



Course 5

Trustworthy and Explainable AI in Health

March 13th, 2026

Χαρίδημος Κονδυλάκης
Αναπληρωτής Καθηγητής, Πανεπιστήμιο Κρήτης
Συνεργαζόμενος Ερευνητής, Ίδρυμα Τεχνολογίας και Έρευνας



ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ
UNIVERSITY OF CRETE



FORTH
INSTITUTE OF COMPUTER SCIENCE

NEWS

Home | Coronavirus | Climate | Video | World | UK | Business | Tech | Science | Stories | Entertainment & Arts

Health | Coronavirus

AI 'outperforms' doctors diagnosing breast cancer

Fergus Walsh
Medical correspondent
@BBCFergusWalsh

2 January 2020

AI in Healthcare

News | Opinion | Sport | Culture | Lifestyle

Coronavirus World UK Environment Science Global development Football Tech Business Obituaries

Artificial intelligence (AI)

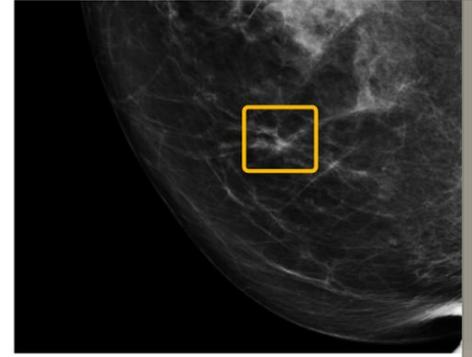
This article is more than 3 years old

AI system as good as experts at recognising skin cancers, say researchers

Deep learning-based system could be further developed for smartphones, increasing access to screening and aiding early detection of cancers

Nicola Davis
@NicolaKSDavis
Wed 25 Jan 2017 18:00 GMT

267 39



BBC Sign in

Home News Sport Reel Worklife

NEWS

Home | War in Ukraine | Climate | Video | World | UK | Business | Tech | Science | Stories

Tech

NHS uses AI scan to detect hidden heart disease

29 March 2021 · Comments

GETTY IMAGES

The technology could help save "thousands of lives"

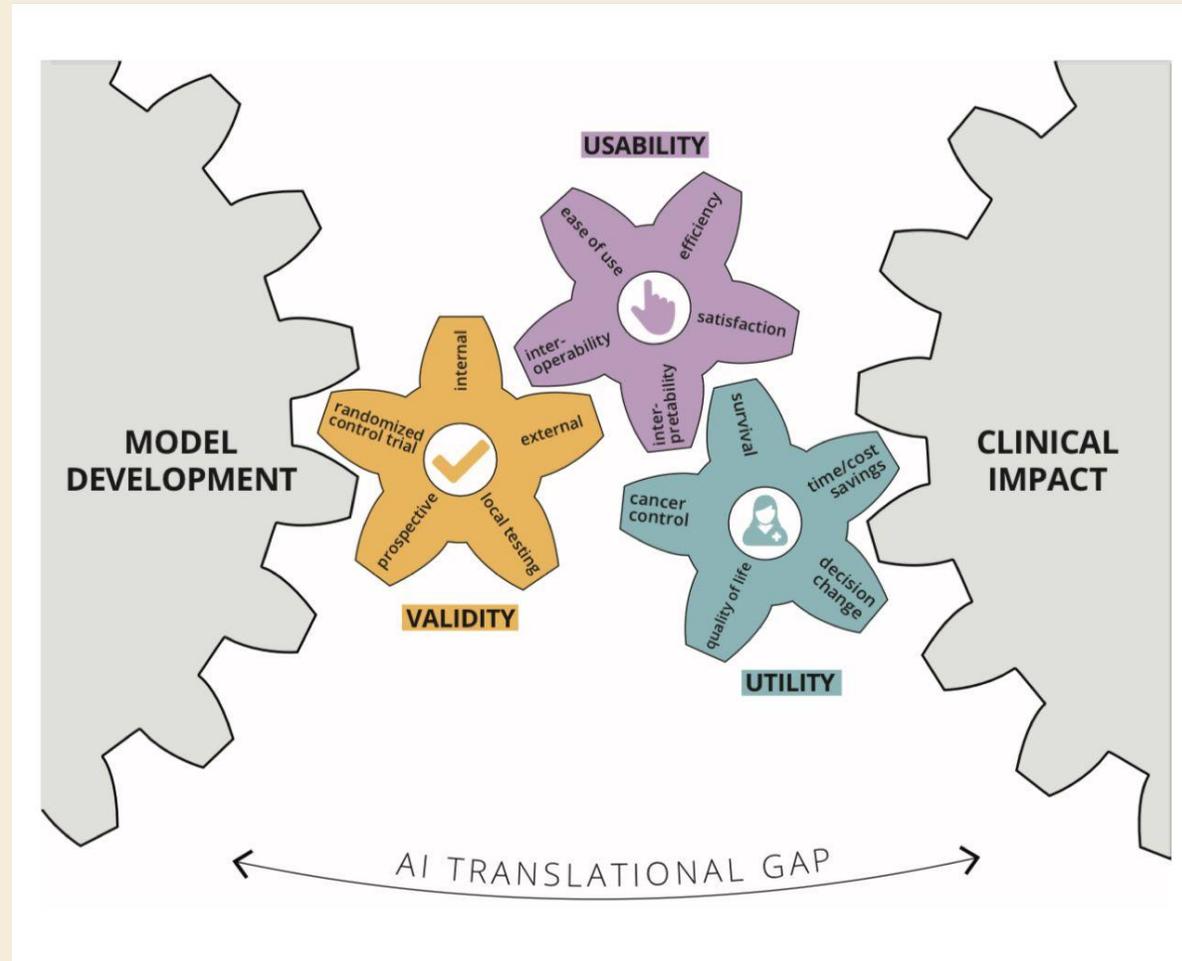
Tumours missed by 6 clinicians
Detected by the AI tool



Less than **0.2%** of Radiomic signatures are translated in the clinical practice !!!



