Future-ai

FUTURE-AI: International consensus guideline for trustworthy and deployable artificial intelligence in healthcare

KARIM LEKADIR*†, Universitat de Barcelona, Spain
AASA FERAGEN, Technical University of Denmark, Denmark
ABDUL JOSEPH FOFANAH, Milton Margai Technical University, Sierra Leone
ALEJANDRO F FRANGI, University of Leeds, United Kingdom and KU Leuven, Belgium
ALENA BUYX, Technical University of Munich, Germany
ANAIS EMELIE, Universitat de Barcelona, Spain
ANDREA LARA, Galileo University, Guatemala
ANTONIO R PORRAS, University of Colorado Anschutz Medical Campus, United States
AN-WEN CHAN, University of Toronto, Canada
ARCADI NAVARRO, Universitat Pompeu Fabra, Spain and Pasqual Maragall Foundation, Spain
BEN GLOCKER, Imperial College London, United Kingdom
BENARD O BOTWE, University of Ghana, Ghana and University of London, UK
BISHESH KHANAL, NepAl Applied Mathematics and Informatics Institute for Research (NAAMII), Nepal
BRIGIT BEGER, European Heart Network, Belgium
CAROL C WU, University of Texas MD Anderson Cancer Center, United States
CELIA CINTAS, IBM Research Africa, Kenya
CURTIS P LANGLOTZ, Stanford University School of Medicine, United States
DANIEL RUECKERT, Technical University Munich, Germany and Imperial College London, UK
DEOGRATIAS MZURIKWAO, Muhimbili University of Health and Allied Sciences, Tanzania
DIMITRIOS I FOTIADIS, Foundation for Research and Technology - Hellas (FORTH), Greece
DOSZHAN ZHUSSUPOV, Almaty AI Lab, Kazakhstan
ENZO FERRANTE, Universidad Nacional del Litoral, Argentina
ERIK MEIJERING, University of New South Wales, Australia
EVA WEICKEN, Fraunhofer Heinrich Hertz Institute, Germany
FABIO A GONZÁLEZ, Universidad Nacional de Colombia, Colombia
FOLKERT W ASSELBERGS, University of Amsterdam, The Netherlands and University College London, UK
FRED PRIOR, University of Arkansas for Medical Sciences, United States
GABRIEL P KRESTIN, Erasmus MC University Medical Center, the Netherlands
GARY COLLINS, University of Oxford, UK
GELETAW S TEGENAW, Jimma University, Ethiopia
GEORGIOS KAISSIS, Technical University Munich, Germany
GIANLUCA MISURACA, Universidad Politécnica de Madrid, Spain
GIANNA TSAKOU, Research and Development Lab, Greece
GIRISH DWIVEDI, The University of Western Australia, Australia
HARIDIMOS KONDYLAKIS, Hellenic Mediterranean University, Greece
HARSHA JAYAKODY, University of Colombo, Sri Lanka
HENRY C WOODRUF, Maastricht University, the Netherlands
HUGO JWL AERTS, Harvard Medical School, United States
IAN WALSH, Technology and Research (A*STAR), Singapore
IOANNA CHOUVARDA, Aristotle University of Thessaloniki, Greece
IRÈNE BUVAT, Inserm, France
ISLEM REKIK, Imperial College London, UK and Istanbul Technical University, Turkey

PETER GORDEBEKE, European Institute for Biomedical Imaging Research, Austria
PHILIPPE LAMBIN, Maastricht University, the Netherlands
PIETA BROWN, Orion Health, New Zealand
PURANG ABOLMAESUMI, the University of British Columbia, Canada
QI DOU, The Chinese University of Hong Kong, China
QINGHUA LU, Data61, The Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia
RICHARD OSUALA, Universitat de Barcelona, Spain
ROSE NAKASI, Makerere University, Uganda
S KEVIN ZHOU, University of Science and Technology of China, China
SANDY NAPEL, Stanford University, United States
SARA COLANTONIO, Institute of Information Science and Technologies of the National Research Council of Italy, Italy
SHADI ALBARQOUNI, University Hospital Bonn, Germany
SMRITI JOSHI, Universitat de Barcelona, Spain
STACY CARTER, University of Wollongong, Australia
STEFAN KLEIN, Erasmus MC University Medical Center, the Netherlands
STEFFEN E PETERSEN, Queen Mary University of London, UK
SUSANNA AUSSÓ, TIC Salut Social Foundation, Spain
SUYASH AWATE, Indian Institute of Technology Bombay, India
TAMMY RIKLIN RAVIV, Ben-Gurion University, Israel
TESSA COOK, University of Pennsylvania, United States
TINASHE E M MUTSVANGWA, University of Cape Town, South Africa
WENDY A ROGERS, Macquarie University, Australia
WIRO J NIESSEN, Erasmus MC University Medical Center, the Netherlands
XÈNIA PUIG-BOSCH, Universitat de Barcelona, Spain
YI ZENG, Chinese Academy of Sciences, China
YUNUSA G MOHAMMED, Gombe State University, Nigeria
YVES SAINT JAMES AQUINO, University of Wollongong, Australia
ZOHAIB SALAHUDDIN, Maastricht University, the Netherlands
MARTIJN P A STARMANS, Erasmus MC University Medical Center, the Netherlands

*Correspondence author.
†The complete list of affiliations can be found in Appendix B

.

Abstract:
Background: Despite major advances in artificial intelligence (AI) for medicine and healthcare, the deployment and adoption of AI technologies remain limited in real-world clinical practice. In recent years, concerns have been raised about the technical, clinical, ethical and legal risks associated with medical AI. To increase adoption in the real world, it is essential that medical AI tools are trusted and accepted by patients, clinicians, health organisations and authorities. This paper describes the FUTURE-AI guideline as the first international consensus framework for guiding the development and deployment of trustworthy AI tools in healthcare.
Methods: The FUTURE-AI consortium was founded in 2021 and currently comprises 118 inter-disciplinary experts from 51 countries representing all continents, including AI scientists, clinicians, ethicists, and social scientists. Over a two-year period, the consortium defined guiding principles and best practices for trustworthy AI through an iterative process comprising an in-depth literature review, a modified Delphi survey, and online consensus meetings.
Findings: The FUTURE-AI framework was established based on six guiding principles for trustworthy AI in healthcare, i.e. Fairness, Universality, Traceability, Usability, Robustness and Explainability. Through consensus, a set of 28 best practices were defined, addressing technical, clinical, legal and socio-ethical dimensions of trustworthy AI. The recommendations cover the entire lifecycle of medical AI, from design, development and validation to regulation, deployment, and monitoring. Interpretation: FUTURE-AI is a risk-informed, assumption-free guideline which provides a structured approach for constructing medical AI tools that will be trusted, deployed and adopted in real-world practice. Researchers are encouraged to take the recommendations into account in proof-of-concept stages to facilitate future translation towards clinical practice of medical AI.
Funding: Support for this work was partially provided by the European Union's Horizon 2020 under Grant Agreement No. 952103 (EuCanImage), No.952159 (ProCAncer-I), No.952172
(CHAIMELEON), No. 826494 (PRIMAGE) and No. 952179 (INCISIVE).

1 INTRODUCTION

Despite major advances in the field of medical AI, the deployment and adoption of AI technologies remain limited in real-world clinical practice. In recent years, concerns have been raised about the technical, clinical, ethical and social risks associated with medical AI [45, 79]. In particular, existing research has shown that medical AI tools can be prone to errors and patient harm, biases and increased health inequalities, lack of transparency and accountability, as well as data privacy and security breaches [13, 14, 32, 33, 60].

To increase adoption in the real world, it is essential that medical AI tools are trusted and accepted by patients, clinicians, health organisations and authorities. However, there is an absence of clear, widely accepted guidelines on how medical AI tools should be designed, developed, evaluated and deployed to be trustworthy, i.e. technically robust, clinically safe, ethically sound and legally compliant. To have a real impact at scale, such guidelines for trustworthy and responsible AI must be obtained through wide consensus involving international and inter-disciplinary experts.

In other domains, international consensus guidelines have made lasting impacts. For example, the FAIR guideline [81] for data management has been widely adopted by researchers, organisations and authorities, as they provided a logical framework for standardising and enhancing the tasks of data collection, curation, organisation and storage. While it can be argued that the FAIR principles do not cover every aspect of data management, as they focus more on findability, accessibility, interoperability and reusability of the data, and less on privacy and security, they delivered a code of practice that is now widely accepted and applied.

For medical AI, initial efforts have focused on providing recommendations for the reporting of AI studies for different medical domains or clinical tasks (e.g. TRIPOD-AI [16], CLAIM [58], CONSORT-AI [49], DECIDE-AI [77], PROBAST-AI [16], CLEAR [39]). These guidelines do not provide best practices for the actual development and deployment of the AI tools but promote standardised and complete reporting of their development and evaluation. Recently, several researchers have published promising ideas on possible best practices for medical AI [11, 42, 51, 64, 67, 80]. However, these proposals have not been established through wide international consensus and do not cover the whole lifecycle of medical AI (i.e. from design, development and evaluation to deployment, usage and monitoring). In 2020, a comprehensive self-assessment checklist for trustworthy AI was defined through consensus by Europe's High-Level Expert Group on Artificial Intelligence, but it covered AI in general and did not address the specific risks and challenges of AI in medicine and healthcare [3].

This paper addresses an important gap in the field of medical AI, by delivering the very first international, consensus guideline for trustworthy medical AI that covers the entire AI lifecycle (Figure 1).

The FUTURE-AI consortium was initiated in 2021 and currently comprises 118 international and inter-disciplinary experts from 51 countries (Figure 1), representing all continents (Europe, North America, South America, Asia, Africa, and Oceania). Additionally, the members represent a variety of disciplines (e.g. data science, medical research, healthcare, computer engineering, medical ethics, social sciences) and data domains (e.g. radiology, genomics, mobile health, electronic health records, surgery, pathology). To develop the FUTURE-AI framework, we drew inspiration from the FAIR principles for data management and defined concise recommendations structured according to six guiding principles, i.e. Fairness, Universality, Traceability, Usability, Robustness, and Explainability (Figure 2).

Fig. 1. Geographical distribution of the multi-disciplinary experts.

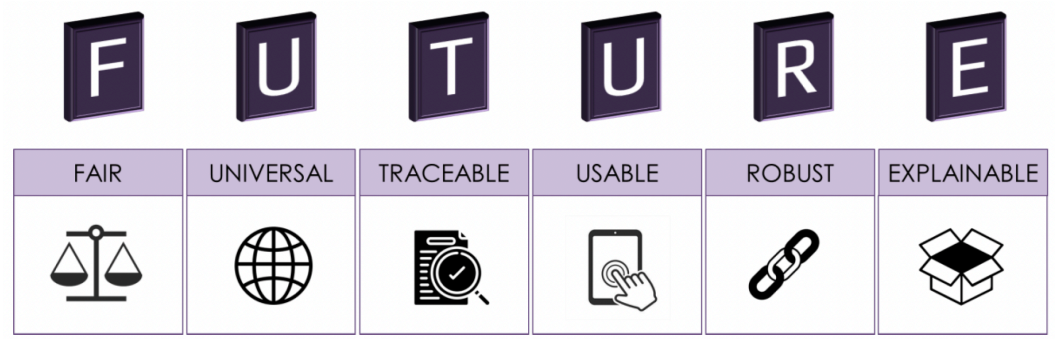

Fig. 2. Organisation of the FUTURE-AI framework for trustworthy medical AI according to six guiding principles, i.e. Fairness, Universality, Traceability, Usability, Robustness and Explainability.

