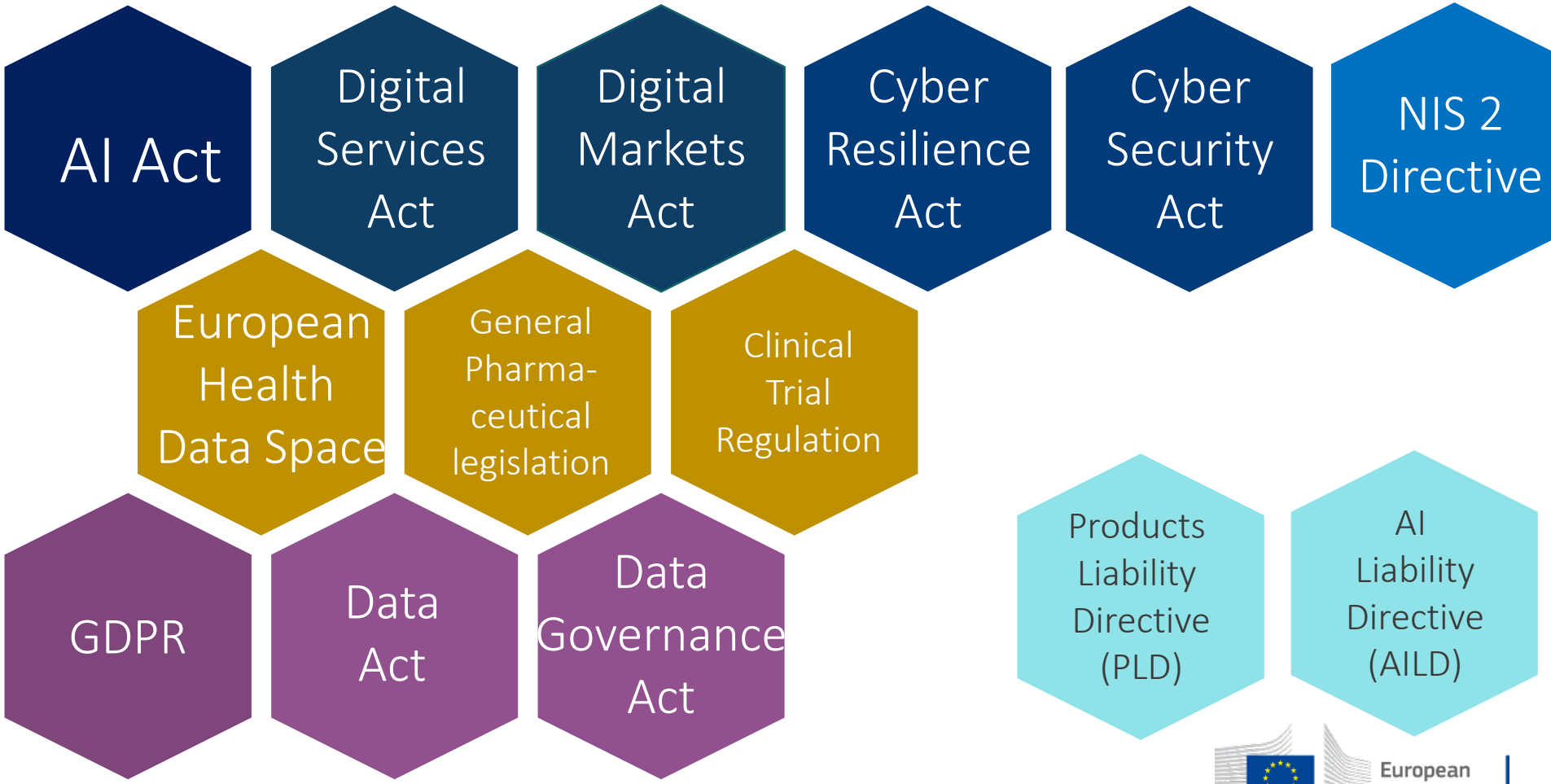
The emerging EU landscape of legislations at the intersection of digital & health / technologies