AI 2.0: Thinking beyond predictions



AI Chasm

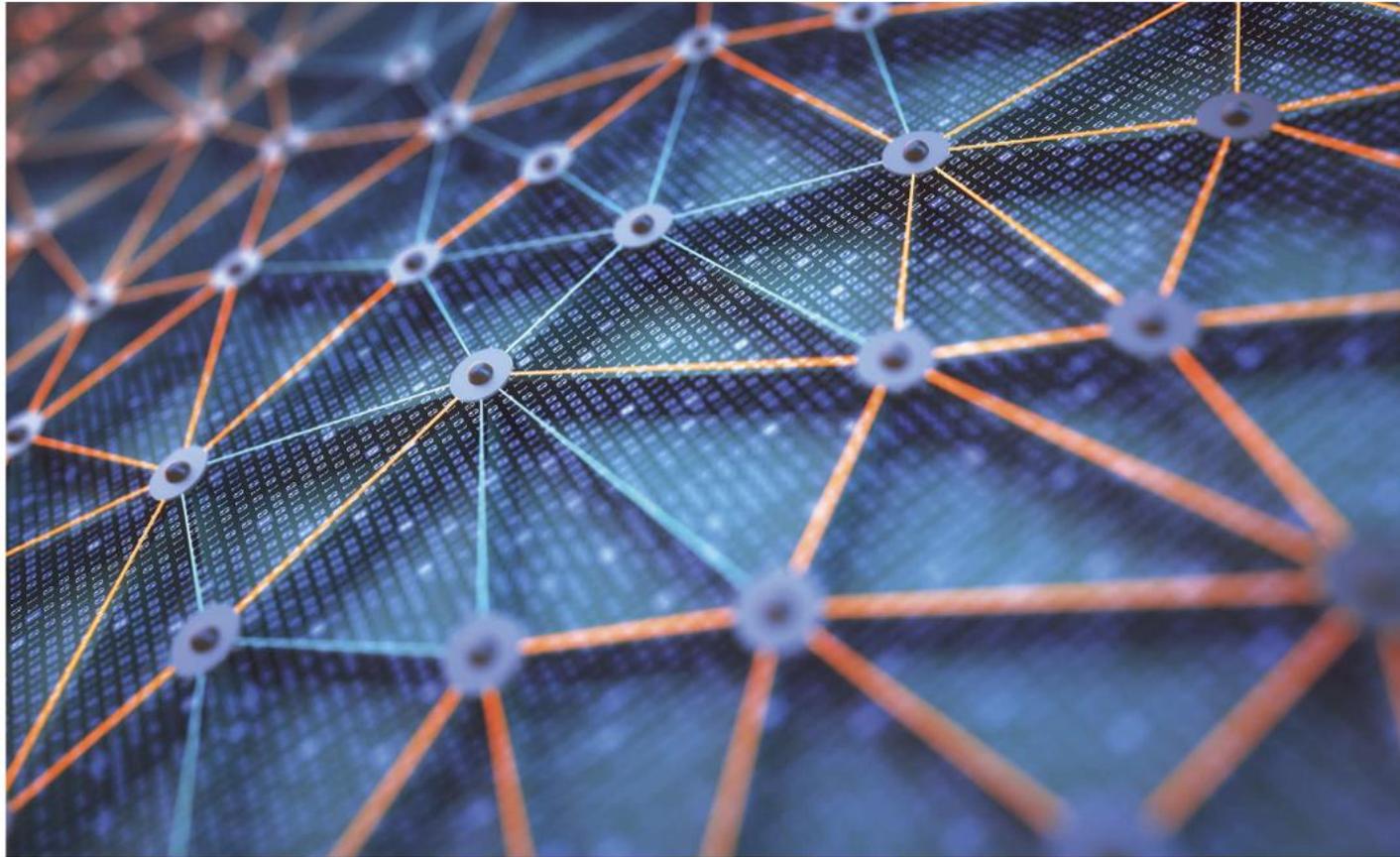


Despite indications that some **AI-based algorithms now match the accuracy of human experts** within preclinical *in silico* studies

Little high-quality evidence for improved clinical performance or patient outcomes in clinical studies

- **lack of necessary expertise needed for translating a tool into practice,**
- **lack of funding available for translation,**
- **lack of trust**

Work / Technology & tools



KTSIMAGE/GETTY

The use of artificial intelligence in medicine is growing rapidly.

THE REPRODUCIBILITY ISSUES THAT HAUNT HEALTH-CARE AI

Health-care systems are rolling out artificial-intelligence tools for diagnosis and monitoring. But how reliable are the models? **By Emily Sohn**

“Almost all of these award-winning models failed miserably. That was kind of surprising to us.”

Sad but (sometimes) true

npj | precision oncology

Published in partnership with The Hormel Institute, University of Minnesota

Comment



<https://doi.org/10.1038/s41698-024-00553-6>

All models are wrong and yours are useless: making clinical prediction models impactful for patients

Florian Markowetz

communications medicine

[Explore content](#) ▾ [About the journal](#) ▾ [Publish with us](#) ▾

[nature](#) > [communications medicine](#) > [articles](#) > article

Article | [Open access](#) | Published: 11 March 2025

Low responsiveness of machine learning models to critical or deteriorating health conditions

[Tanmoy Sarkar Pias](#), [Sharmin Afrose](#), [Moon Das Tuli](#), [Ipsita Hamid Trisha](#), [Xinwei Deng](#), [Charles B. Nemeroff](#)
& [Danfeng Daphne Yao](#)

[Communications Medicine](#) 5, Article number: 62 (2025) | [Cite this article](#)

41k Accesses | 2 Citations | 400 Altmetric | [Metrics](#)

What is the trust in AI-assisted health systems

npj | health systems

Perspective



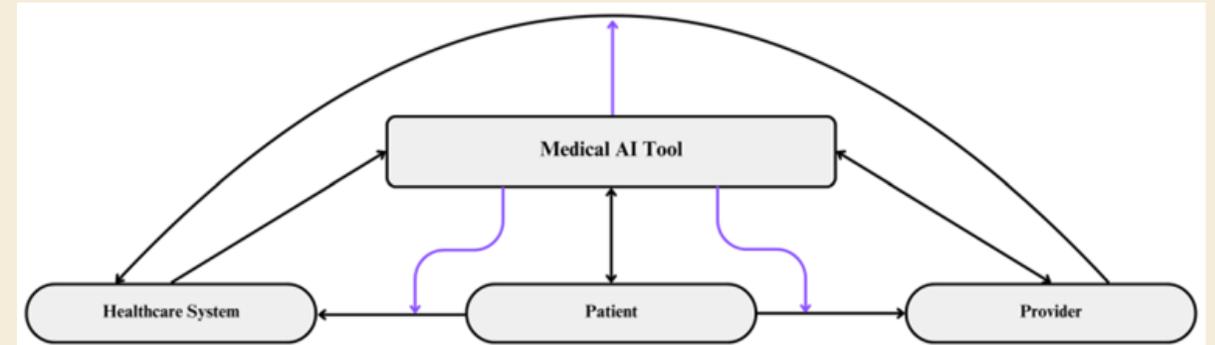
<https://doi.org/10.1038/s44401-025-00016-5>

Trust in AI-assisted health systems and AI's trust in humans

Check for updates

Madeline Sagona^{1,2}, Tinglong Dai^{2,3,4} ✉, Mario Macis^{2,3,5} & Michael Darden^{2,3}

Artificial intelligence (AI) is reshaping healthcare, promising improved diagnostics, personalized treatments, and streamlined operations. Yet a lack of trust remains a persistent barrier to widespread adoption. This Perspective examines the web of trust in AI-assisted healthcare systems, exploring the relationships it shapes, the systemic inequalities it can reinforce, and the technical challenges it poses. We highlight the bidirectional nature of trust, in which both patients and providers must trust AI systems, while these systems rely on the quality of human input to function effectively. Using models of care-seeking behavior, we explore the potential of AI to affect patients' decisions to seek care, influence trust in healthcare providers and institutions, and affect diverse demographic and clinical settings. We argue that addressing trust-related challenges requires rigorous empirical research, equitable algorithm design, and shared accountability frameworks. Ultimately, AI's impact hinges not just on technical progress but on sustaining trust, which may erode if biases persist, transparency falters, or incentives misalign.

