Prescribing trust:
Exploring the Future of AI in Healthcare

Oliver Díaz

University of Barcelona

Artificial Intelligence in Medicine Lab (BCM-AIM)

oliver.diaz@ub.edu

Disclosure

- I have no financial interests or relationships to disclose.

Objectives of the session

- Talk about TRUST

- Discuss limitations of current AI tools.

  - Why they are not reaching the clinical market?

- Present FUTURE-AI guidelines

What is TRUST?

TRUST.

1. firm belief in the reliability, truth, ability, or strength of someone or something.
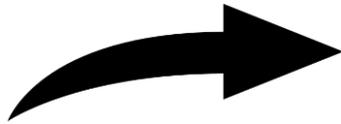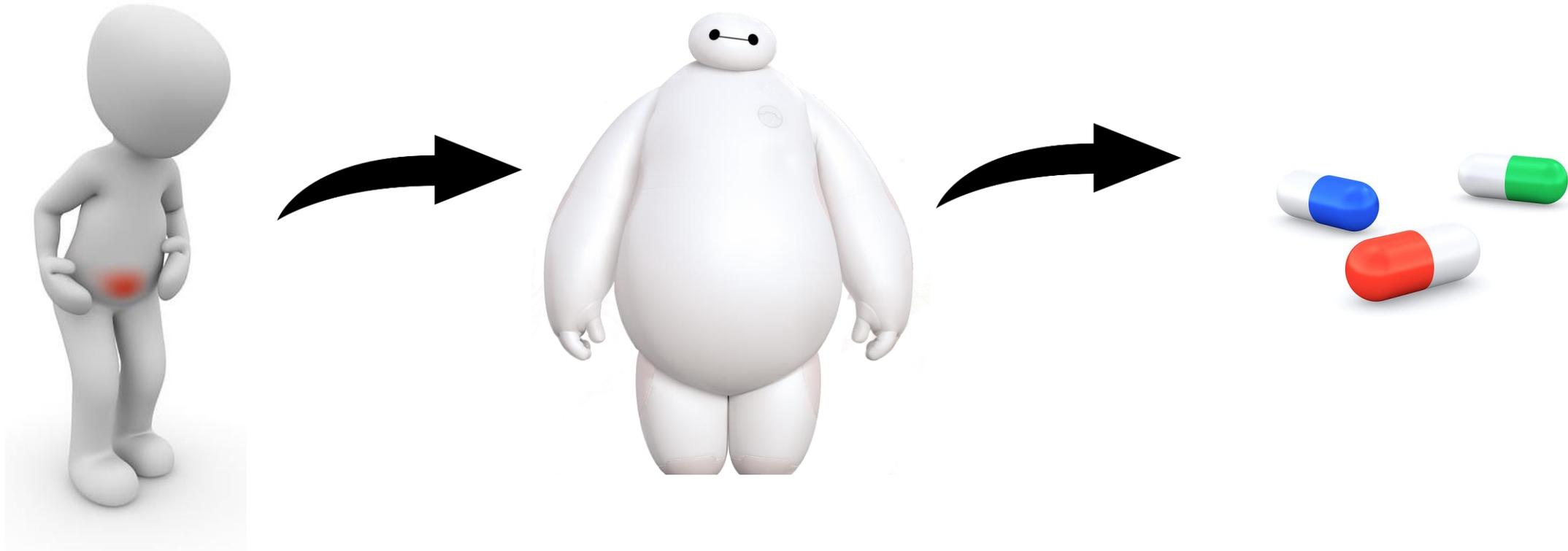
How do we build TRUST?

How do we build TRUST?

How do we build TRUST?

What does TRUST mean?