Digital safety and services in markets

Health data use and pharmaceuticals legislation

Data protection & governance

Liability



European Commission

The emerging EU landscape of legislations at the intersection of digital & health / technologies

Digital safety and services in markets

Health data use and pharmaceuticals legislation

Data protection & governance

Liability

AI Act

Horizontal legislation:
pointing relevant to sectoral legislation for specific requirements

NIS 2 Directive

MDR

IVDR

PPR

IA: e-IFU for medical devices

***Directives**, e.g.*

- *Incorporated medical products, incorporated human blood derivatives, utilisation of tissue of animal origin,*
- *Radio equipment, Machinery*
- *Accessibility requirements for products & services*

AI Liability Directive (AILD)

European Commission

Standardisation for AI applications

Draft Commission mandate to CEN/CENELEC in relation to AI Act

European Standardisation Strategy

CEN/CENELEC Roadmap on AI (2020)

Horizontal standards on trustworthiness
(informed by sectorial needs)

Health | Finance | Transport | Education | ...

Sector-specific guidance and/or standards?

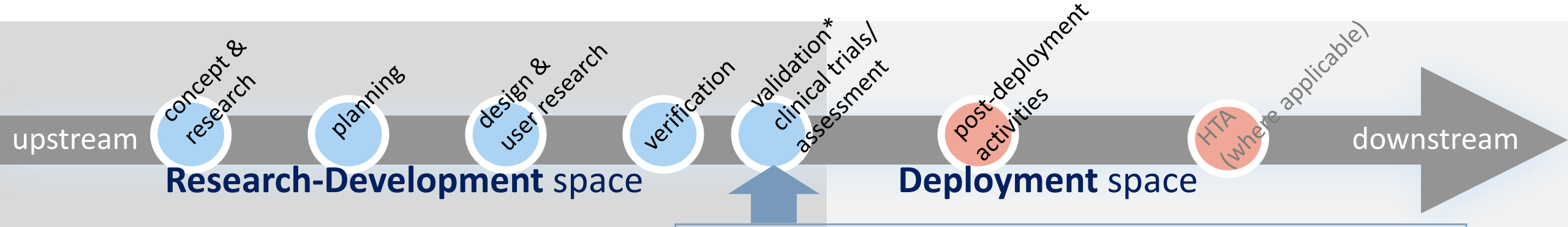
- **ISO/IEC TS 4213** Information technology — Artificial Intelligence — Assessment of ML classification performance
- **ISO/IEC 5259-3** Data quality for analytics and ML — Part 3: Data quality management requirements and guidelines.
- **ISO/IEC 5338** Information technology — Artificial intelligence — AI system life cycle processes.
- **ISO/IEC 5469** Artificial intelligence — Functional safety and AI systems.
- **ISO/IEC 23894-2** Information Technology — Artificial Intelligence — Risk Management
- **ISO/IEC 24027** Information technology — Artificial intelligence (AI) — Bias in AI systems and AI aided decision making.
- **ISO IEC 24029-1** Artificial Intelligence (AI) — Assessment of the robustness of neural networks — Part 1: Overview
- **ISO/IEC 38507** Information technology — Governance of IT — Governance implications of the use of artificial intelligence by organizations
- **ISO/IEC 42001** Information Technology — Artificial intelligence — Management system.

Usefulness for specific needs?

emerging

European Commission

Consensus-based guidelines



The AIMe registry for AI in **biomedical research**

FUTURE-AI: Guiding Principles and Consensus Recommendations for Trustworthy AI in **medical imaging**

IADR E-oral health network (Schwendicke et al.) **dental research**

CLEAR Derm consensus guidelines: checklist for evaluation of **image-based AI reports in dermatology**

CORE-MD recommendations *(EU H2020-funded project)*

CONSORT-AI Reporting of **clinical trial reports** for interventions involving AI

SPIRIT-AI extension: guidelines for **clinical trial protocols** for interventions involving AI

DECIDE-AI Reporting of early-stage **clinical evaluation of decision support systems** driven by AI

equator network — Enhancing the QUAlity and Transparency Of health Research

Upcoming:

*"STARD-AI" * Standards for **Reporting of Diagnostic Accuracy** Study of AI*

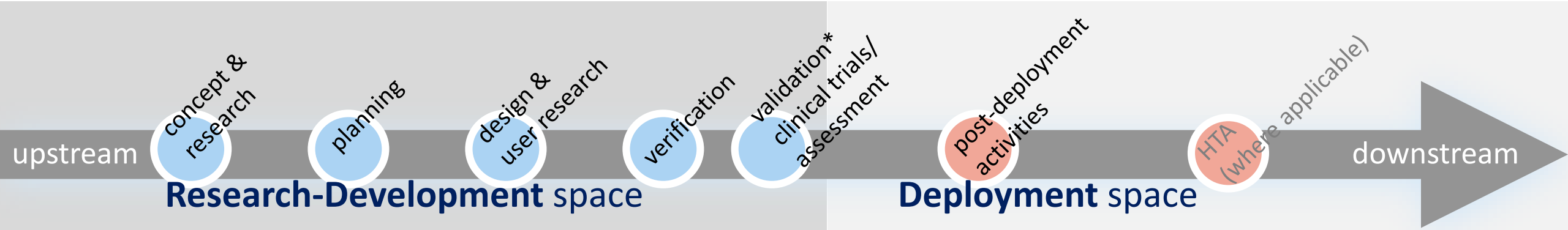
*"PROBAST-AI" * **Prediction model** Risk Of Bias Assessment Tool*

*"TRIPOD-AI" * reporting guideline for **AI-based prediction models** for individual prognosis/diagnosis*

Computer vision in surgery (international collaboration)

See Fraser et al. for synopsis (Table 6)
https://doi.org/10.1080/17434440.2023.2184685

From life cycle to “evidence pathway”



communities

- Developers / designers
- Users, operators, HC professionals
- Health system operators
- Regulators & NBs
- HTA assessors

standards

- Existing ones (e.g. MD software) useful Sufficient though?
- Many AI standards to emerge: ISO + ITU -> CEN/CENELEC
- Process-centric
- Horizontal versus vertical

good science

- Highly distributed information – mainly horizontal – principles but no normative content
- Pragmatic toolbox –provide a pathway for describing AI systems, communicating key elements, evaluating risks and benefits

toolbox

Benefit – Risk pathway

- A pragmatic approach for exploring potential benefits & risks already at the concept stage