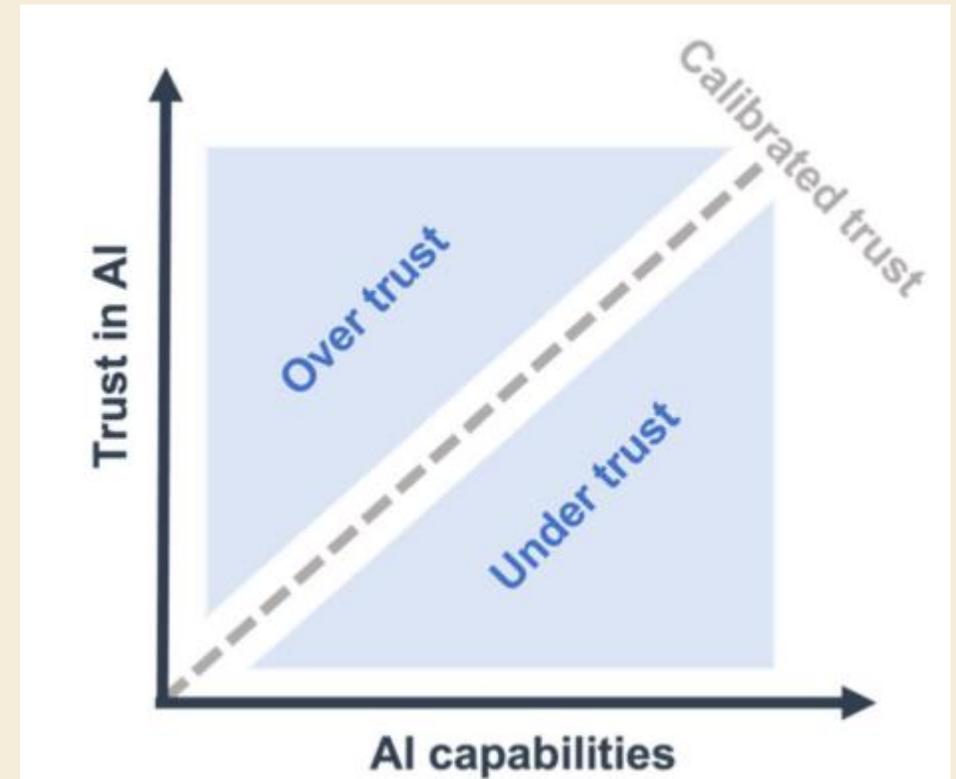


The **web of trust** between key stakeholders in AI assisted healthcare systems, including patients, providers, AI technologies, and healthcare institutions.

Sagona, Madeline, et al. "Trust in AI-assisted health systems and AI's trust in humans." *npj Health Systems* 2.1 (2025): 10.

Why do we need trust between people and AI?

Trust is a **central component** of interaction between people and artificial intelligence (AI) as well as machines and AI since **“incorrect” levels of trust** may cause **misuse, abuse, or disuse** of the technology.



Humanities & Social Sciences
Communications

REVIEW ARTICLE

<https://doi.org/10.1057/s41599-024-04044-8> OPEN

Trust in AI: progress, challenges, and future directions

Salah Afroogh^{1✉}, Ali Akbari², Emmie Malone³, Mohammadali Kargar⁴ & Hananeh Alambeigi⁵

Afroogh, Saleh, et al. "Trust in AI: progress, challenges, and future directions." *Humanities and Social Sciences Communications* 11.1 (2024): 1-30.

OPEN ACCESS

Check for updates

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

Karim Lekadir,^{1,2} Alejandro F Frangi,^{3,4} Antonio R Porras,⁵ Ben Glocker,⁶ Celia Cintas,⁷ Curtis P Langlotz,⁸ Eva Welcken,⁹ Folkert W Asselbergs,^{10,11} Fred Prior,¹² Gary S Collins,¹³ Georgios Kallissis,¹⁴ Gianna Tsakou,¹⁵ Irène Buvat,¹⁶ Jayashree Kalpathy-Cramer,¹⁷ John Mongan,¹⁸ Julia A Schnabel,¹⁹ Kaisar Kushibar,¹ Katrine Riklund,²⁰ Kostas Marias,²¹ Lameck M Amugongo,²² Lauren A Fromont,²³ Lena Maier-Hein,²⁴ Leonor Cerdá-Alberich,²⁵ Luis Martí-Bonmati,²⁶ M Jorge Cardoso,²⁷ Maciej Bobowicz,²⁸ Mahsa Shabani,²⁹ Manolis Tsiknakis,²¹ Maria A Zuluaga,³⁰ Marie-Christine Fritzsche,³¹ Marina Camacho,¹ Marius George Linguraru,³² Markus Wenzel,⁹ Marleen De Bruijne,³³ Martin G Tolsgaard,³⁴ Melanie Gotsauf,³⁵ Mónica Cano Abadía,³⁵ Nikolaos Papanikolaou,³⁶ Noussair Lazrak,¹ Oriol Pujol,¹ Richard Osuala,¹ Sandy Napel,³⁷ Sara Colantonio,³⁸ Smriti Joshi,¹ Stefan Klein,³³ Susanna Aussó,³⁹ Wendy A Rogers,⁴⁰ Zohaib Salahuddin,⁴¹ Martijn P A Starmans³³; on behalf of the FUTURE-AI Consortium

For numbered affiliations see end of the article
 Correspondence to: K Lekadir (ORCID 0000-0002-9456-1612)
 Additional material is published online only. To view please visit the journal online.
 Cite this as: *BMJ* 2025;388:e081554
<http://dx.doi.org/10.1136/bmj-2024-081554>
 Accepted: 10 January 2025

Despite major advances in artificial intelligence (AI) research for healthcare, the deployment and adoption of AI technologies remain limited in clinical practice. This paper describes the FUTURE-AI framework, which provides guidance for the development and deployment of trustworthy AI tools in healthcare. The FUTURE-AI Consortium was founded in 2021 and comprises 117 interdisciplinary experts from 50 countries representing all continents, including AI scientists, clinical researchers, biomedical ethicists, and social scientists. Over a two year period, the FUTURE-AI guideline was

established through consensus based on six guiding principles—fairness, universality, traceability, usability, robustness, and explainability. To operationalise trustworthy AI in healthcare, a set of 30 best practices were defined, addressing technical, clinical, socioethical, and legal dimensions. The recommendations cover the entire lifecycle of healthcare AI, from design, development, and validation to regulation, deployment, and monitoring.

Introduction

In the field of healthcare, artificial intelligence (AI)—that is, algorithms with the ability to self-learn logic—and data interactions have been increasingly used to develop computer aided models, for example, disease diagnosis, prognosis, prediction of therapy response or survival, and patient stratification.¹ Despite major advances, 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 societal risks associated with healthcare AI.^{2,3} In particular, existing research has shown that AI tools in healthcare 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.⁴⁻⁸

To increase adoption in the real world, it is essential that 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 healthcare AI tools should be designed, developed, evaluated, and deployed to be trustworthy—that is, technically robust, clinically safe,

SUMMARY POINTS

Despite major advances in medical artificial intelligence (AI) research, clinical adoption of emerging AI solutions remains challenging owing to limited trust and ethical concerns
 The FUTURE-AI Consortium unites 117 experts from 50 countries to define international guidelines for trustworthy healthcare AI
 The FUTURE-AI framework is structured around six guiding principles: fairness, universality, traceability, usability, robustness, and explainability
 The guideline addresses the entire AI lifecycle, from design and development to validation and deployment, ensuring alignment with real world needs and ethical requirements
 The framework includes 30 detailed recommendations for building trustworthy and deployable AI systems, emphasising multistakeholder collaboration
 Continuous risk assessment and mitigation are fundamental, addressing biases, data variations, and evolving challenges during the AI lifecycle
 FUTURE-AI is designed as a dynamic framework, which will evolve with technological advancements and stakeholder feedback