- **Have confidence** in the diagnosis/results
- **Believe** the process
- **Rely on** the tool/person
- **Have faith in** AI/technique
- **Entrust** the output of the model/doctor
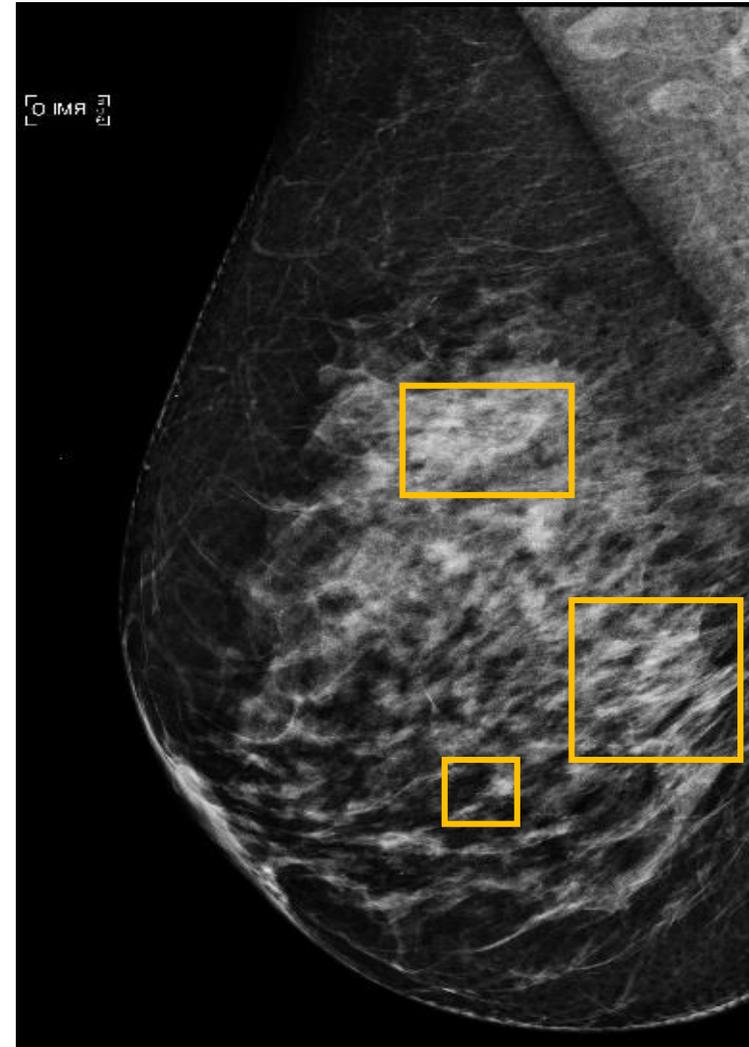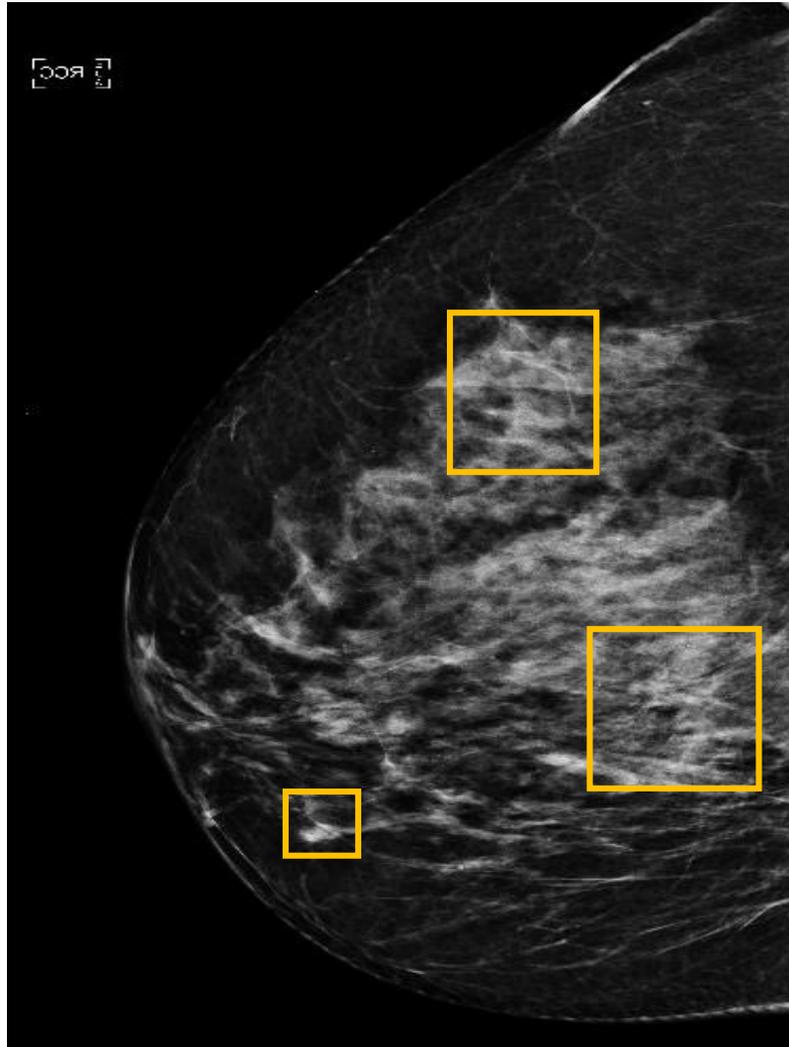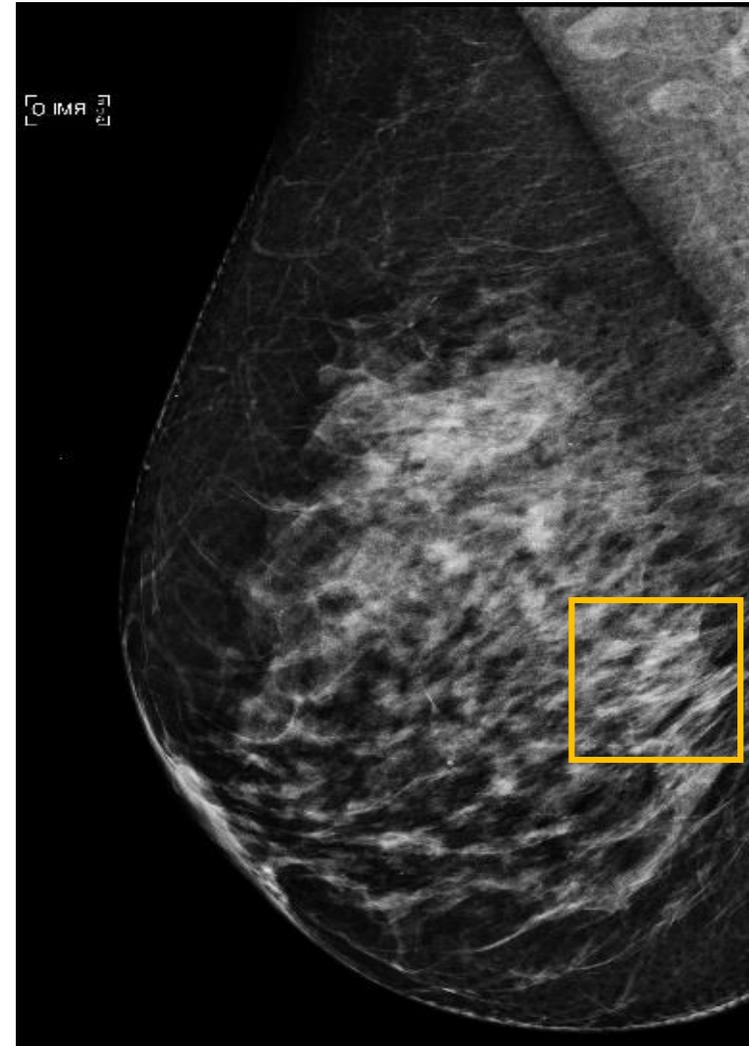- **Confidence** in the person/AI

Role-playing game

Let's have a look at this case

48yo woman. Asymptomatic. Heterogeneously dense breast (ACR C).
Courtesy of Dr. Javier Del Riego – Hospital Universitari Parc Taulí.
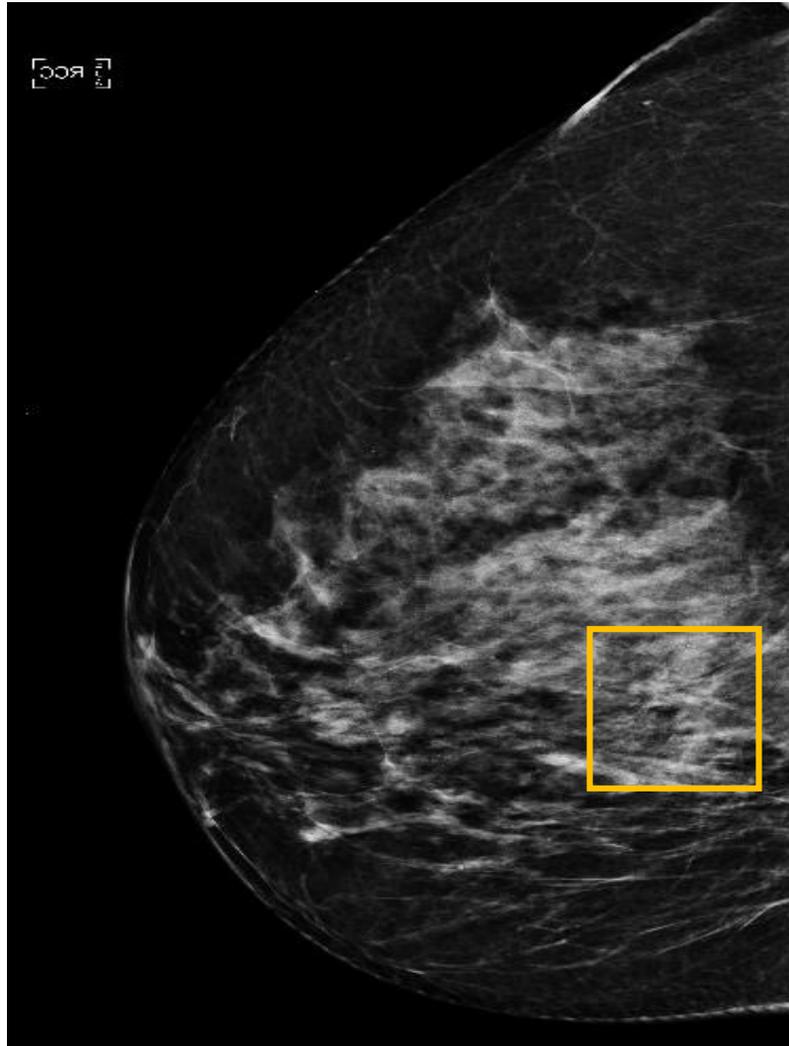
Let's have a look at this case