2 MATERIALS AND METHODS

The FUTURE-AI framework was defined over a 24-month period through a modified Delphi approach (Table 1).

FUTURE-AI was initiated with an in-depth literature review on medical AI and trustworthiness, which resulted in the identification of key dimensions of relevance to trustworthy AI, including robustness, safety, security, fairness, transparency, traceability, accountability, generalisability, explainability, usability and responsible AI. To facilitate the use of the framework, some keywords were grouped, selected and re-ordered to obtain a reduced set of six guiding principles (Fairness, Universality, Traceability, Usability, Robustness, Explainability), which form the basis of the FUTURE-AI acronym.

Working groups composed of three experts each (including clinicians, data scientists and computer engineers) explored the six principles separately and proposed an initial set of best practices, by using AI for medical imaging as an initial use case. Furthermore, the working groups discussed

Table 1. Summary of the methodology and timeline for establishing the FUTURE-AI guideline.

	Step	Period
1	Founding of the FUTURE-AI consortium by members of research projects in Europe [18]	June 2021
2	Creation of working groups for each of the six guiding principles	July 2021
3	Proposal of a first set of 55 recommendations through a use-case driven approach focused on AI in medical imaging	September 2021
4	Feedback gathering through a survey with >100 international and multi-disciplinary experts from all continents	November 2021 – March 2022
5	Analysis of results and derivation of a more concise list of 22 recommendations, generalised to AI for healthcare	March 2022 – April 2022
6	Second round of feedback from the experts	May 2022 – July 2022
7	Analysis of results and derivation of an improved and extended list of 30 recommendations	June 2022 – July 2022
8	Third round of feedback, by focusing on the main disagreements and the manuscript's first draft	September 2022 – February 2023
9	Four consensus online meetings to discuss remaining disagreements, resulting in the final 28 recommendations	June 2023
10	Finalisation and presentation of the FUTURE-AI consensus guideline in a journal publication	June 2023

the proposed recommendations, and removed overlaps and redundancies across the six principles, resulting in a first comprehensive set of 54 recommendations.

Subsequently, the FUTURE-AI consortium members provided systematic feedback on the first version of the FUTURE-AI guideline through a comprehensive survey. The experts could comment on each recommendation, rate its importance, propose missing recommendations, and provide additional feedback in free text. Based on the results of the survey, the list of recommendations was deemed too extensive, and was thus substantially reduced from 54 to 22 recommendations, while the scope was carefully broadened from AI for medical imaging to AI for healthcare.

The reduced set of recommendations was sent out to the FUTURE-AI consortium, together with a list of major disagreements that arose, for a second round of feedback and comments. This step resulted in a new version consisting of 30 recommendations, with a new "General" category in addition to the six guiding principles to account for recommendations that are transversal across all the dimensions of trustworthy AI.

Based on the machine learning technology readiness level (ML-TRL) [44], the FUTURE-AI guideline was refined by distinguishing between medical AI tools at the research or proof-of-concept stage (i.e. ML-TRL 1 to 4) and those intended for clinical deployment (i.e. ML-TRL 5 to 9), as they require different levels of compliance. Hence, we asked the members of the consortium to rate the recommendations as recommended vs. highly recommended, for both proof-of-concept (low ML-TRL) and deployable AI tools (high ML-TRL). Finally, iterative discussions on the guideline, disagreements and manuscript were held, including during four dedicated online meetings, resulting in a final set of 28 consensus recommendations, which are listed in Table 2.

3 FUTURE-AI GUIDELINE

In this section, we provide definitions and justifications for each of the six guiding principles and give an overview of the FUTURE-AI recommendations. Table 2 provides a summary of the recommendations, together with the proposed level of compliance (i.e. recommended vs. highly recommended). More detailed descriptions are provided in Table 3 in the Appendix for readers who

Table 2. List of the FUTURE-AI recommendations, together with the expected compliance for both proof-of-concept (Low ML-TRL) and deployable (High ML-TRL) AI tools (+: Recommended, ++: Highly recommended).

		Recommendations	Low ML-TRL	High ML-TRL
F	1	Define any potential sources of bias from an early stage	++	++
	2	Collect data on individuals' attributes, when possible	+	+
	3	Evaluate potential biases and bias correction measures	+	+
U	1	Define intended clinical settings and cross-setting variations	++	++
	2	Use community-defined standards (e.g. clinical definitions, technical standards)	+	+
	3	Evaluate using external datasets and/or multiple sites	++	++
	4	Evaluate and demonstrate local clinical validity	+	++
T	1	Implement a risk management process throughout the AI lifecycle	+	++
	2	Provide documentation (e.g. technical, clinical)	++	++
	3	Define mechanisms for quality control of the AI inputs and outputs	+	++
	4	Implement a system for periodic auditing and updating	+	++
	5	Implement a logging system for usage recording	+	++
	6	Establish mechanisms for human oversight and governance		
U	1	Define intended use and user requirements from an early stage	++	++
	2	Provide training materials and activities (e.g. tutorials, hands-on sessions)	+	++
	3	Evaluate user experience and acceptance with independent end-users	+	++
	4	Evaluate clinical utility and safety (e.g. effectiveness, harm, cost-benefit)	+	++
R	1	Define sources of data variation from an early stage	++	++
	2	Train with representative real-world data	++	++
	3	Evaluate and optimise robustness against real-world variations	++	++
E	1	Define the need and requirements for explainability with end-users	++	++
	2	Evaluate explainability with end-users (e.g. correctness, impact on users)	+	+
General	1	Engage inter-disciplinary stakeholders throughout the AI lifecycle	++	++
	2	Implement measures for data privacy and security	++	++
	3	Define adequate evaluation plan (e.g. datasets, metrics, reference methods)	++	++
	4	Identify and comply with applicable AI regulatory requirements	+	++
	5	Investigate and address ethical issues	+	++
	6	Investigate and address social and societal issues	+	+

may require more information on any recommendation(s). Note that a glossary of the main terms used in this paper is provided in Table 4 in the Appendix, while the main stakeholders of relevance to the FUTURE-AI framework are listed in Table 5 in the Appendix.

3.1 Fairness

The Fairness principle states that medical AI tools should maintain the same performance across individuals and groups of individuals (including under-represented and disadvantaged groups). AI-driven medical care should be provided equally for all citizens, independently of their sex, gender, ethnicity, age, socio-economic status and (dis)abilities, among other attributes. Fair medical AI tools should be developed such that potential AI biases are minimised as much as possible, or identified and reported.

To this end, three recommendations for Fairness are defined in the FUTURE-AI framework. First, AI developers together with domain experts should define fairness for their specific use case and make an inventory of potential sources of bias (Fairness 1). Accordingly, to facilitate verification of AI fairness and non-discrimination, information on the subjects' relevant attributes should be included in the datasets (Fairness 2). Finally, whenever this data is available, the development team should apply bias detection and correction methods, to obtain the best possible trade-off between fairness and accuracy (Fairness 3).

3.2 Universality

The Universality principle states that a medical AI tool should be generalisable outside the controlled environment where it was built. Specifically, the AI tool should be able to generalise to new patients and new users (e.g. new clinicians), and when applicable, to new clinical sites. Depending on the intended radius of application, medical AI tools should be as interoperable and as transferable as possible, so they can benefit citizens and clinicians at scale. To this end, four recommendations for Universality are defined in the FUTURE-AI framework. First, the AI developers should define the requirements for universality, i.e. the radius of application of their medical AI tool (e.g. clinical centres, countries, clinical settings), and accordingly anticipate any potential obstacles to universality, such as differences in clinical workflows, medical equipment or digital infrastructures (Universality 1). To enhance interoperability, development teams should favour the use of established community-defined standards (e.g. clinical definitions, medical ontologies, data annotations, technical standards) throughout the AI tool's production lifetime (Universality 2). To enhance generalisability, the medical AI tool should be tested with external datasets and, when applicable, across multiple sites (Universality 3). Finally, medical AI tools should be evaluated for their local clinical validity, and if necessary, calibrated so they perform well given the local populations and local clinical workflows (Universality 4).

3.3 Traceability

The Traceability principle states that medical AI tools should be developed together with mechanisms for documenting and monitoring the complete trajectory of the AI tool, from development and validation to deployment and usage. This will increase transparency and accountability by providing detailed and continuous information on the AI tools during their lifetime to clinicians, healthcare organisations, citizens and patients, AI developers and relevant authorities. AI traceability will also enable continuous auditing of AI models [62], identify risks and limitations, and update the AI models when needed. To this end, six recommendations for Traceability are defined in the FUTURE-AI framework. First, a system for risk management should be implemented throughout the AI lifecycle, including risk identification, assessment, mitigation, monitoring and reporting (Traceability 1). To increase transparency, relevant documentation should be provided for the stakeholder groups of interest, including AI information leaflets, technical documentation, and/or scientific publications (Traceability 2). After deployment, continuous quality control of AI inputs and outputs should be implemented, to identify inconsistent input data and implausible AI outputs (e.g. using uncertainty estimation), and to implement necessary model updates (Traceability 3). Furthermore, periodic auditing and updating of AI tools should be implemented (e.g. yearly) to detect and address any potential issue or performance degradation (Traceability 4). To increase traceability and accountability, an AI logging system should be implemented to keep a record of the usage of the AI tool, including for instance, user actions, accessed and used datasets, and identified issues (Traceability 5). Finally, mechanisms for human oversight and governance should be implemented, to enable selected users to flag AI errors or risks, overrule AI decisions, use

human judgment instead, assign roles and responsibilities, and maintain the AI system over time (Traceability 6).

3.4 Usability

The Usability principle states that the end-users should be able to use a medical AI tool to achieve a clinical goal efficiently and safely in their real-world environment. On one hand, this means that end-users should be able to use the AI tool's functionalities and interfaces easily and with minimal errors. On the other hand, the AI tool should be clinically useful and safe, e.g. improve the clinicians' productivity and/or lead to better health outcomes for the patients and avoid harm.