There is a necessity of **structured guidelines** to build AI-based healthcare solutions by design



FAIR	UNIVERSAL	TRACEABLE	USABLE	ROBUST	EXPLAINABLE

The journey of FUTURE-AI

FUTURE-AI: Best practices for trustworthy AI in medicine

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



<https://future-ai.eu/>



117

international and interdisciplinary experts

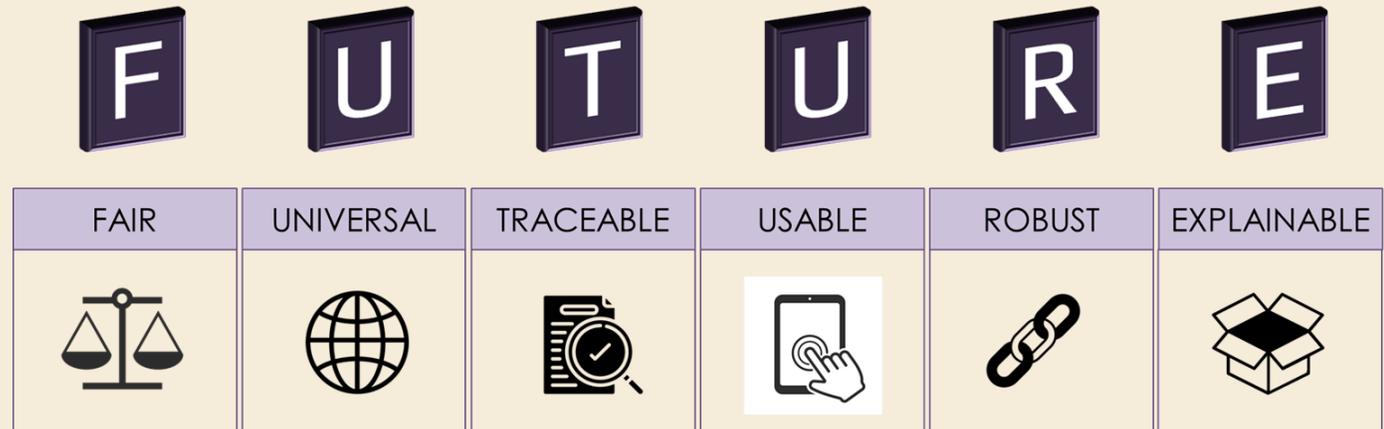
50

countries



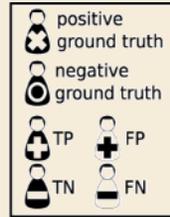
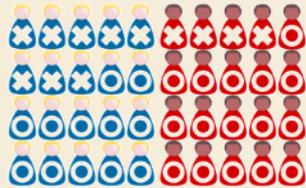
Characteristics of Trustworthy AI

- Fairness
- Universality
- Traceability
- Usability
- Robustness
- Explainability

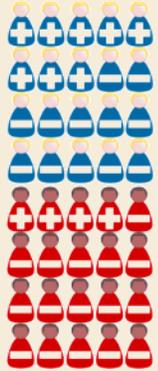


Fairness

population ground truth



model predictions



demographic parity

$$\frac{\# \text{TP} + \# \text{FP}}{\# \text{TP} + \# \text{FN} + \# \text{FP} + \# \text{FN}} = \frac{\# \text{TN} + \# \text{FN}}{\# \text{TN} + \# \text{FN} + \# \text{FP} + \# \text{FN}}$$

not fulfilled

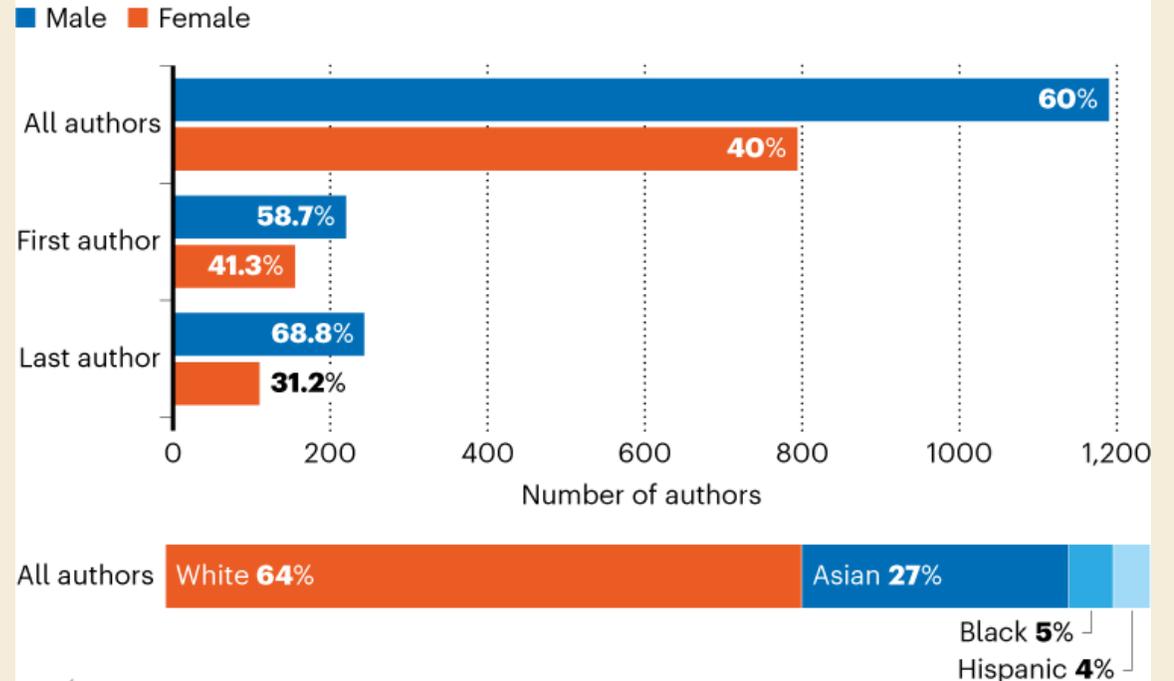
equal opportunity

$$\frac{\# \text{TP}}{\# \text{TP} + \# \text{FN}} = \frac{\# \text{TN}}{\# \text{TN} + \# \text{FN}}$$

fulfilled

GAPS IN REPRESENTATION

An analysis of 375 research papers on the fairness of artificial intelligence showed a lack of ethnic and gender diversity among authors.



<https://www.nature.com/articles/d41586-023-00935-z>

Ricci Lara, M.A., Echeveste, R. & Ferrante, E. Addressing fairness in artificial intelligence for medical imaging. *Nat Commun* 13, 4581 (2022).