48yo woman. Asymptomatic. Heterogeneously dense breast (ACR C). Courtesy of Dr. Javier Del Riego – Hospital Universitari Parc Taulí.

Would you TRUST this tool?

48yo woman. Asymptomatic. Heterogeneously dense breast (ACR C).
Courtesy of Dr. Javier Del Riego – Hospital Universitari Parc Taulí.

What about now?

97.6% confidence!

48yo woman. Asymptomatic. Heterogeneously dense breast (ACR C).
Courtesy of Dr. Javier Del Riego – Hospital Universitari Parc Taulí.

How would you make it TRUSTWORTHY?
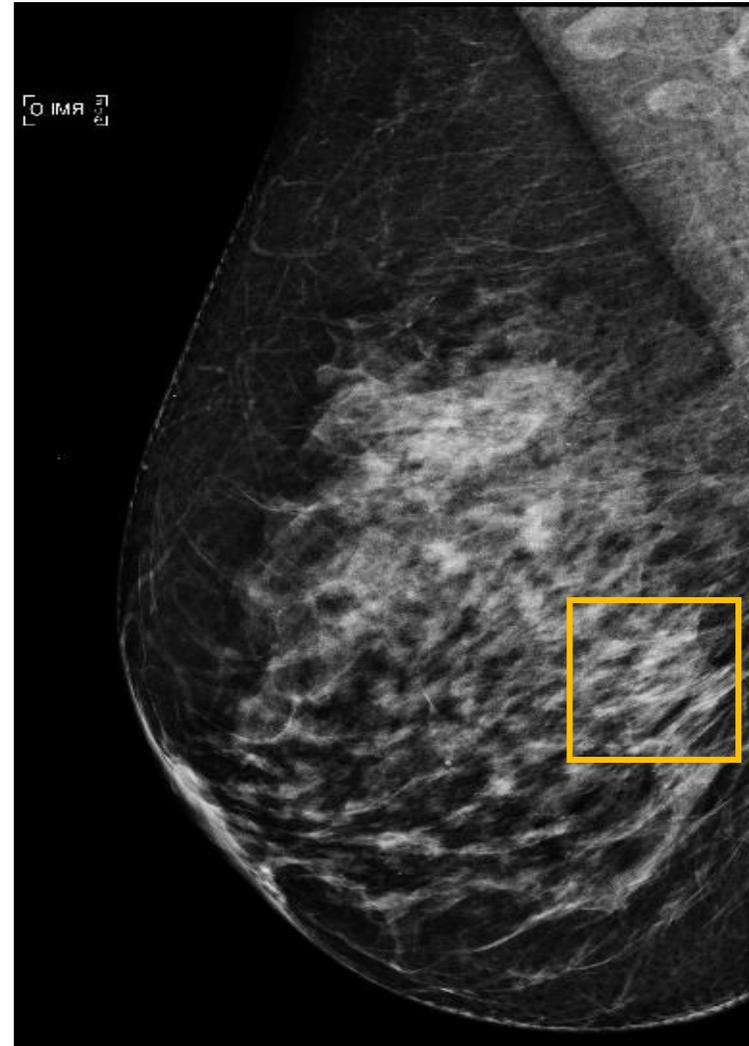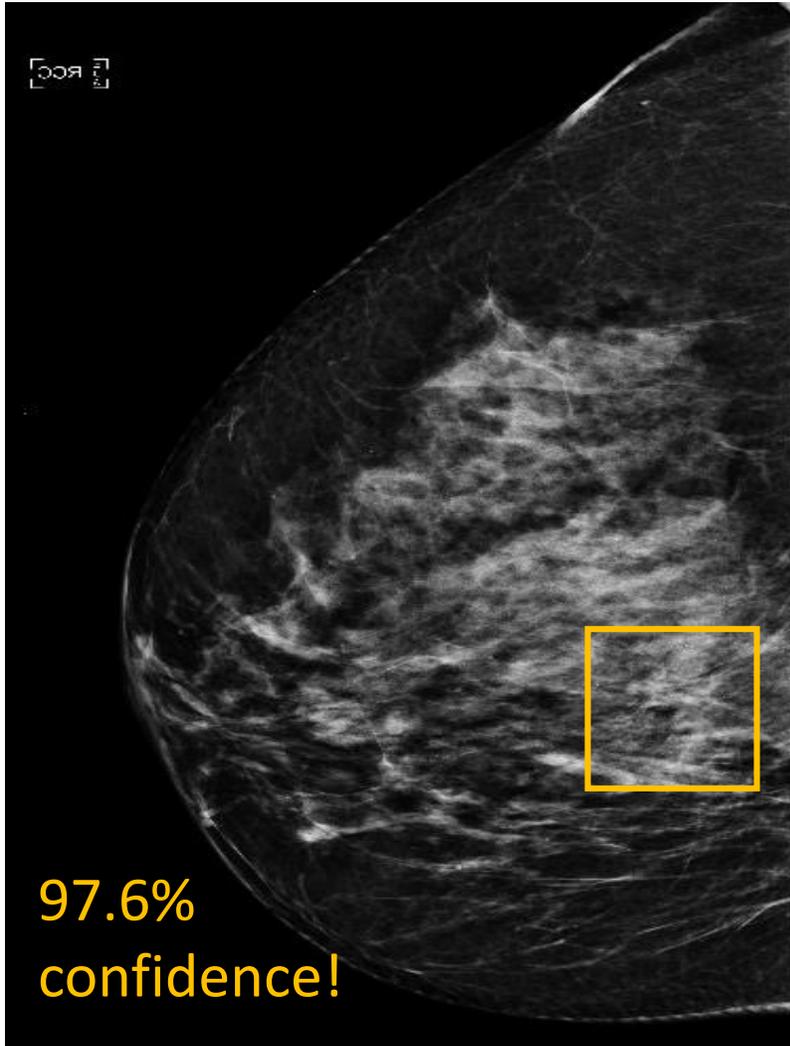
97.6% confidence!
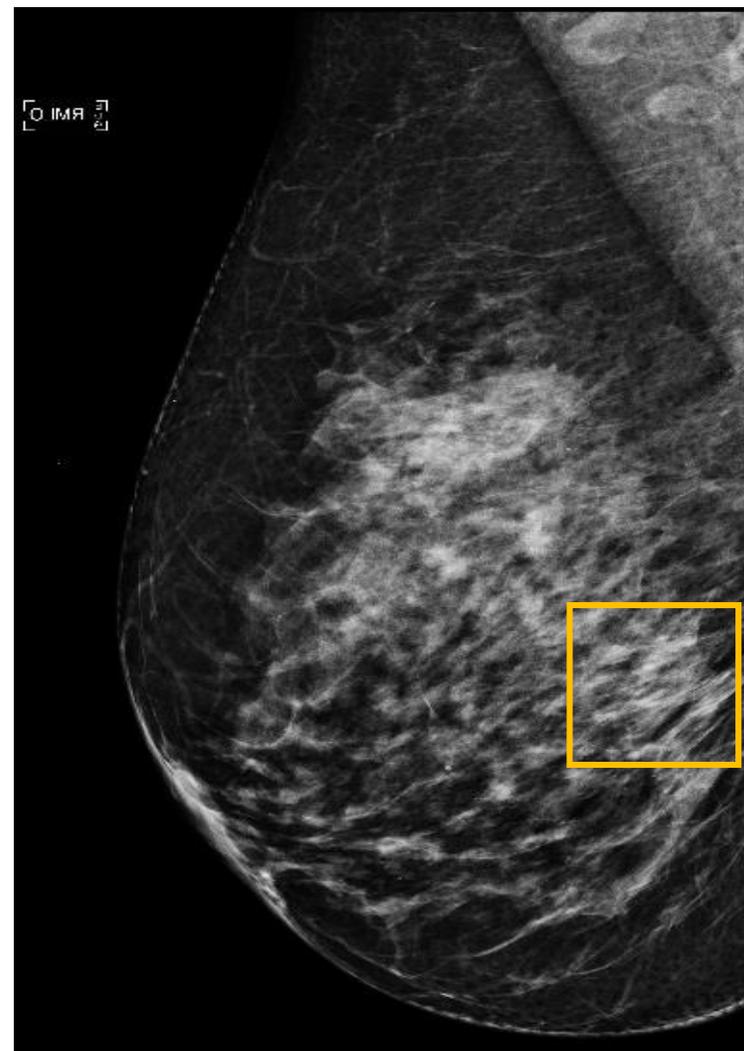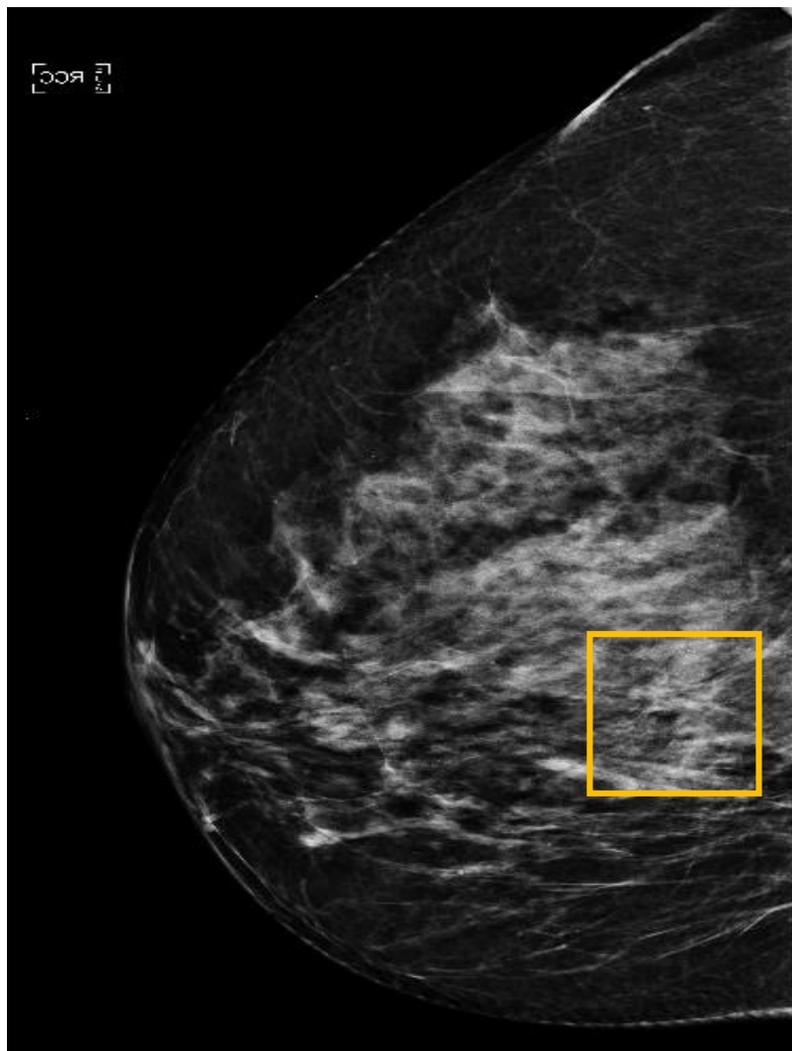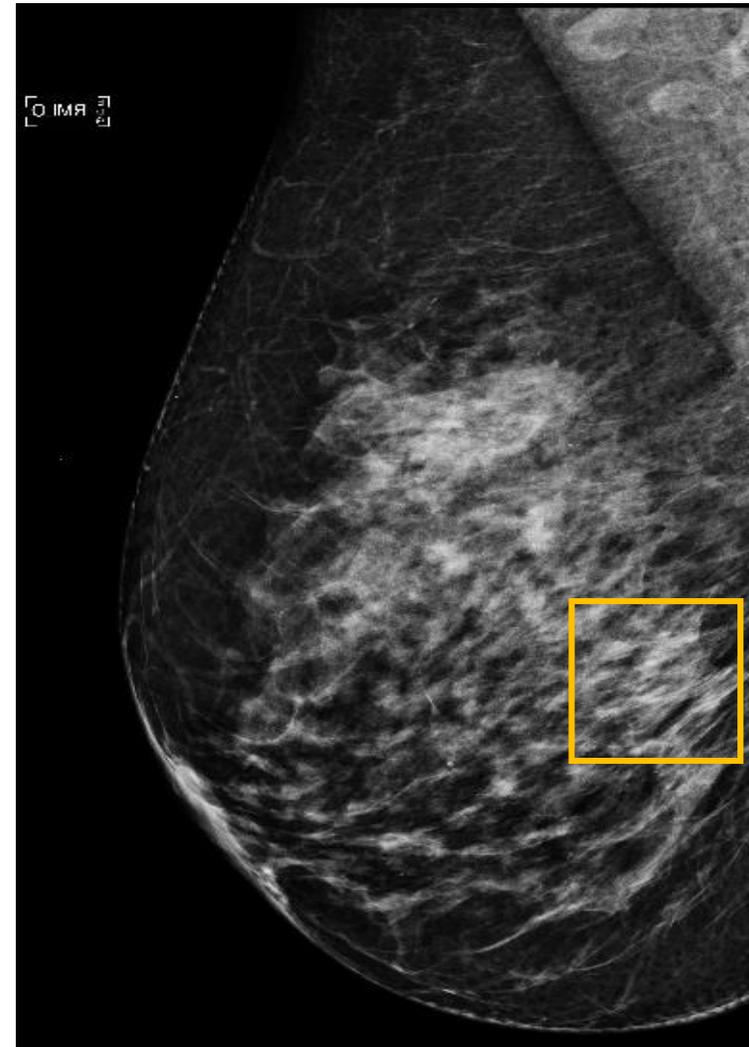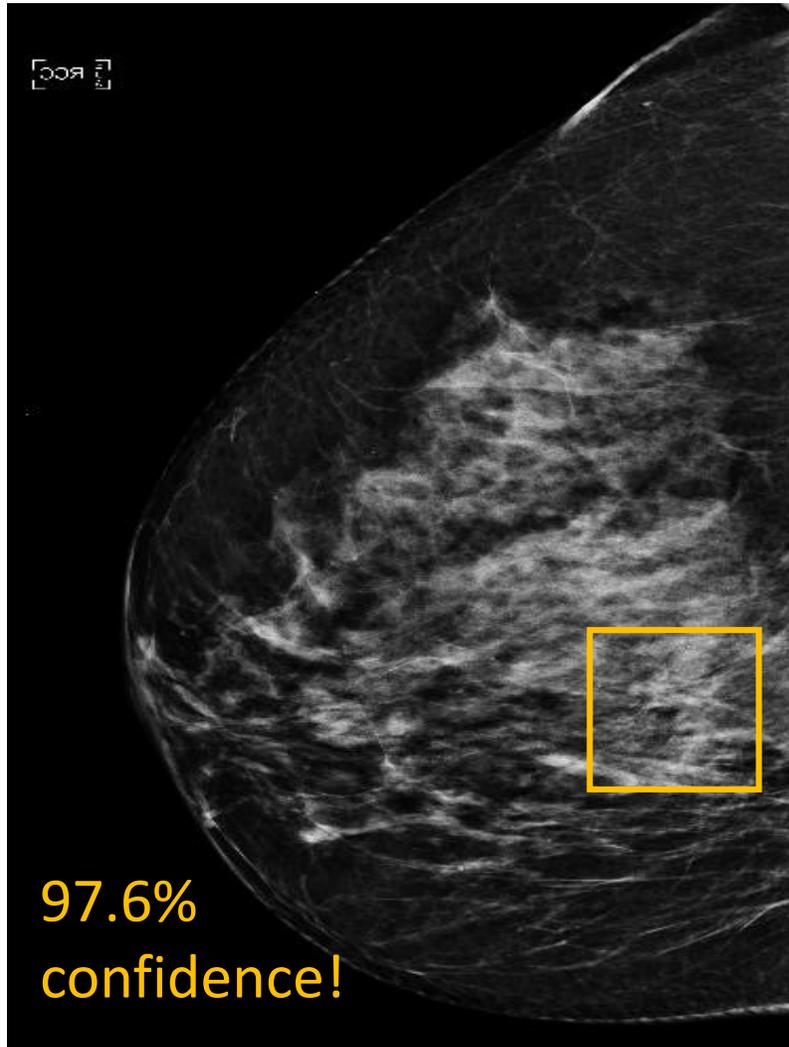
48yo woman. Asymptomatic. Heterogeneously dense breast (ACR C). Courtesy of Dr. Javier Del Riego – Hospital Universitari Parc Taulí.