To this end, four recommendations for Usability are defined in the FUTURE-AI framework. First, through a human-centred approach, target end-users (e.g. general practitioners, specialists, nurses, patients, hospital managers) should be engaged from an early stage to define the AI tool's intended use, user requirements and human-AI interfaces (Usability 1). Second, training materials and training activities should be provided for all intended end-users, to ensure adequate usage of the AI tool, minimise errors and thus patient harm, and increase AI literacy (Usability 2). At the evaluation stage, the usability within the local clinical workflows, including human factors that may impact the usage of the AI tool [72] (e.g. satisfaction, confidence, ergonomics, learnability), should be assessed with representative and diverse end-users (Usability 3). Furthermore, the clinical utility and safety of the AI tools should be evaluated and compared with the current standard of care, to estimate benefits as well as potential harms for the citizens, clinicians and/or health organisations (Usability 4).

3.5 Robustness

The Robustness principle refers to the ability of a medical AI tool to maintain its performance and accuracy under expected or unexpected variations in the input data. Existing research has shown that even small, imperceptible variations in the input data may lead AI models into incorrect decisions [23]. Biomedical and health data can be subject to significant variations in the real world (both expected and unexpected), which can affect the performance of AI tools. Hence, it is important that medical AI tools are designed and developed to be robust against real-world variations, as well as evaluated and optimised accordingly.

To this end, three recommendations for Robustness are defined in the FUTURE-AI framework. At the design phase, the development team should first define robustness requirements for the medical AI application in question, by making an inventory of the potential sources of variation e.g. data-, equipment-, clinician-, patient- and centre-related variations (Robustness 1). Accordingly, the training datasets should be carefully selected, analysed and enriched to reflect these real-world variations as much as possible (Robustness 2). Subsequently, the robustness of the AI tool, as well as measures to enhance robustness, should be iteratively evaluated under conditions that reflect the variations of real-world clinical practice (Robustness 3).

3.6 Explainability

The Explainability principle states that medical AI tools should provide clinically meaningful information about the logic behind the AI decisions. While medicine is a high-stake discipline that requires transparency, reliability and accountability, machine learning techniques often produce complex models which are black boxes in nature. Explainability is considered desirable from a technological, medical, ethical, legal as well as patient perspective [4]. Explainability is a complex task which has challenges that need to be carefully addressed during AI development and evaluation to ensure that AI explanations are clinically meaningful and beneficial to the end-users [29].

To this end, two recommendations for Explainability are defined in the FUTURE-AI framework. At the design phase, it should be first established with end-users and domain experts whether explainable AI is needed for the medical AI tool in questions. In this case, the specific goal and approaches for explainability should be defined (Explainability 1). After their implementation, the selected approaches for explainability should be evaluated, both quantitatively using in silico methods [34], as well qualitatively with end-users to assess their impact on the user's satisfaction and performance (Explainability 2).

3.7 General recommendations

Finally, six general recommendations are defined in the FUTURE-AI framework, which apply across all principles of trustworthy AI in healthcare. First, AI developers should actively engage interdisciplinary stakeholders throughout the production lifecycle, including healthcare professionals, patient representatives, ethicists and social scientists, data managers and legal experts (General 1). During the whole lifecycle from development to deployment, adequate measures should be put in place to ensure data protection and security, such as data de-identification and minimisation, privacy-enhancing techniques, and defences against malicious attacks (General 2). During all evaluation tasks, appropriate evaluation datasets, metrics and reference methods should be carefully selected to gather strong evidence on the medical AI tool's trustworthiness (General 3). The AI development teams should verify and understand the applicable AI regulations from an early stage, so they can anticipate and meet their legal obligations (General 4). All general and application-specific ethical issues should be investigated, discussed and integrated into the practical development of the AI tool, through continuous interactions with domain specialists and ethicists [53] (General 5). Finally, to ensure a positive impact on citizens and society, social and societal issues should be investigated and addressed (e.g. the tool's impact on working conditions, relationships between citizens and health services, upskilling or deskilling of citizens and healthcare professionals [5], environmental sustainability) (General 6).

4 DISCUSSION

Despite the tremendous amount of research in medical AI in recent years, currently, only a limited number of AI tools have made the transition to clinical practice. While many studies have demonstrated the huge potential of AI to improve healthcare, significant clinical, technical, socio-ethical and legal challenges persist.

In this paper, we presented the results of an international effort to establish a consensus guideline for developing trustworthy and deployable AI tools in healthcare. Through an iterative process that lasted 24 months, the FUTURE-AI framework was established, comprising a well-structured, self-contained set of 28 recommendations, which covers the whole lifecycle of medical AI. By dividing the recommendations across six guiding principles, the pathways towards trustworthy AI are clearly characterised to facilitate their use throughout the AI tool's lifecycle.

By the end of the process, all the recommendations were approved with less than 5% disagreement among all FUTURE-AI members. The FUTURE-AI consortium provided knowledge and expertise across a wide range of disciplines and stakeholders, resulting in consensus and wide support, both geographically and across domains. Hence, the FUTURE-AI guideline can benefit a wide range of stakeholders, as detailed in Table 5 in the Appendix.

FUTURE-AI is a risk-informed framework. It proposes to assess application-specific risks and challenges early in the process (e.g. risk of discrimination, lack of generalisability, data drifts, lack of acceptance by end-users, potential harm for patients, lack of transparency, data security vulnerabilities, ethical risks), then implement tailored measures to reduce these risks (e.g. collect data on individuals' attributes to assess and mitigate bias). This is also a risk-benefit balancing

exercise, as the specific measures to be implemented have benefits and potential weaknesses that the developers need to assess and balance. For example, collecting data on individuals' attributes may increase the risk of re-identification, but can enable to reduce the risk of bias and discrimination. Hence, in FUTURE-AI, risk management (as recommended in Traceability 1) must be a continuous and transparent process throughout the AI tool's lifecycle.

Furthermore, FUTURE-AI is an assumption-free, highly collaborative framework. It recommends to continuously engage with multi-disciplinary stakeholders to understand application-specific needs, risks and solutions (General 1). This is crucial to remove assumptions and investigate all possible risks and factors that may reduce trust in a given AI tool. For example, instead of making any assumption on possible sources of bias (e.g. sex or age), FUTURE-AI recommends that the developers engage with healthcare professionals, domain experts, representative citizens, and/or ethicists early in the process to investigate in depth the application-specific sources of bias, that may include factors well beyond standards attributes (e.g. breast density for AI applications in breast cancer).

For deployable AI tools, 24 recommendations out of 28 are rated as highly recommended (Table 2). For research and proof-of-concept AI tools, only 12 recommendations are rated as highly recommended, but we advise that researchers use as many elements as possible from the FUTURE-AI guideline to facilitate future transitions towards real-world practice.

The FUTURE-AI guideline was defined in a generic manner to ensure it can be applied across a variety of domains (e.g. radiology, genomics, mobile health, electronic health records). However, for many recommendations, their applicability varies across medical use cases. Hence, the first recommendation in each of the FUTURE-AI framework's principles is to identify the specificities to be addressed, such as the types of biases (Fairness 1), the clinical settings (Universality 1), or the need and approaches for explainable AI (Explainability 1).

The FUTURE-AI framework provides a set of general recommendations on how to enhance the trustworthiness of medical AI tools but does not impose any specific techniques for implementing each recommendation. While some examples of techniques are provided in the Appendix (Table 3), the final implementations should be defined by the developers, who should carefully select the most adequate methods given the application domain, clinical use case and data characteristics, as well as the advantages and limitations of each method. While we obtained a large consensus, some AI experts may disagree with some of the recommendations or may consider that some recommendations are either missing or not fully addressed. For example, while we propose mechanisms to enhance traceability and governance (e.g. AI logging), the issue of liability is yet to be addressed (e.g. who should be responsible for periodic auditing of the AI tools, who should be accountable when there is an error). Some of these key issues will require further investigations by multi-disciplinary researchers in the field of trustworthy AI, as well as by legal experts, regulators and authorities.

Aware of this limitation, we propose FUTURE-AI as a dynamic, living framework. Progressive development and adoption of medical AI tools will lead to new needs, challenges, and opportunities. To refine the FUTURE-AI guideline and learn from other voices, we set up a dedicated webpage (www.future-ai.eu) through which we invite the community to join the FUTURE-AI network and provide feedback based on their own experience and perspective. On the website, we include a FUTURE-AI self-assessment checklist, which comprises a set of questions and examples to facilitate and illustrate the use of the FUTURE-AI recommendations.

Additionally, we plan to organise regular outreach events such as webinars and workshops to exchange with medical AI researchers, manufacturers, evaluators, end-users and regulators.

REFERENCES

[1] [n. d.]. European Commission, Directorate-General for Communications Networks, Content and Technology, The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self assessment, Publications Office, 2020, https://data.europa.eu/doi/10.2759/002360.

[2] [n. d.]. ISO/IEC JTC 1/SC 42 - artificial intelligence. ISO. May 27, 2023. Accessed May 31, 2023. https://www.iso.org/committee/6794475.html.

[3] Pekka Ala-Pietilä, Yann Bonnet, Urs Bergmann, Maria Bielikova, Cecilia Bonefeld-Dahl, Wilhelm Bauer, Loubna Bouarfa, Raja Chatila, Mark Coeckelbergh, Virginia Dignum, et al. 2020. *The assessment list for trustworthy artificial intelligence (ALTAI).* European Commission.

[4] Julia Amann, Alessandro Blasimme, Effy Vayena, Dietmar Frey, and Vince I Madai. 2020. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC medical informatics and decision making* 20, 1 (2020), 1–9.

[5] Yves Saint James Aquino, Wendy A Rogers, Annette Braunack-Mayer, Helen Frazer, Khin Than Win, Nehmat Houssami, Christopher Degeling, Christopher Semsarian, and Stacy M Carter. 2023. Utopia versus dystopia: Professional perspectives on the impact of healthcare artificial intelligence on clinical roles and skills. *International Journal of Medical Informatics* 169 (2023), 104903.

[6] Matthew Arnold, Rachel KE Bellamy, Michael Hind, Stephanie Houde, Sameep Mehta, Aleksandra Mojsilović, Ravi Nair, K Natesan Ramamurthy, Alexandra Olteanu, David Piorkowski, et al. 2019. FactSheets: Increasing trust in AI services through supplier's declarations of conformity. *IBM Journal of Research and Development* 63, 4/5 (2019), 6–1.

[7] Leila Arras, Ahmed Osman, and Wojciech Samek. 2022. CLEVR-XAI: A benchmark dataset for the ground truth evaluation of neural network explanations. *Information Fusion* 81 (2022), 14–40.

[8] Solon Barocas, Moritz Hardt, and Arvind Narayanan. 2017. Fairness in machine learning. *Nips tutorial* 1 (2017), 2017.

[9] Rachel KE Bellamy, Kuntal Dey, Michael Hind, Samuel C Hoffman, Stephanie Houde, Kalapriya Kannan, Pranay Lohia, Jacquelyn Martino, Sameep Mehta, Aleksandra Mojsilovic, et al. 2018. AI Fairness 360: An extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias. *arXiv preprint arXiv:1810.01943* (2018).