Universality

nature

[Explore content](#) ▾ [About the journal](#) ▾ [Publish with us](#) ▾

[nature](#) > [articles](#) > [article](#)

Article | Published: 01 January 2020

International evaluation of an AI system for breast cancer screening

[Scott Mayer McKinney](#) , [Marcin Sieniek](#), [Varun Godbole](#), [Jonathan Godwin](#), [Natasha Antropova](#), [Hutan Ashrafian](#), [Trevor Back](#), [Mary Chesus](#), [Greg S. Corrado](#), [Ara Darzi](#), [Mozziyar Etemadi](#), [Florenca Garcia-Vicente](#), [Fiona J. Gilbert](#), [Mark Halling-Brown](#), [Demis Hassabis](#), [Sunny Jansen](#), [Alan Karthikesalingam](#), [Christopher J. Kelly](#), [Dominic King](#), [Joseph R. Ledsam](#), [David Melnick](#), [Hormuz Mostofi](#), [Lily Peng](#), [Joshua Jay Reicher](#), [Bernardino Romera-Paredes](#), [Richard Sidebottom](#), [Mustafa Suleyman](#), [Daniel Tse](#) , [Kenneth C. Young](#), [Jeffrey De Fauw](#) & [Shravya Shetty](#) 

[— Show fewer authors](#)

[Nature](#) 577, 89–94 (2020) | [Cite this article](#)

Test datasets

		
Number of women	25,856	3,097
Interpretation	Double reading	Single reading
Screening interval	3 years	1 or 2 years
Cancer follow-up	39 months	27 months
Number of cancers	414 (1.6%)	686 (22.2%)

AI vs. Doctor(s)

Equivalent

Superior

Traceability

Article

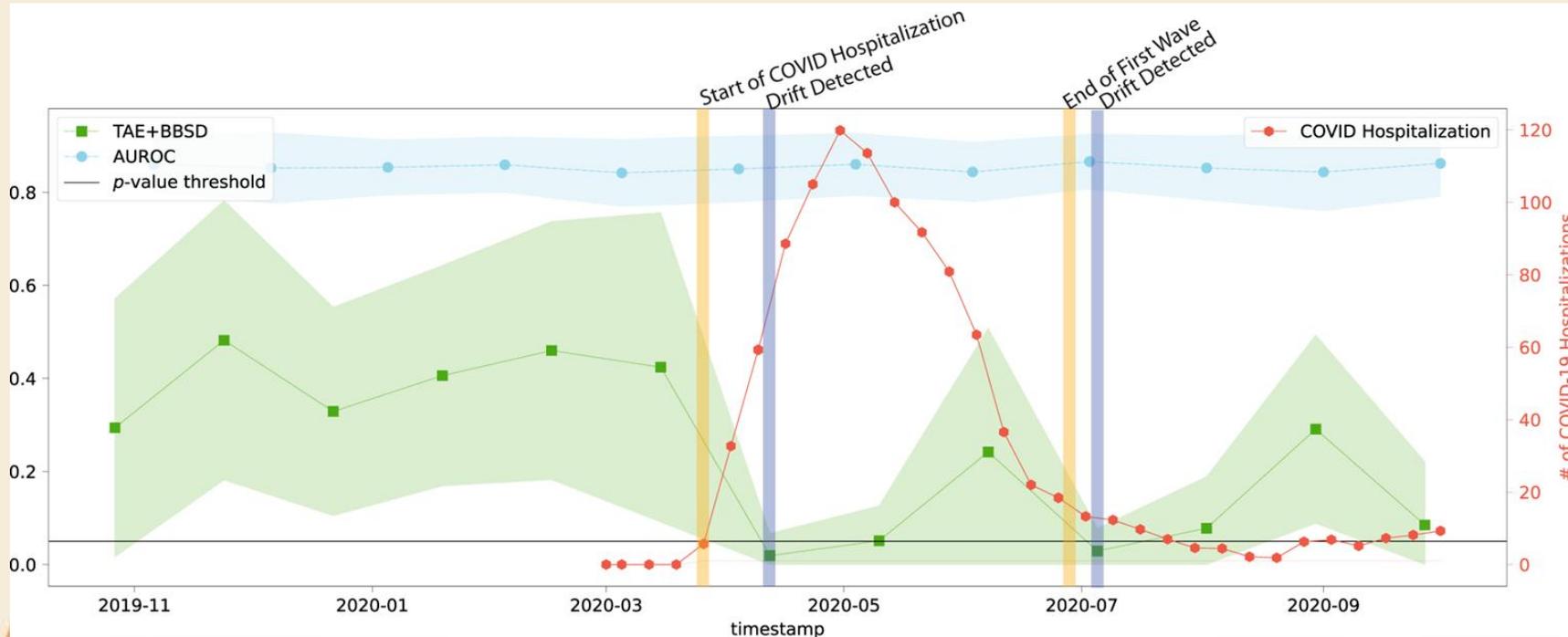
<https://doi.org/10.1038/s41467-024-46142-w>

Empirical data drift detection experiments on real-world medical imaging data

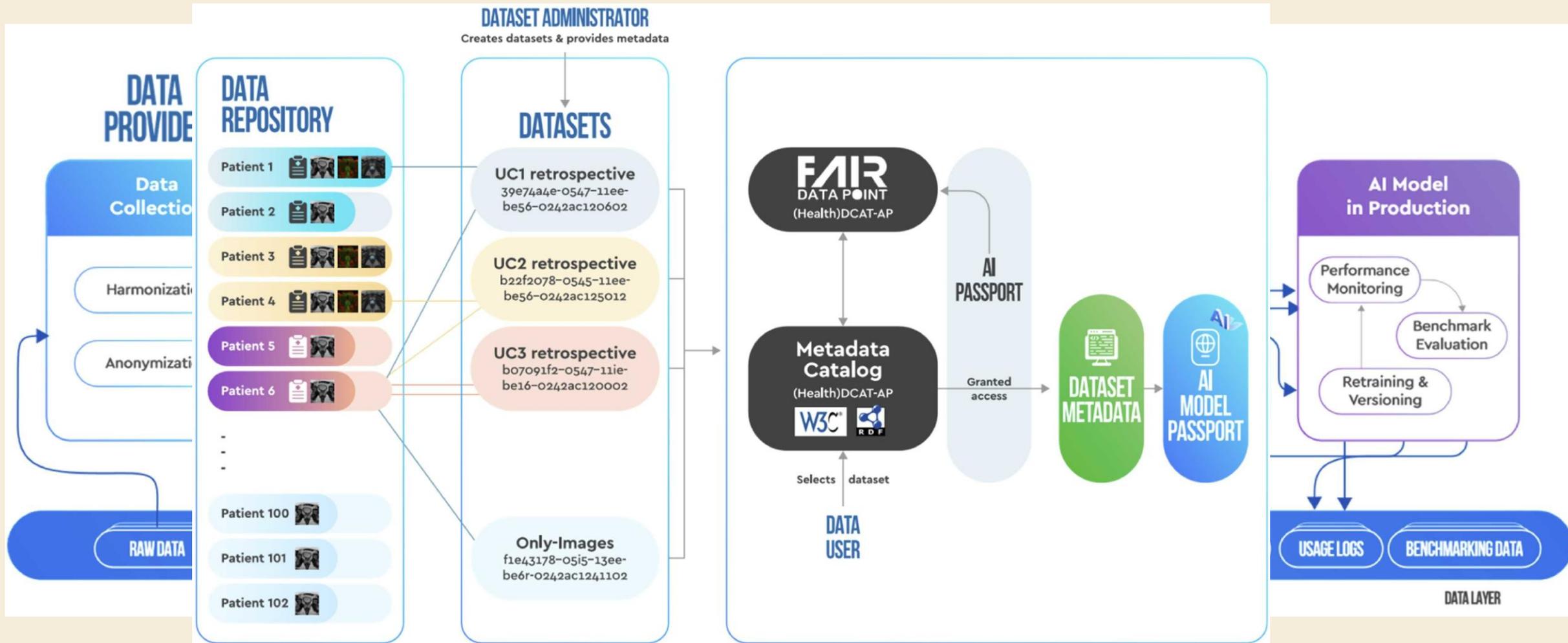
Received: 31 July 2023

Ali Kore¹, Elyar Abbasi Bavit², Vallijah Subasri³, Moustafa Abdalla⁴, Benjamin Fine^{5,6}, Elham Dolatabadi^{1,7} & Mohamed Abdalla⁵ ✉