Why is this a problem?

Keywords: "Artificial Intelligence" AND "Medical Imaging" AND "2000 – 2023"

2019: **4,094** publications

2020: **6.051** publications

2021: **8,076** publications

2022: **9,004** publications

2023 (Jan-Oct): **6,893** publications
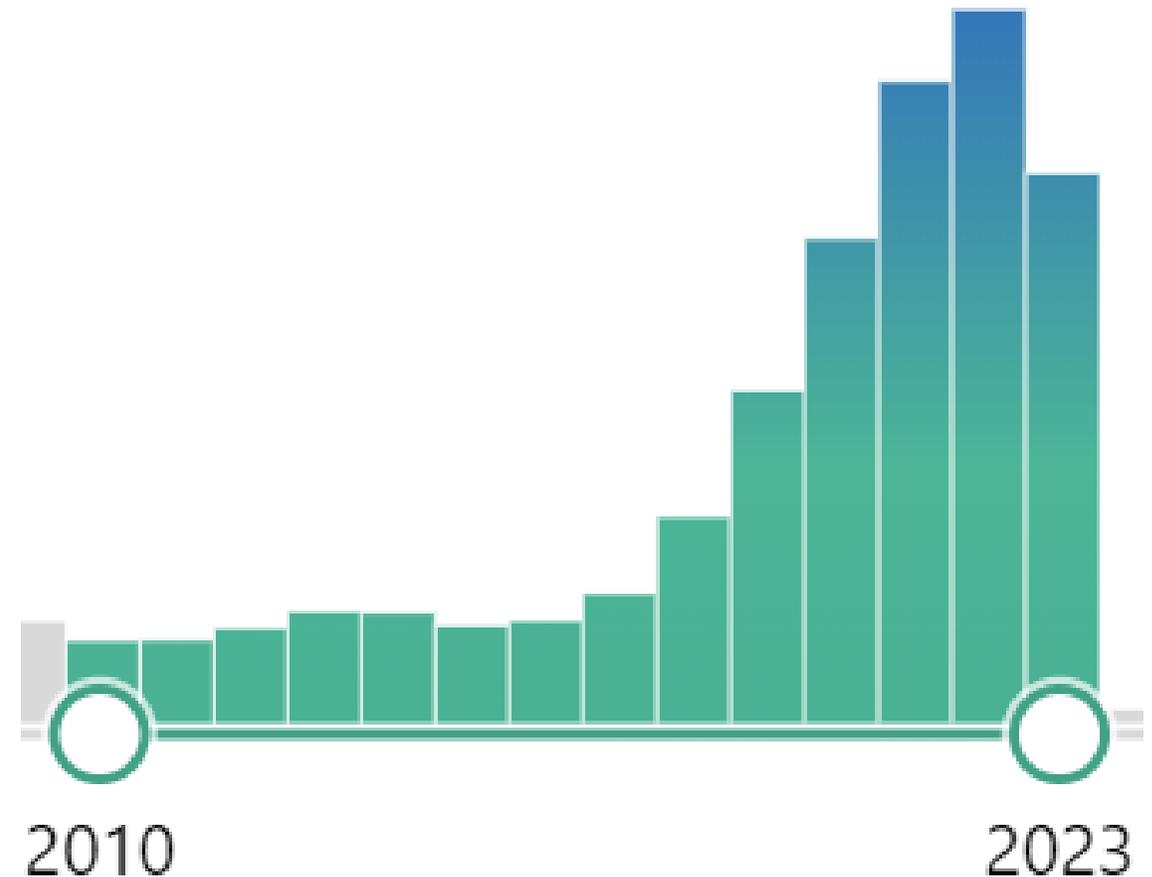
Source: PubMed (2000 – 2023)