[10] Oliver Bodenreider, Ronald Cornet, and Daniel J Vreeman. 2018. Recent developments in clinical terminologies—SNOMED CT, LOINC, and RxNorm. *Yearbook of medical informatics* 27, 01 (2018), 129–139.

[11] Tyler J Bradshaw, Ronald Boellaard, Joyita Dutta, Abhinav K Jha, Paul Jacobs, Quanzheng Li, Chi Liu, Arkadiusz Sitek, Babak Saboury, Peter JH Scott, et al. 2022. Nuclear medicine and artificial intelligence: best practices for algorithm development. *Journal of Nuclear Medicine* 63, 4 (2022), 500–510.

[12] Federico Cabitza, Andrea Campagner, Felipe Soares, Luis García de Guadiana-Romualdo, Feyissa Challa, Adela Sulejmani, Michela Seghezzi, and Anna Carobene. 2021. The importance of being external. methodological insights for the external validation of machine learning models in medicine. *Computer Methods and Programs in Biomedicine* 208 (2021), 106288.

[13] Leo Anthony Celi, Jacqueline Cellini, Marie-Laure Charpignon, Edward Christopher Dee, Franck Dernoncourt, Rene Eber, William Greig Mitchell, Lama Moukheiber, Julian Schirmer, Julia Situ, et al. 2022. Sources of bias in artificial intelligence that perpetuate healthcare disparities—A global review. *PLOS Digital Health* 1, 3 (2022), e0000022.

[14] Robert Challen, Joshua Denny, Martin Pitt, Luke Gompels, Tom Edwards, and Krasimira Tsaneva-Atanasova. 2019. Artificial intelligence, bias and clinical safety. *BMJ Quality & Safety* 28, 3 (2019), 231–237.

[15] Roomasa Channa, Risa Wolf, and Michael D Abramoff. 2021. Autonomous artificial intelligence in diabetic retinopathy: from algorithm to clinical application. *Journal of diabetes science and technology* 15, 3 (2021), 695–698.

[16] Gary S Collins, Paula Dhiman, Constanza L Andaur Navarro, Jie Ma, Lotty Hooft, Johannes B Reitsma, Patricia Logullo, Andrew L Beam, Lily Peng, Ben Van Calster, et al. 2021. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. *BMJ open* 11, 7 (2021), e048008.

[17] Gary S Collins, Paula Dhiman, Constanza L Andaur Navarro, Jie Ma, Lotty Hooft, Johannes B Reitsma, Patricia Logullo, Andrew L Beam, Lily Peng, Ben Van Calster, et al. 2021. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. *BMJ open* 11, 7 (2021), e048008.

[18] European Commission. [n. d.]. AI for Health Imaging | Programme | H2020 | CORDIS. https://cordis.europa.eu/programme/id/H2020_DT-TDS-05-2020

[19] Alex J DeGrave, Joseph D Janizek, and Su-In Lee. 2021. AI for radiographic COVID-19 detection selects shortcuts over signal. *Nature Machine Intelligence* 3, 7 (2021), 610–619.

[20] Carsten F Dormann. 2020. Calibration of probability predictions from machine-learning and statistical models. *Global ecology and biogeography* 29, 4 (2020), 760–765.

[21] Jean Feng, Rachael V Phillips, Ivana Malenica, Andrew Bishara, Alan E Hubbard, Leo A Celi, and Romain Pirracchio. 2022. Clinical artificial intelligence quality improvement: towards continual monitoring and updating of AI algorithms

in healthcare. *npj Digital Medicine* 5, 1 (2022), 66.
[22] Kadija Ferryman and Mikaela Pitcan. 2018. Fairness in precision medicine. (2018).
[23] Samuel G Finlayson, John D Bowers, Joichi Ito, Jonathan L Zittrain, Andrew L Beam, and Isaac S Kohane. 2019. Adversarial attacks on medical machine learning. *Science* 363, 6433 (2019), 1287–1289.
[24] Samuel G Finlayson, John D Bowers, Joichi Ito, Jonathan L Zittrain, Andrew L Beam, and Isaac S Kohane. 2019. Adversarial attacks on medical machine learning. *Science* 363, 6433 (2019), 1287–1289.
[25] Shaswath Ganapathi, Jo Palmer, Joseph E Alderman, Melanie Calvert, Cyrus Espinoza, Jacqui Gath, Marzyeh Ghassemi, Katherine Heller, Francis Mckay, Alan Karthikesalingam, et al. 2022. Tackling bias in AI health datasets through the STANDING Together initiative. *Nature Medicine* 28, 11 (2022), 2232–2233.
[26] Yuan Gao, Yuanyuan Wang, and Jinhua Yu. 2019. Optimized Resolution-Oriented Many-to-One Intensity Standardization Method for Magnetic Resonance Images. *Applied Sciences* 9, 24 (2019), 5531.
[27] Lidia Garrucho, Kaisar Kushibar, Socayna Jouide, Oliver Diaz, Laura Igual, and Karim Lekadir. 2022. Domain generalization in deep learning based mass detection in mammography: A large-scale multi-center study. *Artificial Intelligence in Medicine* 132 (2022), 102386.
[28] Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé Iii, and Kate Crawford. 2021. Datasheets for datasets. *Commun. ACM* 64, 12 (2021), 86–92.
[29] Marzyeh Ghassemi, Luke Oakden-Rayner, and Andrew L Beam. 2021. The false hope of current approaches to explainable artificial intelligence in health care. *The Lancet Digital Health* 3, 11 (2021), e745–e750.
[30] Marzyeh Ghassemi, Luke Oakden-Rayner, and Andrew L Beam. 2021. The false hope of current approaches to explainable artificial intelligence in health care. *The Lancet Digital Health* 3, 11 (2021), e745–e750.
[31] Amirata Ghorbani, Abubakar Abid, and James Zou. 2019. Interpretation of neural networks is fragile. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 33. 3681–3688.
[32] Benjamin Haibe-Kains, George Alexandru Adam, Ahmed Hosny, Farnoosh Khodakarami, Massive Analysis Quality Control (MAQC) Society Board of Directors Shraddha Thakkar 35 Kusko Rebecca 36 Sansone Susanna-Assunta 37 Tong Weida 35 Wolfinger Russ D. 38 Mason Christopher E. 39 Jones Wendell 40 Dopazo Joaquin 41 Furlanello Cesare 42, Levi Waldron, Bo Wang, Chris McIntosh, Anna Goldenberg, Anshul Kundaje, et al. 2020. Transparency and reproducibility in artificial intelligence. *Nature* 586, 7829 (2020), E14–E16.
[33] Jianxing He, Sally L Baxter, Jie Xu, Jiming Xu, Xingtao Zhou, and Kang Zhang. 2019. The practical implementation of artificial intelligence technologies in medicine. *Nature medicine* 25, 1 (2019), 30–36.
[34] Anna Hedström, Leander Weber, Daniel Krakowczyk, Dilyara Bareeva, Franz Motzkus, Wojciech Samek, Sebastian Lapuschkin, and Marina M-C Höhne. 2023. Quantus: An explainable ai toolkit for responsible evaluation of neural network explanations and beyond. *Journal of Machine Learning Research* 24, 34 (2023), 1–11.
[35] Anna Hedström, Leander Weber, Daniel Krakowczyk, Dilyara Bareeva, Franz Motzkus, Wojciech Samek, Sebastian Lapuschkin, and Marina M-C Höhne. 2023. Quantus: An explainable ai toolkit for responsible evaluation of neural network explanations and beyond. *Journal of Machine Learning Research* 24, 34 (2023), 1–11.
[36] IEEE. [n. d.]. IEEE SA - IEEE Recommended Practice for the quality management of datasets for Medical Artificial Intelligence. IEEE Standards Association. Accessed May 31, 2023. https://standards.ieee.org/ieee/2801/7459/.
[37] Sayash Kapoor and Arvind Narayanan. 2022. Leakage and the reproducibility crisis in ML-based science. *arXiv preprint arXiv:2207.07048* (2022).
[38] Sara Kaviani, Ki Jin Han, and Insoo Sohn. 2022. Adversarial attacks and defenses on AI in medical imaging informatics: A survey. *Expert Systems with Applications* 198 (2022), 116815.
[39] Burak Kocak, Bettina Baessler, Spyridon Bakas, Renato Cuocolo, Andrey Fedorov, Lena Maier-Hein, Nathaniel Mercaldo, Henning Müller, Fanny Orlhac, Daniel Pinto dos Santos, et al. 2023. CheckList for EvaluAtion of Radiomics research (CLEAR): a step-by-step reporting guideline for authors and reviewers endorsed by ESR and EuSoMII. *Insights into Imaging* 14, 1 (2023), 1–13.
[40] Benjamin Kompa, Jasper Snoek, and Andrew L Beam. 2021. Second opinion needed: communicating uncertainty in medical machine learning. *NPJ Digital Medicine* 4, 1 (2021), 4.
[41] Florian Königstorfer and Stefan Thalmann. 2022. AI Documentation: A path to accountability. *Journal of Responsible Technology* 11 (2022), 100043.
[42] David B Larson, Hugh Harvey, Daniel L Rubin, Neville Irani, R Tse Justin, and Curtis P Langlotz. 2021. Regulatory frameworks for development and evaluation of artificial intelligence–based diagnostic imaging algorithms: summary and recommendations. *Journal of the American College of Radiology* 18, 3 (2021), 413–424.
[43] David B Larson, Hugh Harvey, Daniel L Rubin, Neville Irani, R Tse Justin, and Curtis P Langlotz. 2021. Regulatory frameworks for development and evaluation of artificial intelligence–based diagnostic imaging algorithms: summary and recommendations. *Journal of the American College of Radiology* 18, 3 (2021), 413–424.
[44] Alexander Lavin, Ciarán M Gilligan-Lee, Alessya Visnjic, Siddha Ganju, Dava Newman, Sujoy Ganguly, Danny Lange, Atílím Güneş Baydin, Amit Sharma, Adam Gibson, et al. 2022. Technology readiness levels for machine learning