Accepted: 14 February 2024



Traceability



Usability

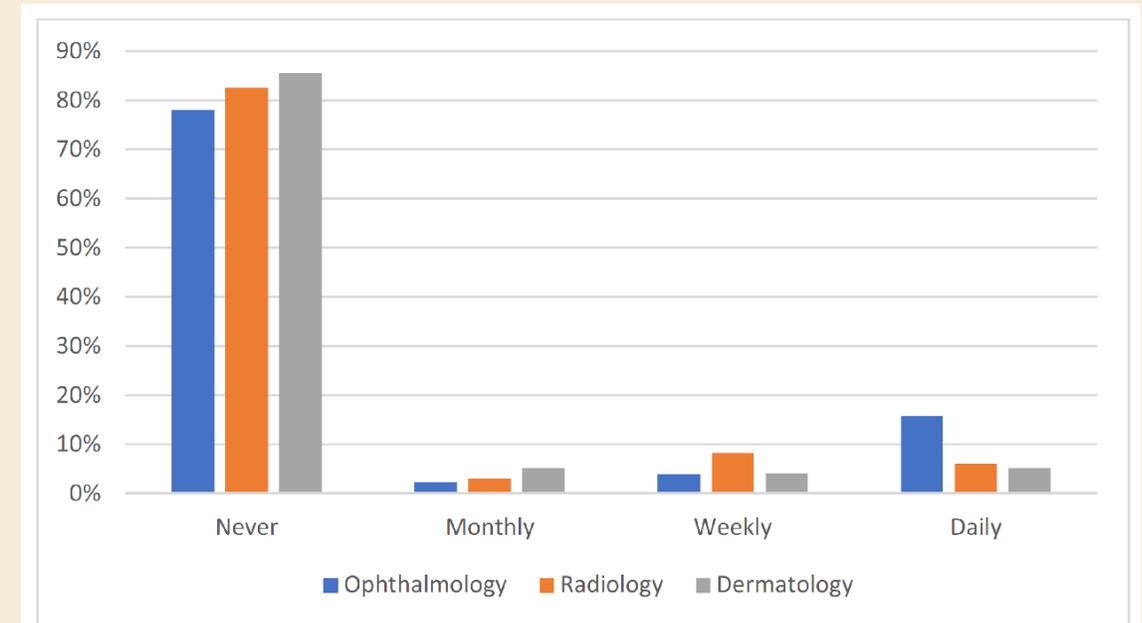


Figure 2. Current frequency of artificial intelligence use in clinical practice.

Scheetz, J. et al. 2021, A survey of clinicians on the use of AI, Scientific reports, 11(1), p.5193.

Robustness

nature

[Explore content](#) ▾ [About the journal](#) ▾ [Publish with us](#) ▾

[nature](#) > [articles](#) > [article](#)

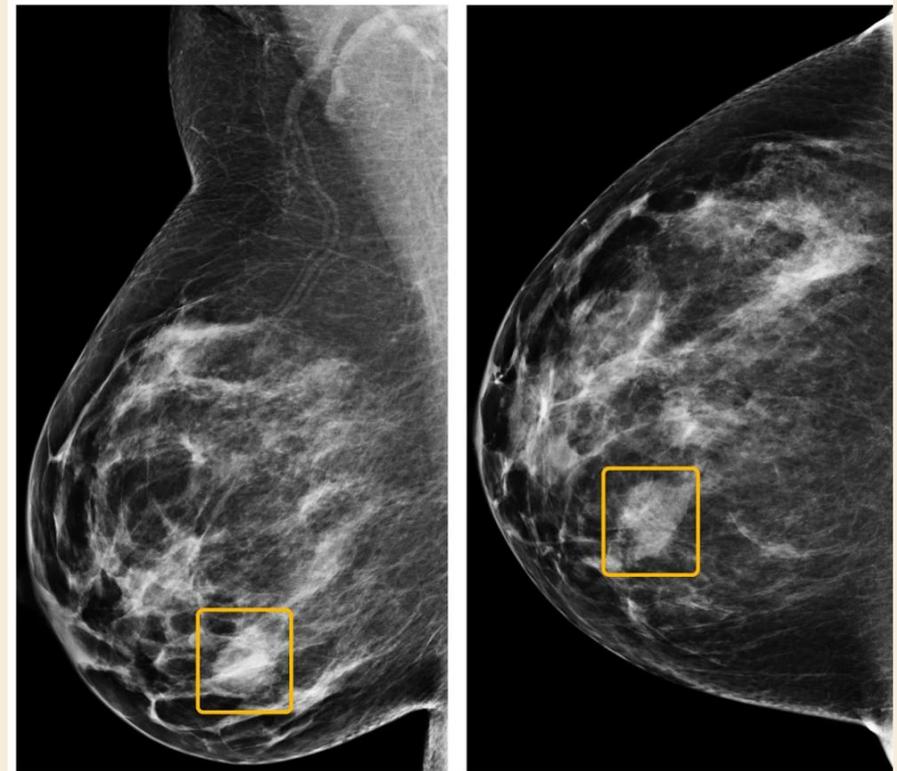
Article | Published: 01 January 2020

International evaluation of an AI system for breast cancer screening

[Scott Mayer McKinney](#) , [Marcin Sieniek](#), [Varun Godbole](#), [Jonathan Godwin](#), [Natasha Antropova](#), [Hutan Ashrafian](#), [Trevor Back](#), [Mary Chesus](#), [Greg S. Corrado](#), [Ara Darzi](#), [Mozziyar Etemadi](#), [Florenca Garcia-Vicente](#), [Fiona J. Gilbert](#), [Mark Halling-Brown](#), [Demis Hassabis](#), [Sunny Jansen](#), [Alan Karthikesalingam](#), [Christopher J. Kelly](#), [Dominic King](#), [Joseph R. Ledsam](#), [David Melnick](#), [Hormuz Mostofi](#), [Lily Peng](#), [Joshua Jay Reicher](#), [Bernardino Romera-Paredes](#), [Richard Sidebottom](#), [Mustafa Suleyman](#), [Daniel Tse](#) , [Kenneth C. Young](#), [Jeffrey De Fauw](#) & [Shravya Shetty](#) 

— Show fewer authors

[Nature](#) 577, 89–94 (2020) | [Cite this article](#)



Tumours detected by 6 clinicians
Missed by the AI tool

Explainability

REVIEW ARTICLE OPEN

Check for updates

Explainable artificial intelligence for mental health through transparency and interpretability for understandability

Dan W. Joyce^{1,2}, Andrey Kormilitzin¹, Katharine A. Smith^{1,3,4} and Andrea Cipriani^{1,3,4}