FDA-cleared AI radiology tools

https://aicentral.acrdsi.org

IA radiology products with CE mark

AI for Radiology
an implementation guide

Products   Companies   Blogs   About ⌄   Contact

Subscribe to our monthly newsletter   Subscribe

220 IA solutions!

Products

Find the artificial intelligence based software for radiology that you are looking for.
All products listed are available for the European market (CE marked).

Subspeciality:	Modality:	CE:ⓘ	CE class:ⓘ	FDA class:ⓘ	Sort by:
All ⌄	All ⌄	All ⌄	All ⌄	All ⌄	last modified ⌄

Search...   Search

220/220 results

24 AI tools for breast

www.AIforRadiology.com

Is FDA or CE enough to TRUST a tool?

IMAGING INFORMATICS AND ARTIFICIAL INTELLIGENCE

Check for updates

Artificial intelligence in radiology: 100 commercially available products and their scientific evidence

Kicky G. van Leeuwen [1] · Steven Schalekamp [1] · Matthieu J. C. M. Rutten [1,2] · Bram van Ginneken [1] · Maarten de Rooij [1]

- 64% of AI solutions do not have scientific evidence of their effectiveness.

- Only 18% have shown potential impact in clinical practice.

- Wide variety of technology deployment strategies, prices, etc.

Pending issues for trustworthy AI

- Lack of Safety
- Lack of Transparency
- Lack of Accountability
- Lack of Explainability
- Lack of Fairness
- Lack of Usability
- Lack of Security
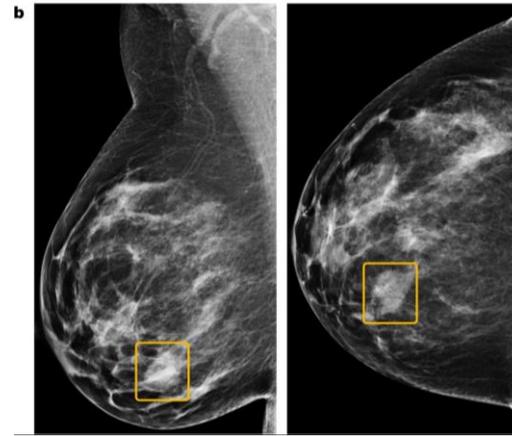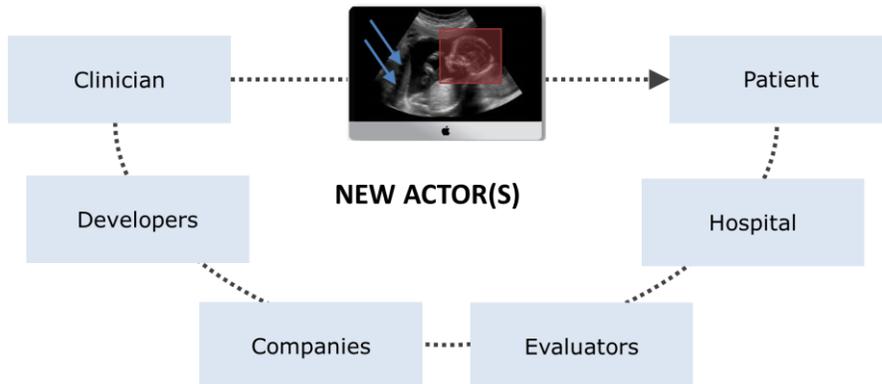
Case missed by the AI model but identified by all 6 readers

McKinney, S.M. et al. 2020. International evaluation of an AI system for breast cancer screening. Nature, 577(7788)

The Machine
Making sense of AI

Google's breast cancer-predicting AI research is useless without transparency, critics say

ARTICLES
https://doi.org/10.1038/s41591-021-01595-0

nature medicine

OPEN

Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations

Laleh Seyyed-Kalantari [1,2], Haoran Zhang[3], Matthew B. A. McDermott[3], Irene Y. Chen[3] and Marzyeh Ghassemi [2,3]

Artificial intelligence (AI) systems have increasingly achieved expert-level performance in medical imaging applications. However, there is growing concern that such AI systems may reflect and amplify human bias, and reduce the quality of their performance in historically under-served populations such as female patients, Black patients, or patients of low socioeconomic status. Such biases are especially troubling in the context of underdiagnosis, whereby the AI algorithm would inaccurately label an individual with a disease as healthy, potentially delaying access to care. Here, we examine algorithmic underdiagnosis in chest X-ray pathology classification across three large chest X-ray datasets, as well as one multi-source dataset. We find that classifiers produced using state-of-the-art computer vision techniques consistently and selectively underdiagnosed under-served patient populations and that the underdiagnosis rate was higher for intersectional under-served subpopulations, for example, Hispanic female patients. Deployment of AI systems using medical imaging for disease diagnosis with such biases risks exacerbation of existing care biases and can potentially lead to unequal access to medical treatment, thereby raising ethical concerns for the use of these models in the clinic.

NEW ACTOR(S)