systems. *Nature Communications* 13, 1 (2022), 6039.
[45] Karim Lekadir, Gianluca Quaglio, A Tselioudis Garmendia, and Catherine Gallin. 2022. Artificial Intelligence in Healthcare-Applications, Risks, and Ethical and Societal Impacts. *European Parliament* (2022).
[46] Andreanne Lemay, Katharina Hoebel, Christopher P Bridge, Brian Befano, Silvia De Sanjosé, Didem Egemen, Ana Cecilia Rodriguez, Mark Schiffman, John Peter Campbell, and Jayashree Kalpathy-Cramer. 2022. Improving the repeatability of deep learning models with Monte Carlo dropout. *npj Digital Medicine* 5, 1 (2022), 174.
[47] Xiaoxiao Li, Ziteng Cui, Yifan Wu, Lin Gu, and Tatsuya Harada. 2021. Estimating and improving fairness with adversarial learning. *arXiv preprint arXiv:2103.04243* (2021).
[48] Daniel M Lima, Jose F Rodrigues-Jr, Agma JM Traina, Fabio A Pires, and Marco A Gutierrez. 2019. Transforming two decades of ePR data to OMOP CDM for clinical research. *Stud Health Technol Inform* 264, August (2019), 233–7.
[49] Xiaoxuan Liu, Samantha Cruz Rivera, David Moher, Melanie J Calvert, Alastair K Denniston, Hutan Ashrafian, Andrew L Beam, An-Wen Chan, Gary S Collins, Ara DarziJonathan J Deeks, et al. 2020. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *The Lancet Digital Health* 2, 10 (2020), e537–e548.
[50] Xiaoxuan Liu, Samantha Cruz Rivera, David Moher, Melanie J Calvert, Alastair K Denniston, Hutan Ashrafian, Andrew L Beam, An-Wen Chan, Gary S Collins, Ara DarziJonathan J Deeks, et al. 2020. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *The Lancet Digital Health* 2, 10 (2020), e537–e548.
[51] Lena Maier-Hein, Bjoern Menze, et al. 2022. Metrics reloaded: Pitfalls and recommendations for image analysis validation. *arXiv. org* 2206.01653 (2022).
[52] Lena Maier-Hein, Bjoern Menze, et al. 2022. Metrics reloaded: Pitfalls and recommendations for image analysis validation. *arXiv. org* 2206.01653 (2022).
[53] Stuart McLennan, Amelia Fiske, Leo Anthony Celi, Ruth Müller, Jan Harder, Konstantin Ritt, Sami Haddadin, and Alena Buyx. 2020. An embedded ethics approach for AI development. *Nature Machine Intelligence* 2, 9 (2020), 488–490.
[54] Stuart McLennan, Amelia Fiske, Daniel Tigard, Ruth Müller, Sami Haddadin, and Alena Buyx. 2022. Embedded ethics: a proposal for integrating ethics into the development of medical AI. *BMC Medical Ethics* 23, 1 (2022), 6.
[55] Agnieszka Mikołajczyk and Michał Grochowski. 2018. Data augmentation for improving deep learning in image classification problem. In *2018 international interdisciplinary PhD workshop (IIPhDW)*. IEEE, 117–122.
[56] Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. Model cards for model reporting. In *Proceedings of the conference on fairness, accountability, and transparency*. 220–229.
[57] Sina Mohseni, Niloofar Zarei, and Eric D Ragan. 2021. A multidisciplinary survey and framework for design and evaluation of explainable AI systems. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 11, 3-4 (2021), 1–45.
[58] John Mongan, Linda Moy, and Charles E Kahn Jr. 2020. Checklist for artificial intelligence in medical imaging (CLAIM): a guide for authors and reviewers. , e200029 pages.
[59] John Mongan, Linda Moy, and Charles E Kahn Jr. 2020. Checklist for artificial intelligence in medical imaging (CLAIM): a guide for authors and reviewers. , e200029 pages.
[60] Blake Murdoch. 2021. Privacy and artificial intelligence: challenges for protecting health information in a new era. *BMC Medical Ethics* 22, 1 (2021), 1–5.
[61] Kee Yuan Ngiam and Wei Khor. 2019. Big data and machine learning algorithms for health-care delivery. *The Lancet Oncology* 20, 5 (2019), e262–e273.
[62] Luis Oala, Andrew G Murchison, Pradeep Balachandran, Shruti Choudhary, Jana Fehr, Alixandro Werneck Leite, Peter G Goldschmidt, Christian Johner, Elora DM Schörverth, Rose Nakasi, et al. 2021. Machine learning for health: algorithm auditing & quality control. *Journal of medical systems* 45 (2021), 1–8.
[63] Luis Oala, Andrew G Murchison, Pradeep Balachandran, Shruti Choudhary, Jana Fehr, Alixandro Werneck Leite, Peter G Goldschmidt, Christian Johner, Elora DM Schörverth, Rose Nakasi, et al. 2021. Machine learning for health: algorithm auditing & quality control. *Journal of medical systems* 45 (2021), 1–8.
[64] Seong Ho Park and Kyunghwa Han. 2018. Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction. *Radiology* 286, 3 (2018), 800–809.
[65] Geoff Pleiss, Manish Raghavan, Felix Wu, Jon Kleinberg, and Kilian Q Weinberger. 2017. On fairness and calibration. *Advances in neural information processing systems* 30 (2017).
[66] Janet Rafner, Dominik Dellermann, Arthur Hjorth, Dora Veraszto, Constance Kampf, Wendy MacKay, and Jacob Sherson. 2022. Deskilling, upskilling, and reskilling: a case for hybrid intelligence. *Morals & Machines* 1, 2 (2022), 24–39.
[67] Sandeep Reddy, Wendy Rogers, Ville-Petteri Makinen, Enrico Coiera, Pieta Brown, Markus Wenzel, Eva Weicken, Saba Ansari, Piyush Mathur, Aaron Casey, et al. 2021. Evaluation framework to guide implementation of AI systems into healthcare settings. *BMJ health & care informatics* 28, 1 (2021).

[68] Pouria Rouzrokh, Bardia Khosravi, Shahriar Faghani, Mana Moassefi, Diana V Vera Garcia, Yashbir Singh, Kuan Zhang, Gian Marco Conte, and Bradley J Erickson. 2022. Mitigating bias in radiology machine learning: 1. Data handling. *Radiology: Artificial Intelligence* 4, 5 (2022), e210290.

[69] Berkman Sahiner, Weijie Chen, Ravi K Samala, and Nicholas Petrick. 2023. Data drift in medical machine learning: implications and potential remedies. *The British Journal of Radiology* (2023), 20220878.

[70] Raghavendra Selvan, Nikhil Bhagwat, Lasse F Wolff Anthony, Benjamin Kanding, and Erik B Dam. 2022. Carbon footprint of selecting and training deep learning models for medical image analysis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 506–516.

[71] Matthew Sperrin, Richard D Riley, Gary S Collins, and Glen P Martin. 2022. Targeted validation: validating clinical prediction models in their intended population and setting. *Diagnostic and Prognostic Research* 6, 1 (2022), 24.

[72] Mark Sujan, Dominic Furniss, Kath Grundy, Howard Grundy, David Nelson, Matthew Elliott, Sean White, Ibrahim Habli, and Nick Reynolds. 2019. Human factors challenges for the safe use of artificial intelligence in patient care. *BMJ health & care informatics* 26, 1 (2019).

[73] Mark Sujan, Dominic Furniss, Kath Grundy, Howard Grundy, David Nelson, Matthew Elliott, Sean White, Ibrahim Habli, and Nick Reynolds. 2019. Human factors challenges for the safe use of artificial intelligence in patient care. *BMJ health & care informatics* 26, 1 (2019).

[74] Yingjie Tian and Yuqi Zhang. 2022. A comprehensive survey on regularization strategies in machine learning. *Information Fusion* 80 (2022), 146–166.

[75] Erico Tjoa and Cuntai Guan. 2020. A survey on explainable artificial intelligence (xai): Toward medical xai. *IEEE transactions on neural networks and learning systems* 32, 11 (2020), 4793–4813.

[76] Gaël Varoquaux and Veronika Cheplygina. 2022. Machine learning for medical imaging: methodological failures and recommendations for the future. *NPJ digital medicine* 5, 1 (2022), 48.

[77] Baptiste Vasey, Myura Nagendran, Bruce Campbell, David A Clifton, Gary S Collins, Spiros Denaxas, Alastair K Denniston, Livia Faes, Bart Geerts, Mudathir Ibrahim, et al. 2022. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *Nature medicine* 28, 5 (2022), 924–933.

[78] Kerstin N Vokinger, Stefan Feuerriegel, and Aaron S Kesselheim. 2021. Mitigating bias in machine learning for medicine. *Communications medicine* 1, 1 (2021), 25.

[79] Sebastian Vollmer, Bilal A Mateen, Gergo Bohner, Franz J Király, Rayid Ghani, Pall Jonsson, Sarah Cumbers, Adrian Jonas, Katherine SL McAllister, Puja Myles, et al. 2020. Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. *bmj* 368 (2020).

[80] Ian Walsh, Dmytro Fishman, Dario Garcia-Gasulla, Tiina Titma, Gianluca Pollastri, Jennifer Harrow, Fotis E Psomopoulos, and Silvio CE Tosatto. 2021. DOME: recommendations for supervised machine learning validation in biology. *Nature methods* 18, 10 (2021), 1122–1127.

[81] Mark D Wilkinson, Michel Dumontier, Ij J Aalbersberg, G Appleton, M Axton, A Baak, N Blomberg, JW Boiten, LB da Silva Santos, PE Bourne, et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3: 160018.

[82] Kuan Zhang, Bardia Khosravi, Sanaz Vahdati, Shahriar Faghani, Fred Nugen, Seyed Moein Rassoulinejad-Mousavi, Mana Moassefi, Jaidip Manikrao M Jagtap, Yashbir Singh, Pouria Rouzrokh, et al. 2022. Mitigating bias in radiology machine learning: 2. Model development. *Radiology: Artificial Intelligence* 4, 5 (2022), e220010.

[83] Qian Zhou, Zhi-hang Chen, Yi-heng Cao, and Sui Peng. 2021. Clinical impact and quality of randomized controlled trials involving interventions evaluating artificial intelligence prediction tools: a systematic review. *NPJ digital medicine* 4, 1 (2021), 154.

APPENDIX

A TABLES

Table 3. Detailed descriptions of the FUTURE-AI recommendations.

Recommendation	Description
Fairness 1. Define sources of bias	Bias in medical AI is application-specific [22]. At the design phase, the development team should identify possible types and sources of bias for their AI tool [25]. These may include group attributes (e.g. sex, gender, age, ethnicity, socioeconomics, geography), the medical profiles of the individuals (e.g. with comorbidities or disability), as well as human biases during data labeling, data curation, or the selection of the input features.