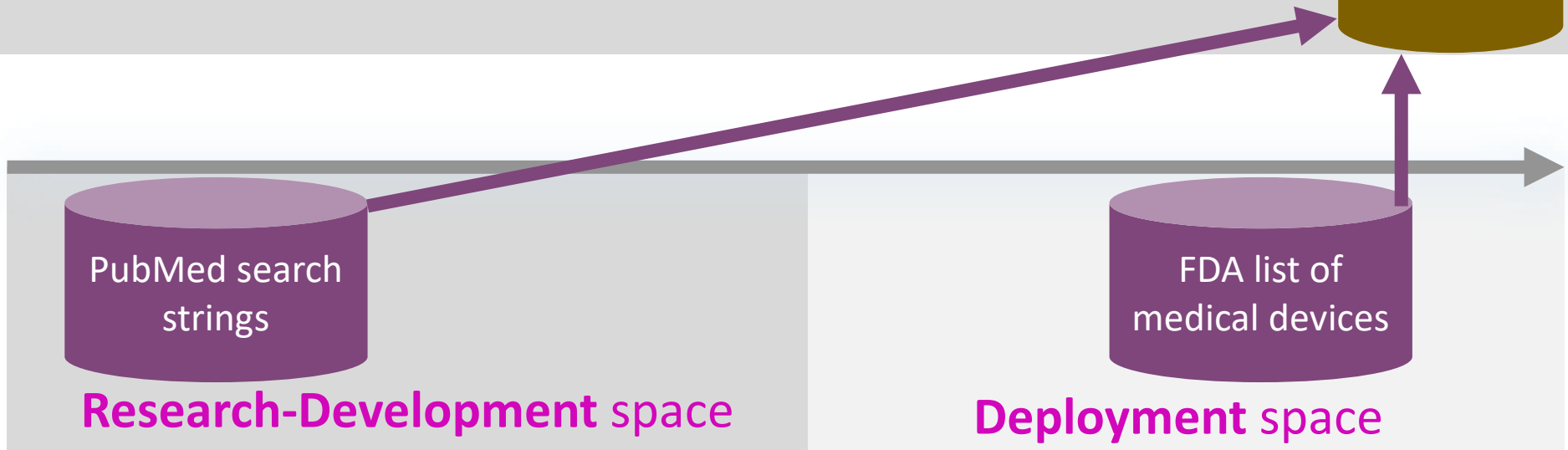
Classification/ Description Matrix

- Classification of AI in the health domain
- Comprehensive description to inform user communities

Ontology

- Terms, relationships, contents, vocabulary – communication!