Example Induction

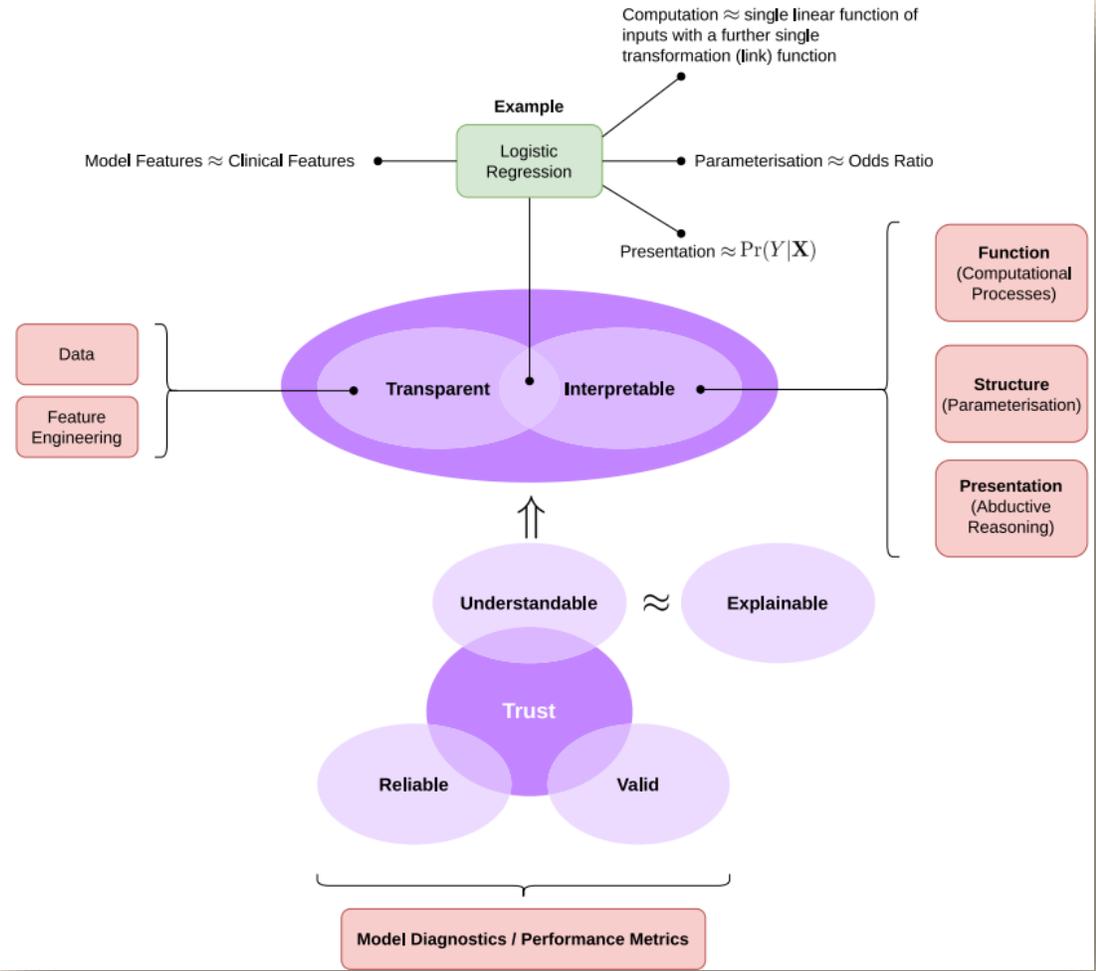
	N = 50	Psychosis
Data	No Abnormal Beliefs	10
	Abnormal Beliefs	40

80% of people with **Psychosis** have **Abnormal Beliefs**
 An individual x has **Psychosis**
 Therefore, the **probability** that x has **Abnormal Beliefs** = 0.8

Example Abduction

	N = 100	Psychosis	Depression
Data	No Abnormal Beliefs	10	45
	Abnormal Beliefs	40	5

80% of people with **Psychosis** have **Abnormal Beliefs**
 10% of people with **Depression** have **Abnormal Beliefs**
 An individual x has **Abnormal Beliefs**
 The probability of **Psychosis** given **Abnormal Beliefs** = 0.89
 The probability of **Depression** given **Abnormal Beliefs** = 0.11
 Therefore, the **diagnosis** that best accounts for x having **Abnormal Beliefs** is **Psychosis**



Joyce, D.W., Kormilitzin, A., Smith, K.A. *et al.* Explainable artificial intelligence for mental health through transparency and interpretability for understandability. *npj Digit. Med.* 6, 6 (2023).

Characteristics of Trustworthy AI

Based on ethical principles and fundamental rights:

FAIRNESS



Right to non-discrimination

UNIVERSALITY



Right to equity

TRACEABILITY



Right to accountability

USABILITY



Right to autonomy

ROBUSTNESS



Right to safety

EXPLAINABILITY



Right to transparency



Characteristics of Trustworthy AI

	Clusters of requirements	Core principle
1	Diversity, Inclusiveness, Non-discrimination, Bias, Equity	<u>F</u> airness
2	Generalisability, Adaptability, Interoperability, Applicability	<u>U</u> niversality
3	Transparency, Monitoring, Auditing, Accountability	<u>T</u> raceability
4	Human-centred AI, User engagement, Accessibility, Efficiency	<u>U</u> sability
5	Reliability, Resilience, Safety, Security	<u>R</u> obustness
6	Interpretability, Understandability, Transparency	<u>E</u> xplainability



Table 2 | List of FUTURE-AI recommendations, together with the expected compliance for both research and deployable artificial intelligence (AI) tools (+: recommended, ++: highly recommended)

Recommendations	Research	Deployable
Fairness		
1. Define any potential sources of bias from an early stage	++	++
2. Collect information on individuals' and data attributes	+	+
3. Evaluate potential biases and, when needed, bias correction measures	+	++
Universality		
1. Define intended clinical settings and cross setting variations	++	++
2. Use community defined standards (eg, clinical definitions, technical standards)	+	+
3. Evaluate using external datasets and/or multiple sites	++	++
4. Evaluate and demonstrate local clinical validity	+	++
Traceability		
1. Implement a risk management process throughout the AI lifecycle	+	++
2. Provide documentation (eg, 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 AI governance	+	++
Usability		
1. Define intended use and user requirements from an early stage	++	++
2. Establish mechanisms for human-AI interactions and oversight	+	++
3. Provide training materials and activities (eg, tutorials, hands-on sessions)	+	++
4. Evaluate user experience and acceptance with independent end users	+	++
5. Evaluate clinical utility and safety (eg, effectiveness, harm, cost-benefit)	+	++
Robustness		
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	++	++
Explainability		
1. Define the need and requirements for explainability with end users	++	++
2. Evaluate explainability with end users (eg, correctness, impact on users)	+	+
General		
1. Engage interdisciplinary stakeholders throughout the AI lifecycle	++	++
2. Implement measures for data privacy and security	++	++
3. Implement measures to address identified AI risks	++	++
4. Define adequate evaluation plan (eg, datasets, metrics, reference methods)	++	++

BMJ: first published as 10.1136/bmj-2024-081554 on 5 February 2025. Downloaded from https://www.bmj.com/ on 05 February 2025. Protected by copyright, including for uses related to text and data mining, AI technologies, and similar technologies.