Continued on next page

Table 3 – *Continued from previous page*

Recommendation	Description
Fairness 2. Collect data on attributes	To identify biases and apply measures for increased fairness, relevant attributes of the individuals, such as sex, gender, age, ethnicity, risk factors, comorbidities or disabilities, should be collected. This should be subject to informed consent and approval by ethics committees to ensure an appropriate balance between the benefits for non-discrimination and risks for re-identification.
Fairness 3. Evaluate & correct biases	When possible, i.e. the individuals' attributes are included in the data, bias detection methods should be applied by using fairness metrics [8, 9]. To correct for any identified biases, mitigation measures should be applied (e.g. data re-sampling, bias-free representations, equalised odds post-processing) [47, 65, 68, 78, 82] and tested to verify their impact on both the tool's fairness and the model's accuracy. Importantly, any potential bias should be documented and reported to inform the end-users and citizens (see Traceability 2).
Universality 1. Define clinical settings	At the design phase, the development team should specify the clinical settings in which the AI tool will be applied (e.g. primary healthcare centres, hospitals, home care, low vs. high-resource settings, one or multiple countries), and anticipate potential obstacles to universality (e.g. differences in clinical definitions, medical equipment or IT infrastructures across settings).
Universality 2. Use existing standards	To ensure the quality and interoperability of the AI tool, it should be developed based on existing community-defined standards. These may include clinical definitions, medical ontologies (e.g. SNOMED CT [10], OMOP [48]), interface standards (e.g. DICOM, FHIR HL7), data annotations, evaluation criteria [52], and technical standards (e.g. IEEE [36] or ISO [2]).
Universality 3. Evaluate using external data	To assess generalisability, technical validation of the AI tools should be performed with external datasets that are distinct from those used for training [12]. These may include reference or benchmarking datasets which are representative for the task in question (i.e. approximating the expected real-world variations). Except for AI tools intended for single centres, the clinical evaluation studies should be performed at multiple sites [71] to assess performance and interoperability across clinical workflows. If the tool's generalisability is limited, mitigation measures (e.g. transfer learning or domain adaptation) should be considered, applied and tested.
Universality 4. Evaluate local clinical validity	Clinical settings vary in many aspects, such as populations, equipment, clinical workflows, and end-users. Hence to ensure trust at each site, the AI tools should be evaluated for their local clinical validity [43]. In particular, the AI tool should fit the local clinical workflows and perform well on the local populations. If the performance is decreased when evaluated locally, re-calibration of the AI model should be performed (e.g., through model fine-tunning or retraining).
Traceability 1. Implement risk management	Throughout the AI tool's lifecycle, the development team should analyse potential risks, assess each risk's likelihood, effects and risk-benefit balance, define risk mitigation measures, monitor the risks and mitigations continuously, and maintain a risk management file. The risks may include those explicitly covered by the FUTURE-AI guiding principles (e.g. bias, harm), but also application-specific risks. Other risks to consider include human factors that may lead to misuse of the AI tool (e.g. not following the instructions, receiving insufficient training), application of the AI tool to individuals who are not within the target population, use of the tool by others than the target end-users (e.g. technician instead of physician), hardware failure, incorrect data annotations or input values, and adversarial attacks. Mitigation measures may include warnings to the users, system shutdown, re-processing of the input data, the acquisition of new input data, or the use of an alternative procedure or human judgment only.
Traceability 2. Provide documentation	To increase transparency, traceability, and accountability, adequate documentation should be created and maintained for the AI tool [41], which may include (i) an AI information leaflet to inform citizens and healthcare professionals about the tool's intended use, risks (e.g. biases) and instructions for use; (ii) a technical document [6, 28, 56] to inform AI developers, health organisations and regulators about the AI model's properties (e.g. hyperparameters), training and testing data, evaluation criteria and results, biases and other limitations, and periodic audits and updates; (iii) a publication based on existing AI reporting standards [17, 50, 59], and (iv) a risk management file (see Traceability 1).

Continued on next page

Table 3 – *Continued from previous page*

Recommendation	Description
Traceability 3. Implement continuous quality control	The AI tool should be developed and deployed with mechanisms for continuous monitoring and quality control of the AI inputs and outputs [63], such as to identify missing or out-of-range input variables, inconsistent data formats or units, incorrect annotations or data pre-processing, and erroneous or implausible AI outputs. For quality control of the AI decisions, uncertainty estimates should be provided (and calibrated [20]) to inform the end-users on the degree of confidence in the results [40]. Finally, when necessary, model updates should be applied to address any identified limitations and enhance the AI models over time [21].
Traceability 4. Implement periodic auditing	The AI tool should be developed and deployed with a configurable system for periodic auditing [63], which should define site-specific datasets and timelines for periodic evaluations (e.g. every year). The periodic auditing should enable the identification of data or concept drifts, newly occurring biases, performance degradation [69] or changes in the decision making of the end-users. Accordingly, necessary updates to the AI models or AI tools should be applied [21].
Traceability 5. Implement AI logging	To increase traceability and accountability, an AI logging system should be implemented to trace the user's main actions in a privacy-preserving manner, specify the data that is accessed and used, record the AI predictions and clinical decisions, and log any encountered issues. Time-series statistics and visualisations should be used to inspect the usage of the AI tool over time.
Traceability 6. Implement human oversight	Given the high-stake nature of medical AI, human oversight is essential and increasingly required by policy makers and regulators [1, 43]. Human-AI interfaces and human-in-the-loop mechanisms should be designed and implemented to perform specific quality checks (e.g. to flag biases, errors or implausible explanations), and to overrule the AI decisions when necessary. Furthermore, governance of the AI tool in the health organisation should be specified, including roles and responsibilities for performing risk management, periodic auditing, human oversight, and AI tool maintenance.
Usability 1. Define user requirements	The AI developers should engage clinical experts, end-users (e.g. patients, physicians) and other relevant stakeholders (e.g. data managers, administrators) from an early stage, to compile information on the AI tool's intended use and end-user requirements (e.g. human-AI interfaces), as well as on human factors that may impact the usage of the AI tool [73] (e.g. ergonomics, intuitiveness, experience, learnability).
Usability 2. Provide training	To facilitate best usage of the AI tool, minimise errors and harm, and increase AI literacy, the developers should provide training materials (e.g. tutorials, manuals, examples) in accessible language and/or training activities (e.g. hands-on sessions), taking into account the diversity of end-users (e.g. clinical specialists, nurses, technicians, citizens or administrators).
Usability 3. Evaluate clinical usability	To facilitate adoption, the usability of the AI tool should be evaluated in the real world with representative and diverse end-users (e.g. with respect to sex, gender, age, clinical role, digital proficiency, (dis)ability). The usability tests should gather evidence on the user's satisfaction, performance and productivity. These tests should also verify whether the AI tool impacts the behaviour and decision making of the end-users.
Usability 4. Evaluate clinical utility	The AI tool should be evaluated for its clinical utility and safety. The clinical evaluations of the AI tool should show benefits for the clinician (e.g. increased productivity, improved care), for the patient (e.g. earlier diagnosis, better outcomes), and/or for the healthcare organisation (e.g. reduced costs, optimised workflows), when compared to the current standard of care. Additionally, it is important to show that the AI tool is safe and does not cause harm to individuals (or specific groups), such as through a randomised clinical trial [83].
Robustness 1. Define sources of data variation	At the design phase, an inventory should be made of the application-specific sources of variation that may impact the AI tool's robustness in the real world. These may include differences in equipment, technical fault of a machine, data heterogeneities during data acquisition or annotation, and/or adversarial attacks [24].
Robustness 2. Train with representative data	Clinicians, citizens and other stakeholders are more likely to trust the AI tool if it is trained on data that adequately represents the variations encountered in real-world clinical practice [61]. Hence, the training datasets should be carefully selected, analysed and enriched according to the sources of variation identified at the design phase (see Robustness 1).

Continued on next page

Table 3 – *Continued from previous page*

Recommendation	Description
Robustness 3. Evaluate & optimise robustness	Evaluation studies should be implemented to evaluate the AI tool's robustness (including stress tests and repeatability tests [46]), by considering all potential sources of variation (see Robustness 1), such as data-, equipment-, clinician-, patient- and centre-related variations. Depending on the results, mitigation measures should be implemented to optimise the robustness of the AI model, such as regularisation [74], data augmentation [55], data harmonisation [26], or domain adaptation [27].
Explainability 1. Define explainability needs	At the design phase, it should be established if explainability is required for the AI tool. In this case, the specific requirements for explainability should be defined with representative experts and end-users, including (i) the goal of the explanations (e.g. global description of the model's behaviour vs. local explanation of each AI decision), (ii) the most suitable approach for AI explainability [75], and (iii) the potential limitations to anticipate and monitor (e.g. over-reliance of the end-users on the AI decision [30]).
Explainability 2. Evaluate explainability	The explainable AI methods should be evaluated, first quantitatively by using in silico methods to assess the correctness of the explanations [7, 35], then qualitatively with end-users to assess their impact on user satisfaction, confidence and clinical performance [57]. The evaluations should also identify any limitations of the AI explanations (e.g. they are clinically incoherent [19] or sensitive to noise or adversarial attacks [31], they unreasonably increase the confidence in the AI-generated results [15]).
General 1. Engage stakeholders continuously	Throughout the AI tool's lifecycle, the AI developers should continuously engage with interdisciplinary stakeholders, such as healthcare professionals, citizens, patient representatives, expert ethicists, data managers and legal experts. This interaction will facilitate the understanding and anticipation of the needs, obstacles and pathways towards acceptance and adoption.
General 2. Ensure data protection	Adequate measures to ensure data privacy and security should be put in place throughout the AI lifecycle. These may include privacy-enhancing techniques (e.g. differential privacy, encryption), data protection impact assessment and appropriate data governance after deployment (e.g. logging system for data access, see Traceability 5). If de-identification is implemented (e.g. pseudonymisation, k-anonymity), the balance between the health benefits for citizens and the risks for re-identification should be carefully assessed and considered. Furthermore, the manufacturers and deployers should implement and regularly evaluate measures for protecting the AI tool against malicious attacks, such as by using system-level cybersecurity solutions or application-specific defense mechanisms [38] (e.g. attack detection or mitigation).
General 3. Define adequate evaluation plan	To increase trust and adoption, an appropriate evaluation plan should be defined (including test data, metrics and reference methods). First, adequate test data should be selected for assessing each dimension of trustworthy AI. In particular, the test data should be well separated from the training to prevent data leakage [37]. Furthermore, adequate evaluation metrics should be carefully selected, taking into account their benefits and potential flaws [76]. Finally, benchmarking with respect to reference AI tools or standard practice should be performed to enable a comparative assessment of model performance.
General 4. Comply with AI regulations	The development team should identify the applicable AI regulations depending on the relevant jurisdictions. This should be done at an early stage to anticipate regulatory obligations based on the medical AI tool's intended classification and risks.
General 5. Investigate ethical issues	In addition to the well-known ethical issues that arise in medical AI (e.g. privacy, transparency, equity, autonomy), AI developers, domain specialists and professional ethicists should identify, discuss and address all application-specific ethical, social and societal issues as an integral part of the development and deployment of the AI too [54].
General 6. Investigate social issues	Social and societal implications should be considered and addressed when developing the AI tool, to ensure a positive impact on citizens and society. Relevant issues include the impact of the AI tool on the working conditions and power relations, on the new skills (or deskilling) of the healthcare professionals and citizens [66], and on future interactions between citizens, health professionals and social carers. Furthermore, for environmental sustainability, AI developers should consider strategies to reduce the carbon footprint of the AI tool [70].

Table 4. A glossary of main terms used in the FUTURE-AI guideline (ranked alphabetically).