Clinician → Patient

Developers

Companies — Evaluators

Hospital

Ransomware attack spreads chaos at a major hospital in Barcelona

E.g. Non-optimal evaluation metrics
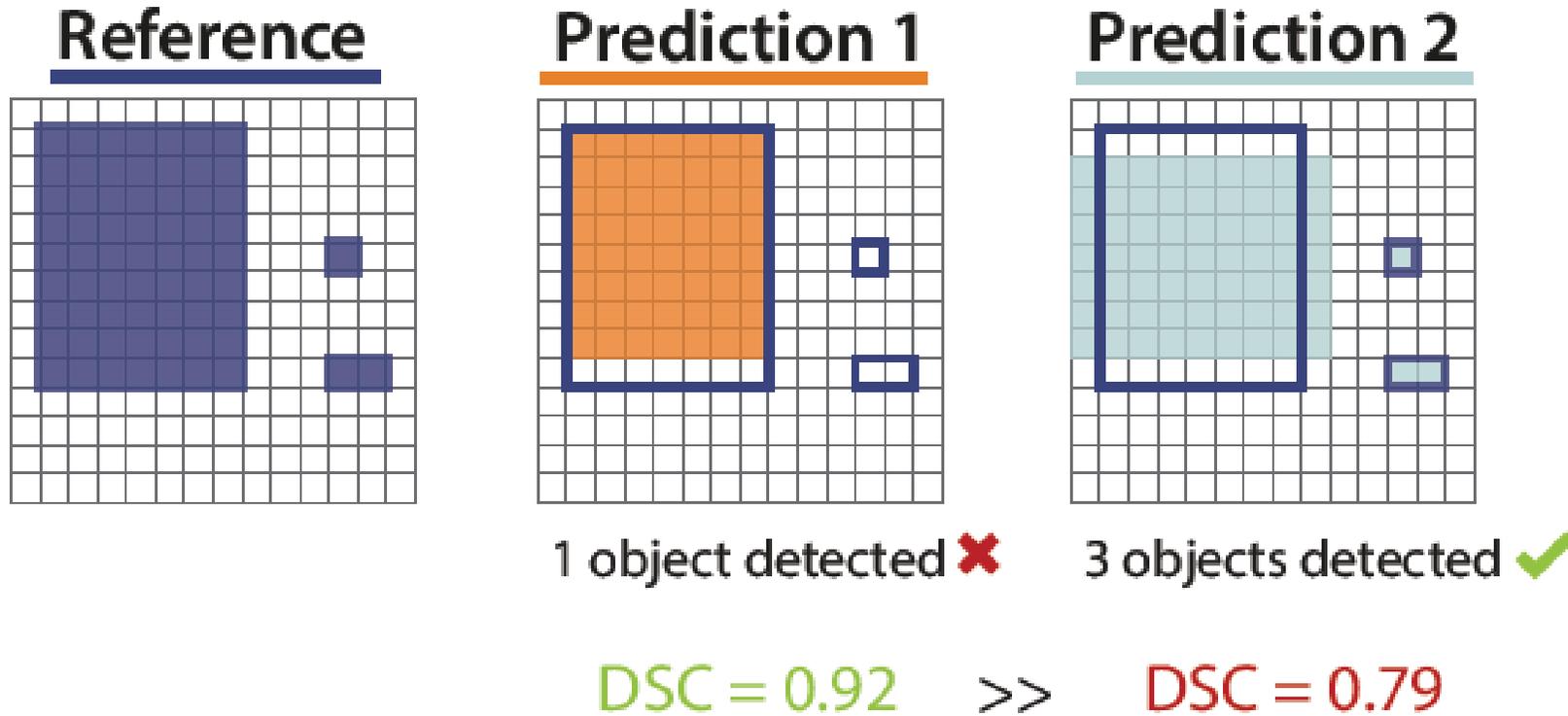
DSC is heavily biased towards a single object and is therefore not appropriate for measuring detection of multiple objects.

Maier-Hein et al. "Metrics reloaded: Pitfalls and recommendations for image analysis validation." arXiv. org 2206.01653 (2022).

Efforts from European Union

(July 2020)

https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342

International Initiatives

Open access — **Protocol**

BMJ Open 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

Gary S Collins [1,2] Paula Dhiman [1,2] Constanza L Andaur Navarro [3] Jie Ma [1] Lotty Hooft [3,4] Johannes B Reitsma [3] Patricia Logullo [1,2] Andrew L Beam [5,6] Lily Peng [7] Ben Van Calster [8,9,10] Maarten van Smeden [3] Richard D Riley [11] Karel GM Moons [3,4]

Kocak et al. Insights into Imaging (2023) 14:75
https://doi.org/10.1186/s13244-023-01415-8

Insights into Imaging
ESR EUROPEAN SOCIETY OF RADIOLOGY

GUIDELINE — **Open Access**

CheckList for EvaluAtion of Radiomics research (CLEAR): a step-by-step reporting guideline for authors and reviewers endorsed by ESR and EuSoMII

Burak Kocak[1*] Bettina Baessler[2] Spyridon Bakas[3,4,5] Renato Cuocolo[6] Andrey Fedorov[7] Lena Maier-Hein[8,9] Nathaniel Mercaldo[10,11] Henning Müller[12,13] Fanny Orlhac[14] Daniel Pinto dos Santos[15,16] Arnaldo Stanzione[17] Lorenzo Ugga[17] and Alex Zwanenburg[18,19,20]

Metrics Reloaded: Recommendations for image analysis validation

LENA MAIER-HEIN[*†], German Cancer Research Center (DKFZ), Germany, Heidelberg University, Germany, and National Center for Tumor Diseases (NCT), Germany
ANNIKA REINKE[†], German Cancer Research Center (DKFZ), Germany and Heidelberg University, Germany

Radiology: Artificial Intelligence — **EDITORIAL**

Checklist for Artificial Intelligence in Medical Imaging (CLAIM): A Guide for Authors and Reviewers

John Mongan, MD, PhD • Linda Moy, MD • Charles E. Kahn, Jr, MD, MS

AAPM SCIENTIFIC REPORT — **MEDICAL PHYSICS**

AAPM task group report 273: Recommendations on best practices for AI and machine learning for computer-aided diagnosis in medical imaging

Lubomir Hadjiiski[1] | Kenny Cha[2] | Heang-Ping Chan[3] | Karen Drukker[4] | Lia Morra[5] | Janne J. Näppi[6] | Berkman Sahiner[7] | Hiroyuki Yoshida[8] | Quan Chen[9] | Thomas M. Deserno[10] | Hayit Greenspan[11] | Henkjan Huisman[12] | Zhimin Huo[13] | Richard Mazurchuk[14] | Nicholas Petrick[15] | Daniele Regge[16,17] | Ravi Samala[18] | Ronald M. Summers[19] | Kenji Suzuki[20] | Georgia Tourassi[21] | Daniel Vergara[22] | Samuel G. Armato III[23]

FUTURE-AI: Guiding Principles and Consensus Recommendations for Trustworthy Artificial Intelligence in Medical Imaging

Karim Lekadir[a,*], Richard Osuala[a], Catherine Gallin[a], Noussair Lazrak[a], Kaisar Kushibar[a], Gianna Tsakou[b], Susanna Aussó[e], Leonor Cerdá Alberich[d], Kostas Marias[c], Manolis Tsiknakis[c], Sara Colantonio[g], Nickolas Papanikolaou[h], Zohaib Salahuddin[f], Henry C Woodruff[f], Philippe Lambin[f], Luis Martí-Bonmatí[d]

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, Universality, Traceability, Usability, Robustness and Explainability.

www.future-ai.eu

FUTURE-AI

International consensus

FUTURE-AI

Fair	• Increase fairness across groups • Increase fairness across subjects
Universal	• Increase interoperability • Increase transferability
Traceable	• Facilitate accountability • Identify data or concept drifts
Usable	• Increase clinical usability • Increases clinical adoption
Robust	• Robust to heterogeneous conditions • Robust to security threats
Explainable	• Increase transparency • Increase acceptance

FUTURE-AI

		Recommendations	Low ML-TRL	High ML-TRL
Fairness	1	Define any potential sources of bias from an early stage	++	++
	2	Collect data on individuals' attributes	+	+
	3	Evaluate potential biases and bias correction measures	+	++

Recommendation	Description
Fairness 1 Define sources of bias	Bias in medical AI is application-specific[1]. At the design phase, the development team should identify possible types and sources of bias for their AI tool[2]. 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 labelling, data curation, or the selection of the input features.