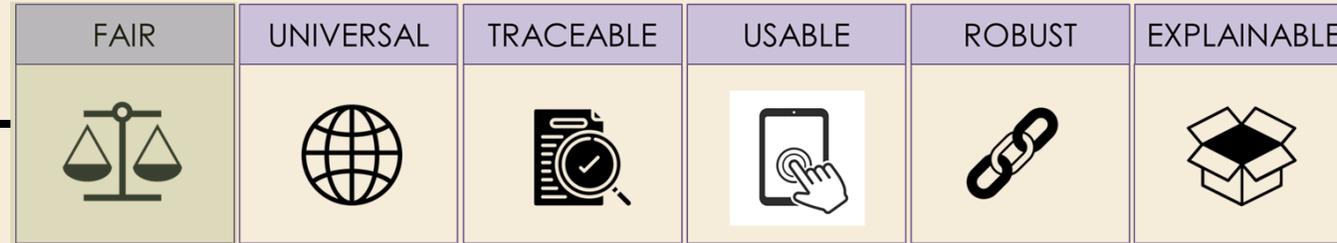
AI	TRACEABLE	USABLE	ROBUST	EXPLAINABLE



OPERATIONALISATION



Operationalisation



Recommendations	Research	Deployable
Fairness		
1. Define any potential sources of bias from an early stage	++	++
2. Collect information on individuals' and data attributes	+	+
3. Evaluate potential biases and, when needed, bias correction measures	+	++

Recommendations	Operations	Examples
Define any potential sources of bias (fairness 1)	Engage relevant stakeholders to define the sources of bias	Patients, clinicians, epidemiologists, ethicists, social carers ^{97 98}
	Define standard attributes that might affect the AI tool's fairness	Sex, age, socioeconomic status ⁹⁹
	Identify application specific sources of bias beyond standard attributes	Skin colour for skin cancer detection, ^{100 101} breast density for breast cancer detection ³⁴
	Identify all possible human biases	Data labelling, data curation ⁹⁹



Operationalisation

Recommendations	Methods	Explanation	Example Refs
Define intended use and user requirements from an early stage.	Record user needs through focus groups, design a user-friendly UI, and integrate clinical information systems.	Tailor the system to specific clinical needs to enhance usability and adoption. User-centered design and adherence to human factors engineering principles can improve the usability and adoption of radiology AI systems.	[62], [92]–[94]
Establish mechanisms for human-AI interactions and oversight.	Provide dashboards, automate system log analysis, and implement feedback mechanisms and uncertainty estimates.	These tools support error detection, provide transparency into decision-making processes, and facilitate clinical decision-making by enabling clinicians to assess the AI's reliability.	[95]
Provide training materials and activities.	Create training resources like user guides and workshops, and implement training courses for radiologists.	Training materials reduce perceived complexity and enhance user understanding, while courses ensure efficient usage and awareness of AI system limitations.	[62], [96]–[101]
Evaluate user experience and acceptance with independent end-users.	Conduct performance testing, user satisfaction studies, usability questionnaires and interviews, and efficiency evaluations.	User experience testing in clinical environments identifies usability issues, evaluates user satisfaction, and highlights areas for improvement. Involving diverse clinical users in design phases can further support system adoption.	[95], [102]–[104]
Evaluate clinical utility and safety (e.g., effectiveness, harm, cost-benefit).	Implement feedback mechanisms, design systems to warn about critical errors, and address risk management processes.	Ensuring minimal errors and providing feedback mechanisms enhance the clinical utility and safety of AI systems. These actions also foster trust and reliability while addressing skepticism around AI in medical imaging.	[105]–[109]

TABLE IV
RECOMMENDATIONS AND METHODS FOR USABILITY.

(1) DESIGN STAGE

- G1: Engage inter-disciplinary stakeholders
- Us1: Define intended use and user requirements
- Un1: Define clinical settings and user needs
- R1: Define all sources of data for training and validation
- F1: Define all sources of bias and uncertainty
- E1: Define explainability needs and requirements
- Un2: Use community-defined standards and guidelines
- G6: Investigate application-specific requirements
- G7: Investigate social and societal implications
- T1: Implement a risk management process

Guidelines for trustworthy AI in FUTURE-AI Guidelines

Bosch, Noussair Lazrak, Oliver Diaz, Kaisar Kushibar, PA Starmans, Sara Colantonio, Nikos Tachos, Smriti Ina Tsakou, Susanna Aussó, Leonor Cerdá Alberich, Marias, Manolis Tsiknakis, Dimitrios I. Fotiadis, Luis if, Karim Lekadir

Trustworthiness, Recommendations, Guidelines, Healthcare

I. INTRODUCTION

Artificial intelligence (AI) is considered a highly promising and disruptive technology for future healthcare, particularly in medical imaging. AI developments have significantly impacted image reconstruction, medical image segmentation, image-based diagnosis, and decision on treatment planning [1].

AI can enhance medical imaging by improving the acquisition, processing, and interpretation of images, helping clinicians diagnose and manage patients more efficiently and accurately. However, despite these advances, the adoption of AI in clinical practice remains limited, marking a substantial gap between technical proof-of-concepts and clinical implementation. A survey in Australia and New Zealand revealed that although most radiologists strongly believe AI could improve their field, over 80% have not used AI in their daily practice [2]. Another survey by the European Society of Radiology showed that only 40% of radiologists had experience with AI in their practice, while only 10% had interest in acquiring AI-based tools [3].

In addition to radiologists, AI implementation also impacts radiographers and medical physicists, who are integral to imaging workflows. Radiographers have indirect influence over data quality (e.g., by managing the settings) and acquisition protocols, while medical physicists ensure safety and calibration, both of which affect AI performance. Recent studies highlight the training needs and workflow challenges these professionals face [4], [5]. Addressing these issues calls for a multidisciplinary approach, with all clinical stakeholders engaged from the design stage onward.

Concerns about AI include mainly potential risks, ethical implications, and a lack of trust due to its complexity and opacity. There is a risk of AI generating undetected errors, especially when applied to imaging conditions differing from those used for training. Furthermore, imbalanced imaging databases can lead to biased AI algorithms, exacerbating health disparities [6]–[8]. There are also concerns about the impact of AI on radiologists' decision-making and interpretation skills [9]. Current AI solutions often lack mechanisms for ongoing

Take-Away Messages

1. Awareness of Trustworthy AI is critical.
2. FUTURE-AI is a structured guideline for trustworthy AI.
3. It requires active multi-stakeholder engagement.
4. FUTURE-AI is a trustworthy-by-design approach.
5. AI evaluation should be multi-faceted.
6. Trustworthiness is a continuous process.

✓ Special thanks to Karim Lekadir and Nikos Papanikolaou for providing some of the slides



FUTURE AI

RELATED PROJECTS



OUR PARTNERS



JOIN THE FUTURE-AI NETWORK



<https://future-ai.eu/contact-us/>



FOLLOW THE UPDATES

<https://future-ai.eu/news-and-events/>

FUTURE AI Webinar Series



12 FEBRUARY 2026

FUTURE-AI Anniversary Webinar | 12 February 2026

FUTURE-AI Anniversary Webinar: "The Future of AI and the Radiology Workforce" by Curtis P. Langlotz, MD, PhD. We are ...

FUTURE AI Webinar Series



Recording available

FUTURE-AI Webinar #6 | October 31, 2025

FUTURE-AI Webinar: "From Prototype to Practice: Bridging the Medical AI Implementation Gap" by Dr. Judy Wawira Gichoya...

FUTURE AI Webinar Series



Recording available

FUTURE-AI Webinar #5 | July 23, 2025

Upcoming FUTURE-AI Webinar: "A Step for AI Co-Pilot in Robotic Surgery" by Prof. Qi Dou We are pleased to...





Thank you!



ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ
UNIVERSITY OF CRETE



FORTH
INSTITUTE OF COMPUTER SCIENCE