Term	Definition
AI auditing	A periodic evaluation of an AI tool to assess its performance and working conditions over time, and to identify potential problems.
AI deployment	The process of placing a completed AI tool into a live clinical environment where it can be used for its intended purpose.
AI design	Early stage of an AI's production lifetime, during which specifications and plans are defined for the subsequent development of the AI tool
AI development	The process of training AI models and building AI-human interfaces, based on the specifications and plans from the AI design phase.
AI evaluation	The assessment of an AI tool's added value in its intended clinical setting.
AI model	A program trained using a machine learning algorithm to perform a given task based on specific input data.
AI monitoring	The process of tracking the behavior of a deployed AI tool over time, to identify potential degradation in performance and implement mitigation measures such as model updating.
AI regulation	A set of requirements and obligations defined by public authorities, that AI developers, deployers and users must adhere to.
AI risk	Any negative effect that may occur when using an AI tool.
AI tool	A software that comprises the AI model plus a user interface that can be used by the end-users to perform a given AI-powered clinical task.
AI training	The process of using machine learning algorithms to build AI models that learn to perform specific tasks based on existing data samples.
AI updating	The process of re-training or fine-tuning the AI model after some time to improve its performance and correct identified issues.
AI validation	The assessment of an AI model's performance.
Attribute	Personal quality, trait or characteristic of an individual or group of individuals, such as sex, gender, age, ethnicity, socioeconomic status or disability. Protected attributes refer to those attributes that, by law, cannot be discriminated against (i.e. attributes that are protected by law).
Benchmarking	The practice of comparing the performance of multiple AI tools (or an AI tool against the standard practice) based on a common reference dataset and a set of predefined performance criteria and metrics.
Bias	Systematic, prejudiced errors by an AI tool against certain individuals or subgroups due to inadequate data or assumptions used during the training of the machine learning model.
Clinical safety	The capability of an AI tool to keep individuals and patients safe and not to cause them any harm.
Clinical setting	The environment or location where the AI tool will be used, such as a hospital, a radiology department, a primary care centre, or for home-based care.
Clinical utility	The capability of an AI tool to be useful in its intended clinical settings, such as to improve clinical outcomes, to increase the clinicians' productivity, or to reduce healthcare costs.
Concept drift	Changes in relationship between AI model inputs and outputs.
Data drift	Changes in the distribution of the AI model's input data over time.
Data quality control	The process of assessing the quality of the input data, to identify potential defects that may affect the correct functioning of the AI tool.
Deployable AI	AI developed with a high technology readiness level (TRL) (5-9) intended for deployment in clinical practice.
Ethical AI	AI that adheres to key ethical values and human rights, such as the rights to privacy, equity and autonomy.
Explainability	The ability of an AI tool to provide clinically meaningful information about the logic behind the AI decisions.
Fairness	The ability of an AI tool to treat equally individuals with similar characteristics or subgroups of individuals including under-represented groups.

Continued on next page

Table 4 – *Continued from previous page*

Term	Definition
Human oversight	A procedure or set of procedures put in place to ensure an AI tool is used under the supervision of a human (e.g. a clinician), who is able to overrule the AI decisions and take the final clinical decision.
Intended use	Clinical purpose or clinical task that the AI tool aims to realise in its intended clinical setting.
Logging	The process of keeping a log of events that occur while using an AI tool, such as user actions, accessed and used datasets, clinical decisions, and identified issues.
Proof-of-concept AI	AI developed with a low machine learning technology readiness level (ML-TRL) (1-4) to demonstrate the feasibility of a new AI method or new AI concept.
Real world	The clinical environment in which AI tools will be applied in practice, outside the controlled environment of research labs.
Responsible AI	AI that is designed, developed, evaluated, and monitored by employing an appropriate code of conduct and appropriate methods to achieve technical, clinical, ethical, and legal requirements (e.g. efficacy, safety, fairness, robustness, transparency).
Robustness	The ability of an AI tool to overcome expected or unexpected variations, such as due to noise or artefacts in the data.
Third-party evaluator	An independent evaluator who did not participate in any way in the design or development of the AI tool to be evaluated.
Traceability	The ability of an AI tool to be monitored over its complete lifecycle.
Trustworthy AI	AI with proven characteristics such as efficacy, safety, fairness, robustness, transparency, which enable relevant stakeholders such as citizens, clinicians, health organisations and authorities to rely on it and adopt it in real-world practice.
Trustworthy AI vs. Responsible AI	For trustworthy AI, the emphasis is on the characteristics of the AI tool and how they are perceived by the stakeholders of interest (e.g. patients, clinicians), while for responsible AI, the emphasis is on the developers, evaluators and managers of the AI tool, and the code of conduct and methods they employ to obtain trustworthy AI tools.
Universality	The ability of an AI tool to generalise across clinical settings.
Usability	The degree to which an AI tool is fit to be used by end-users in the intended clinical setting.

Table 5. List of stakeholder groups (ranked alphabetically) that can benefit from the FUTURE-AI guideline.

Stakeholders	FUTURE-AI usage
AI ethicists	• To embed ethics into the development of medical AI tools.
AI evaluators/clinical trialists	• To perform more comprehensive, multi-faceted evaluations of medical AI tools based on the principles of trustworthy AI. • To assess the trustworthiness of AI tools.
Citizens and patients	• To increase literacy about medical AI and trustworthy AI. • To increase engagement in the production and evaluation of medical AI tools.
Conferences/journals	• To promote best practices and new methods for trustworthy AI among researchers reading or publishing scientific papers.

Continued on next page

Table 5 – *Continued from previous page*

Stakeholders	FUTURE-AI usage
Data managers	• To support the development and deployment of medical AI tools that are compliant with data protection/governance principles.
Educational institutions	• To educate students from all disciplines (machine learning, computer science, medicine, ethics, social sciences) on the principles and approaches for trustworthy AI.
Funding agencies	• To promote new research projects that integrate best practices and new approaches for responsible AI.
Health organisations	• To guide healthcare organisations in the evaluation, deployment and monitoring of medical AI tools. • To verify the trustworthiness of AI tools.
Healthcare professionals	• To adopt the principles of trustworthy AI and best practices among the healthcare professions. • To engage clinicians in the design, development, evaluation and monitoring of medical AI tools.
IT managers	• To promote IT solutions for the deployment and monitoring of trustworthy and secure AI tools in clinical practice.
Legal experts	• To ensure compliance with applicable laws and regulations related to medical AI and data protection.
Manufacturers of medical AI devices	• To adopt best practices for responsible AI within companies. • To develop and/or commercialise new AI tools that will be accepted, certified and deployed for clinical use.
Public authorities	• To adapt existing regulations and policies on medical AI.
Regulatory bodies	• To enhance the procedures for the evaluation, certification and monitoring of AI tools as medical devices.
Researchers and developers in medical AI.	• To investigate new methods according to the recommendations for trustworthy AI. • To develop proof-of-concepts that can more easily transition into deployable AI tools for clinical practice.

Continued on next page

Table 5 – *Continued from previous page*

Stakeholders	FUTURE-AI usage
Scientific/medical societies	• To promote the principles of trustworthy AI and best practices among scientific and medical communities.
Social scientists	• To ensure social and societal dimensions of medical AI are considered.
Standardisation bodies	• To develop new standards that facilitate the implementation, evaluation and adoption of trustworthy AI tools in healthcare.

B FULL AUTHOR AFFILIATIONS

Karim Lekadir, Artificial Intelligence in Medicine Lab (BCN-AIM), Department of Mathematics and Computer Science, Universitat de Barcelona, Barcelona, Spain
Aasa Feragen, DTU Compute, Technical University of Denmark, Kgs Lyngby, Denmark
Abdul Joseph Fofanah, Department of Mathematics and Computer Science, Faculty of Science and Technology, Milton Margai Technical University, Freetown, Sierra Leone
Alejandro F Frangi, Centre for Computational Imaging & Simulation Technologies in Biomedicine, Schools of Computing and Medicine, University of Leeds, Leeds, United Kingdom and Medical Imaging Research Center (MIRC), Cardiovascular Science and Electronic Engineering Departments, KU Leuven, Leuven, Belgium
Alena Buyx Institute of History and Ethics in Medicine, Technical University of Munich, Munich, Germany
Anais Emelie Artificial Intelligence in Medicine Lab (BCN-AIM), Department of Mathematics and Computer Science, Universitat de Barcelona, Barcelona, Spain
Andrea Lara Faculty of Engineering of Systems, Informatics and Sciences of Computing, Galileo University, Guatemala City, Guatemala
Antonio R Porras Department of Biostatistics and Informatics, Colorado School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, CO, United States
An-Wen Chan Department of Medicine, Women's College Research Institute, University of Toronto, Toronto, Canada
Arcadi Navarro Institució Catalana de Recerca i Estudis Avançats (ICREA) and Universitat Pompeu Fabra, Barcelona, Spain and Barcelona Beta Brain Research Center, Pasqual Maragall Foundation, Barcelona, Spain
Ben Glocker Department of Computing, Imperial College London, London, United Kingdom
Benard O Botwe School of Biomedical & Allied Health Sciences, University of Ghana, Accra, Ghana and Department of Midwifery & Radiography, School of Health & Psychological Sciences, City University of London, UK
Bishesh Khanal NepAl Applied Mathematics and Informatics Institute for research (NAAMII), Kathmandu, Nepal
Brigit Beger European Heart Network, Brussels, Belgium
Carol C Wu Department of Thoracic Imaging, University of Texas MD Anderson Cancer Center, Houston, United States
Celia Cintas IBM Research Africa, Nairobi, Kenya
Curtis P Langlotz Center for Artificial Intelligence in Medicine & Imaging, Stanford University