Case studies



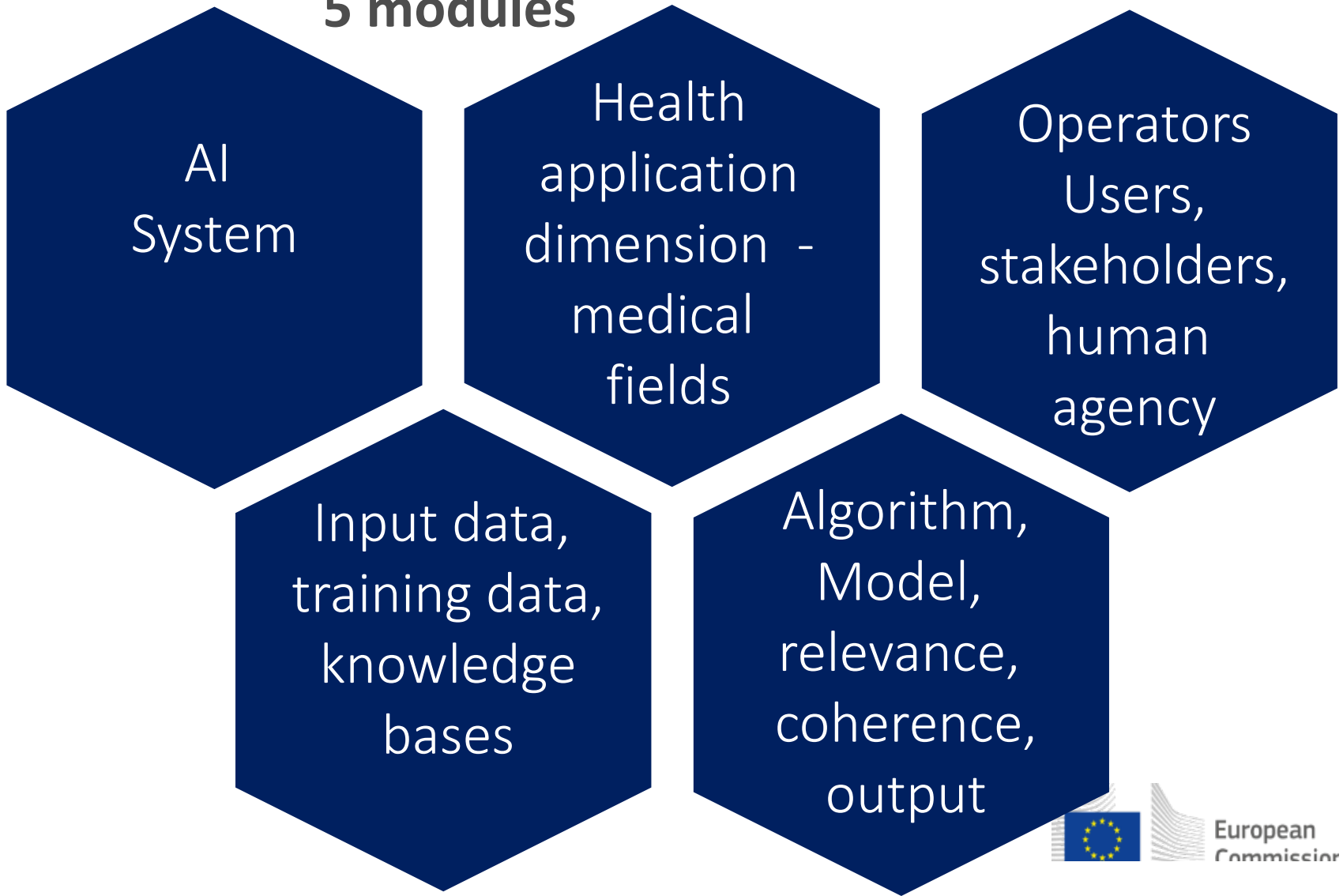
Classification/ description matrix

- Allowing precise description of key elements of an AI system in healthcare and medicine
- Facilitating cross-community knowledge exchange

Sources

- OECD framework for classification of AI systems
- Trustworthy AI High level expert group
- EP STOA paper on AI risks
- EIT taxonomy for AI enabling technologies
- JRC reflections
- IMDRF health impacts (JRC web browser)

5 modules



European Commission

Benefit risk evaluation pathway

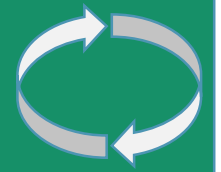
- Charge questions
- Explanations
- References
- Proposals on how to address topics
- Examples for actions