School of Medicine, Stanford, United States
Daniel Rueckert Institute for AI and Informatics in Medicine, Klinikum rechts der Isar, Technical University Munich, Munich, Germany and Department of Computing, Imperial College London, London, UK
Deogratias Mzurikwao Muhimbili University of Health and Allied Sciences, Dar es Salaam, Tanzania
Dimitrios I Fotiadis Unit of Medical Technology and Intelligent Information Systems, Foundation for Research and Technology - Hellas (FORTH), Ioannina, Greece
Doszhan Zhussupov Almaty AI Lab, Almaty, Kazakhstan
Enzo Ferrante CONICET, Universidad Nacional del Litoral, Santa Fe, Argentina
Erik Meijering School of Computer Science and Engineering, University of New South Wales, Sydney, Australia
Eva Weicken Fraunhofer Heinrich Hertz Institute, Berlin, Germany
Fabio A González Computing Systems and Industrial Engineering Dept., Universidad Nacional de Colombia, Bogotá, Colombia
Folkert W Asselbergs Amsterdam University Medical Centers, Department of Cardiology, University of Amsterdam, Amsterdam, The Netherlands and Health Data Research UK and Institute of Health Informatics, University College London, London, United Kingdom
Fred Prior Department of Biomedical Informatics, University of Arkansas for Medical Sciences, Little Rock, United States
Gabriel P Krestin Department of Radiology & Nuclear Medicine, Erasmus MC University Medical Center, Rotterdam, the Netherlands
Gary Collins Centre for Statistics in Medicine, University of Oxford, Oxford, United Kingdom
Geletaw S Tegenaw Faculty of Computing and Informatics, JiT, Jimma University, Jimma, Ethiopia
Georgios Kaissis Institute for AI and Informatics in Medicine, Klinikum rechts der Isar, Technical University Munich, Munich, Germany
Gianluca Misuraca Department of Artificial Intelligence, Universidad Politécnica de Madrid, Madrid, Spain
Gianna Tsakou Gruppo Maggioli, Research and Development Lab, Athens, Greece
Girish Dwivedi Department of Advanced Clinical and Translational Cardiovascular Imaging, The University of Western Australia, Perth, Australia
Haridimos Kondylakis Department of Electrical and Computer Engineering, Hellenic Mediterranean University and Foundation for Research and Technology - Hellas (FORTH), Crete, Greece
Harsha Jayakody Postgraduate Institute of Medicine, University of Colombo, Colombo, Sri Lanka
Henry C Woodruf The D-lab, Department of Precision Medicine, GROW - School for Oncology and Reproduction, Maastricht University, Maastricht, the Netherlands
Hugo JWL Aerts Artificial Intelligence in Medicine Program, Mass General Brigham, Harvard Medical School, Boston, United States
Ian Walsh Bioprocessing Technology Institute, Agency for Science, Technology and Research (A*STAR), Singapore, Singapore
Ioanna Chouvarda School of Medicine, Aristotle University of Thessaloniki, Thessaloniki, Greece
Irène Buvat Institut Curie, Inserm, Orsay, France
Islem Rekik BASIRA Lab, Imperial-X and Computing Department, Imperial College London, UK and]Faculty of Computer and Informatics Engineering, Istanbul Technical University, Turkey
James Duncan Departments of Biomedical Engineering and Radiology & Biomedical Imaging,

Schools of Engineering & Applied Science and Medicine, Yale University, New Haven, United States
Jayashree Kalpathy-Cramer Massachusetts General Hospital, Harvard Medical School, Massachusetts, United States
Jihad Zahir LISI Laboratory, Computer Science Department, Cadi Ayyad University, Marrakech, Morocco
Jinah Park School of Computing, Korea Advanced Institute of Science and Technology, Daejeon, South Korea
John Mongan Department of Radiology and Biomedical Imaging, University of California San Francisco, San Francisco, United States
Judy W Gichoya Department of Radiology & Imaging Sciences, Emory University, Atlanta, United States
Julia A Schnabel Institute of Machine Learning in Biomedical Imaging, Helmholtz Center Munich, Munich, Germany
Kaisar Kushibar Artificial Intelligence in Medicine Lab (BCN-AIM), Department of Mathematics and Computer Science, Universitat de Barcelona, Barcelona, Spain
Katrine Riklund Department of Radiation Sciences, Diagnostic Radiology, Umeå University, Umeå, Sweden
Kensaku Mori Graduate School of Informatics, Nagoya University, Nagoya, Japan
Kostas Marias Department of Electrical and Computer Engineering, Hellenic Mediterranean University and Foundation for Research and Technology - Hellas (FORTH), Crete, Greece
Lameck M Amugongo Department of Software Engineering, Namibia University of Science & Technology, Windhoek, Namibia
Lauren A Fromont Center for Genomic Regulation, The Barcelona Institute of Science and Technology, Barcelona, Spain
Lena Maier-Hein Div. Intelligent Medical Systems (IMSY), German Cancer Research Center (DKFZ), Heidelberg, Germany
Leonor Cerdá Alberich Biomedical Imaging Research Group, La Fe Health Research Institute, Valencia, Spain
Leticia Rittner School of Electrical and Computer Engineering, University of Campinas, Campinas, Brazil
Lighton Phiri Department of Library Information Science, University of Zambia, Lusaka, Zambia
Linda Marrakchi-Kacem National Engineering School of Tunis, University of Tunis El Manar, Tunis, Tunisia
Lluís Donoso-Bach Clinical Advanced Technologies institute (CATI), Hospital Clínic of Barcelona, Barcelona, Spain
Luis Martí-Bonmatí Medical Imaging Department, Hospital Universitario y Politécnico La Fe, Valencia, Spain
M Jorge Cardoso School of Biomedical Engineering & Imaging Sciences, King's College London, London, United Kingdom
Maciej Bobowicz 2nd Division of Radiology, Medical University of Gdansk, Gdansk, Poland
Mahsa Shabani Faculty of Law and Criminology, Ghent University, Ghent, Belgium
Manolis Tsiknakis Department of Electrical and Computer Engineering, Hellenic Mediterranean University and Foundation for Research and Technology - Hellas (FORTH), Crete, Greece
Maria A Zuluaga Data Science Department, EURECOM, Sophia Antipolis, France
Maria Bielikova Kempelen Institute of Intelligent Technologies, Bratislava, Slovakia
Marie-Christine Fritzsche Institute of History and Ethics in Medicine, Technical University of Munich, Munich, Germany
Marius George Linguraru Sheikh Zayed Institute for Pediatric Surgical Innovation, Children's

National Hospital Washington DC, United States
Markus Wenzel Fraunhofer Heinrich Hertz Institute, Berlin, Germany
Marleen De Bruijne Department of Radiology & Nuclear Medicine, Erasmus MC University Medical Center, Rotterdam, the Netherlands
Martin G Tolsgaard Copenhagen Academy for Medical Education and Simulation Rigshospitalet, University of, Copenhagen, Copenhagen, Denmark
Marzyeh Ghassemi Electrical Engineering and Computer Science (EECS) and Institute for Medical Engineering & Science (IMES), Massachusetts Institute of Technology, Cambridge, United States
Md Ashrafuzzaman Department of Biomedical Engineering, Military Institute of Science and Technology, Dhaka, Bangladesh
Melanie Goisauf BBMRI-ERIC, ELSI Services & Research, Graz, Austria
Mohammad Yaqub Mohamed Bin Zayed University of Artificial Intelligence, Abu Dhabi, United Arab Emirates
Mohammed Ammar Engineering Systems and Telecommunication Laboratory, University M'Hamed Bougara, Boumerdes, Algeria
Mónica Cano Abadía BBMRI-ERIC, ELSI Services & Research, Graz, Austria
Mukhtar M E Mahmoud Faculty of Computer Science and Information Technology, University of Kassala, Kassala, Sudan
Mustafa Elattar Center for Informatics Science, Nile University, Sheikh Zayed City, Egypt
Nicola Rieke Healthcare & Life Science EMEA, NVIDIA GmbH, Munich, Germany
Nikolaos Papanikolaou Computational Clinical Imaging Group, Champalimaud Foundation, Lisbon, Portugal
Noussair Lazrak Health, Environment and policy, New York University, New York, United States
Oliver Díaz Artificial Intelligence in Medicine Lab (BCN-AIM), Department of Mathematics and Computer Science, Universitat de Barcelona, Barcelona, Spain
Olivier Salvado Data61, The Commonwealth Scientific and Industrial Research Organisation (CSIRO), Canberra, Australia
Oriol Pujol Artificial Intelligence in Medicine Lab (BCN-AIM), Department of Mathematics and Computer Science, Universitat de Barcelona, Barcelona, Spain
Ousmane Sall Pôle Sciences, Technologies et Numérique, Université Virtuelle du Sénégal, Diamniadio, Senegal
Pamela Guevara Faculty of Engineering, Universidad de Concepción, Concepción, Chile
Peter Gordebeke European Institute for Biomedical Imaging Research, Vienna, Austria
Philippe Lambin The D-lab, Department of Precision Medicine, GROW - School for Oncology and Reproduction, Maastricht University, Maastricht, the Netherlands
Pieta Brown Orion Health, Auckland, New Zealand
Purang Abolmaesumi Department of Electrical and Computer Engineering, the University of British Columbia, Vancouver, BC, Canada
Qi Dou Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, China
Qinghua Lu Data61, The Commonwealth Scientific and Industrial Research Organisation (CSIRO), Canberra, Australia
Richard Osuala Artificial Intelligence in Medicine Lab (BCN-AIM), Department of Mathematics and Computer Science, Universitat de Barcelona, Barcelona, Spain
Rose Nakasi Makerere Artificial Intelligence Lab, Makerere University, Kampala, Uganda
S Kevin Zhou School of Biomedical Engineering & Suzhou Institute for Advanced Research, University of Science and Technology of China, Suzhou, China

Sandy Napel Integrative Biomedical Imaging Informatics at Stanford (IBIIS), Department of Radiology, Stanford University, Stanford CA, United States
Sara Colantonio Institute of Information Science and Technologies of the National Research Council of Italy, Pisa, Italy
Shadi Albarqouni Department of Diagnostic and Interventional Radiology, University Hospital Bonn, Bonn, Germany
Smriti Joshi Artificial Intelligence in Medicine Lab (BCN-AIM), Department of Mathematics and Computer Science, Universitat de Barcelona, Barcelona, Spain
Stacy Carter Australian Centre for Health Engagement, Evidence and Values, School of Health and Society, University of Wollongong, New South Wales, Australia
Stefan Klein Department of Radiology & Nuclear Medicine, Erasmus MC University Medical Center, Rotterdam, the Netherlands
Steffen E Petersen William Harvey Research Institute, Queen Mary University of London, London, United Kingdom
Susanna Aussó Artificial Intelligence in Healthcare Program, TIC Salut Social Foundation, Barcelona, Spain
Suyash Awate, Computer Science and Engineering Department, Indian Institute of Technology Bombay, Mumbai, India
Tammy Riklin Raviv School of Electrical and Computer Engineering, Ben-Gurion University, Beer Sheba, Israel
Tessa Cook Department of Radiology, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, United States
Tinashe E M Mutsvangwa Department of Human Biology, University of Cape Town, Cape Town, South Africa
Wendy A Rogers Department of Philosophy, and School of Medicine, Macquarie University, Sydney, Australia
Wiro J Niessen Department of Radiology & Nuclear Medicine, Erasmus MC University Medical Center, Rotterdam, the Netherlands
Xènia Puig-Bosch Artificial Intelligence in Medicine Lab (BCN-AIM), Department of Mathematics and Computer Science, Universitat de Barcelona, Barcelona, Spain
Yi Zeng Center for Long-term AI, Chinese Academy of Sciences, Beijing, China
Yunusa G Mohammed Department of Human Anatomy, Gombe State University, Gombe, Nigeria
Yves Saint James Aquino Australian Centre for Health Engagement, Evidence and Values, School of Health and Society, University of Wollongong, New South Wales, Australia
Zohaib Salahuddin The D-lab, Department of Precision Medicine, GROW - School for Oncology and Reproduction, Maastricht University, Maastricht, the Netherlands
Martijn P A Starmans Department of Radiology & Nuclear Medicine, Erasmus MC University Medical Center, Rotterdam, the Netherlands