4 stages

1) Gains & benefits

2) Adversity & risks

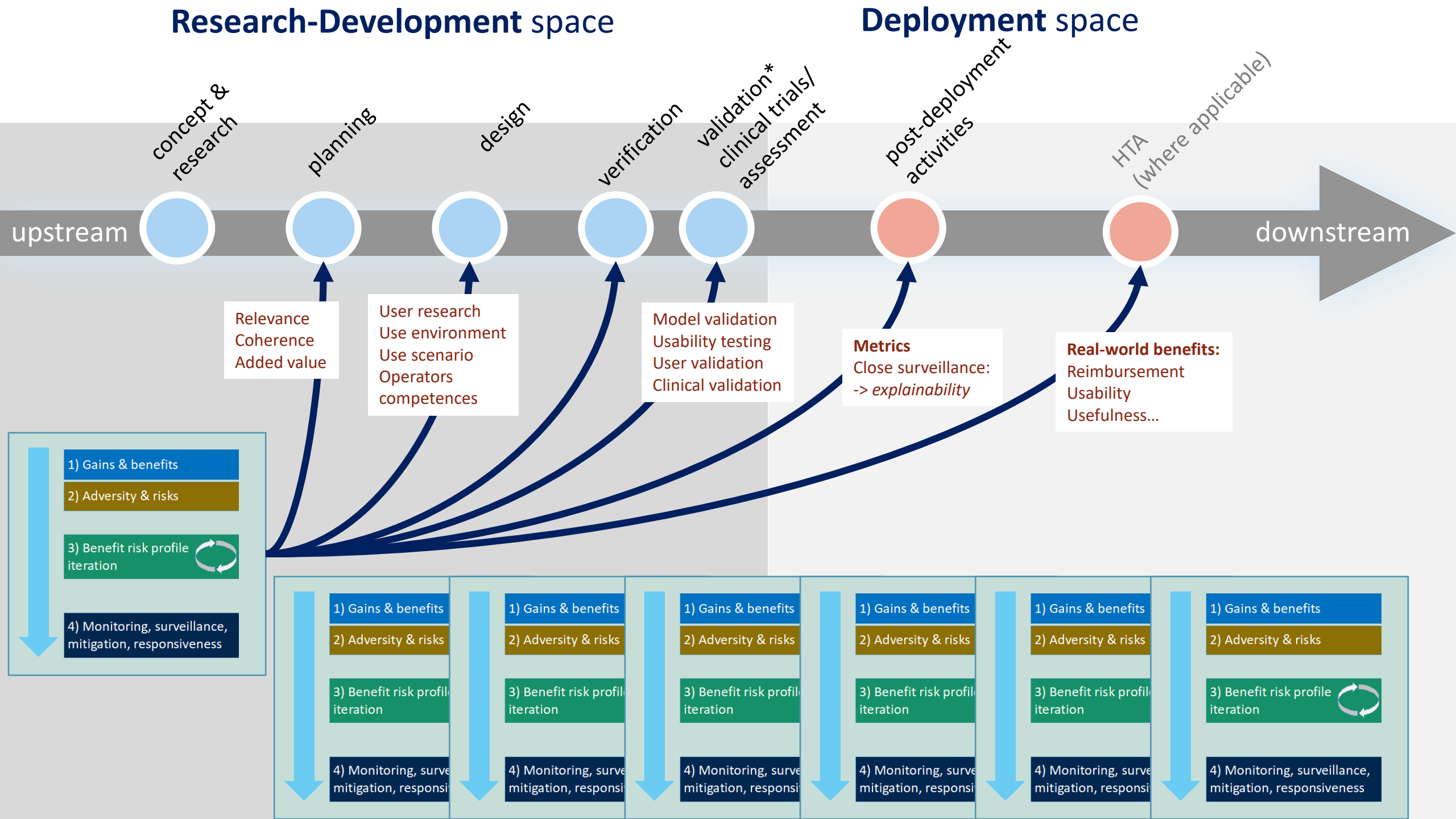
3) Benefit risk profile iteration



4) Monitoring, surveillance, mitigation, responsiveness

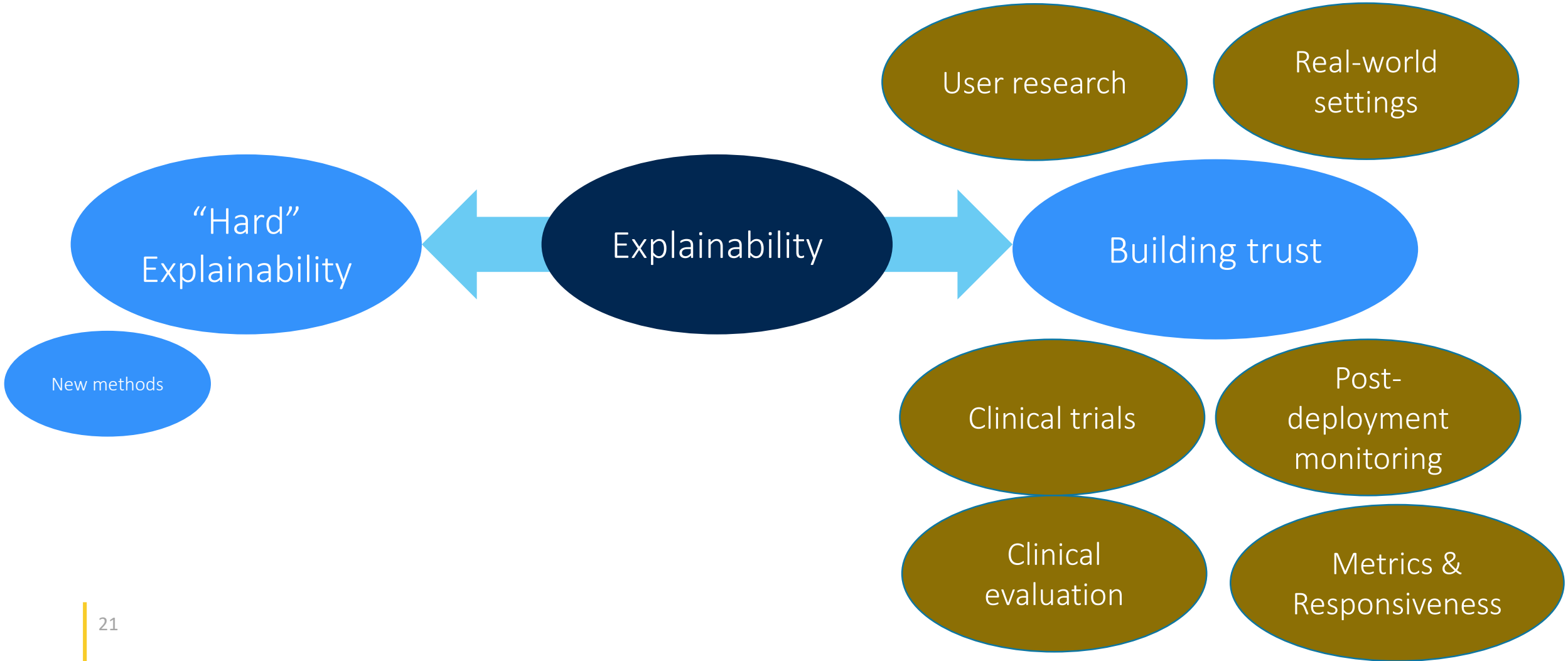
IMDRF health impact terminology set

- Development & production
- Human agency / oversight
- Societal impacts
- Safety, robustness, performance
- Transparency, explainability, intelligibility
- Usability, human factors
- Bias, discrimination, inequity, lack of inclusiveness
- Cybersecurity
- Organisational risks: accountability, liability
- Data: privacy, consent
- Integration into real-world environment


Research-Development space
Deployment space
concept & research
planning
design
verification
validation* clinical trials/ assessment
post-deployment activities
HTA (where applicable)
upstream
downstream
Relevance
Coherence
Added value
User research
Use environment
Use scenario
Operators competences
Model validation
Usability testing
User validation
Clinical validation
Metrics
Close surveillance:
-> explainability
Real-world benefits:
Reimbursement
Usability
Usefulness…
1) Gains & benefits
2) Adversity & risks
3) Benefit risk profile iteration
4) Monitoring, surveillance, mitigation, responsiveness

Benefit risk evaluation pathway & explainability conundrum

Planning ahead



Thank you

claudius.griesinger@ec.europa.eu

Acknowledgements:

Case studies:

Hubert Chassaigne, Alexandre Zenie, Persa Xyderou, Uwe Holzwarth (cancer technologies)

Ontology, Classification/Description matrix, Benefit-Risk pathway*:

Claudius Griesinger, Vittorio Reina (cybersecurity), Hubert Chassaigne, Sandra Coecke*, Thorsten Prinz*#, Hans Wenner*# (cybersecurity),

#) Affiliation T.P. and H.W.: VDE

European Commission

Image credits stated on individual slides, except for slides 3 & 4.
Image credits for these slides:

Orientation & ocular dominance map of the same patch of primary visual cortex from: **Hübener M, Shoham D, Grinvald, Bonhoeffer T (1997)** Spatial relationships among three columnar systems in cat area 17;
J. Neurosci 17(23)9270-9284
Scale bar = 1mm

Image of McCulloch & Pitt:
https://www.semanticscholar.org/paper/On-the-legacy-of-W.S.-McCulloch-Moreno-D%C3%ADaz-Moreno-D%C3%ADaz/8056242a82ecc5e0064d4ff187fb07c5853fe8a6

Image of Konrad Zuse:
ETH Zürich Library. Reproduced from: Copeland, B.J., Sommaruga, G. (2015). The Stored-Program Universal Computer: Did Zuse Anticipate Turing and von Neumann?. In: Sommaruga, G., Strahm, T. (eds) Turing's Revolution. Birkhäuser, Cham.
https://doi.org/10.1007/978-3-319-22156-4_3
https://link.springer.com/chapter/10.1007/978-3-319-22156-4_3

Image of F. Rosenblatt from:
https://news.cornell.edu/stories/2019/09/professors-perceptron-paved-way-ai-60-years-too-soon

Image of Marvin Minsky:
Amy Sussman/Getty Images for Tribeca Film Festival
https://edition.cnn.com/2016/01/26/us/marvin-minsky-obit-feat/index.html