FUTURE-AI

	Criteria	Example
Current models	Performance Metric	• 0.85 AUC • 0.80 F1-Score

Conclusions

- Despite the **hype**, IA still not widely used clinically

- Many groups focused on **trust** rather than **precision**

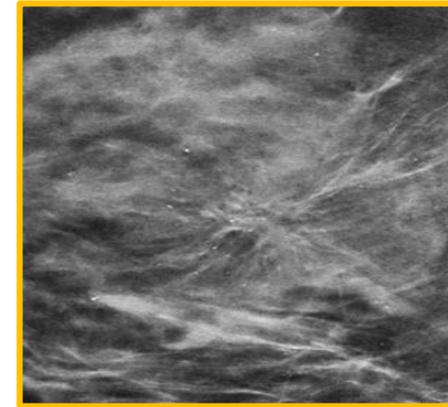- We need to **work collaboratively** to develop, validate and deploy **trustworthy AI tools**
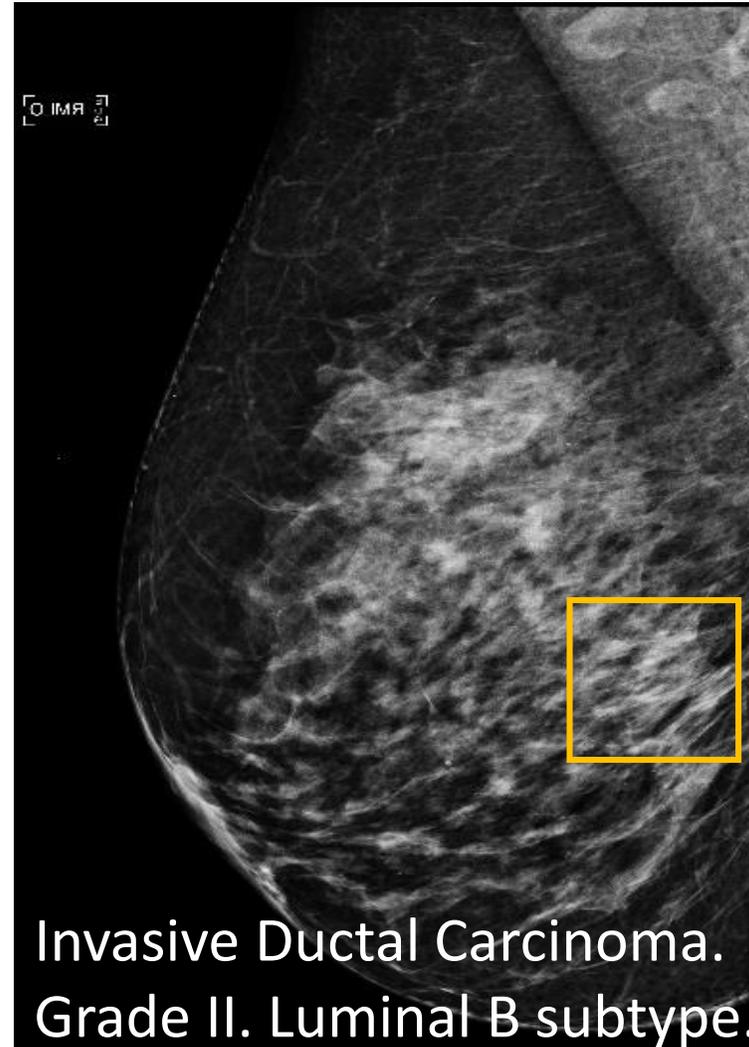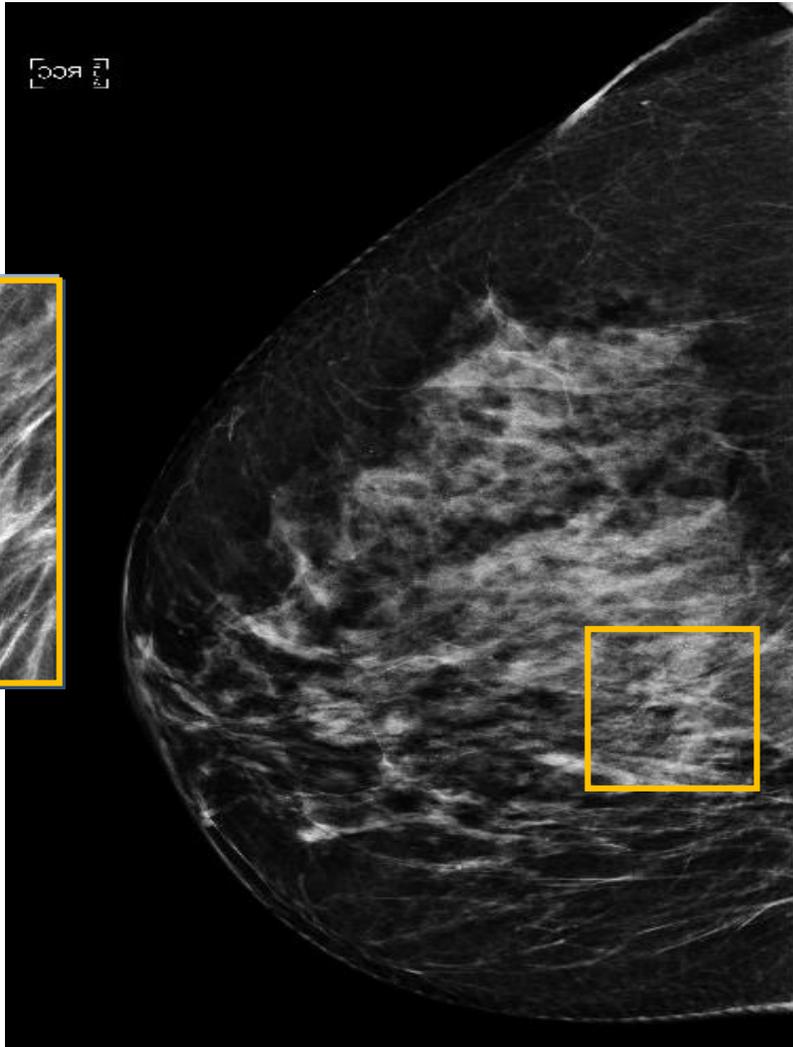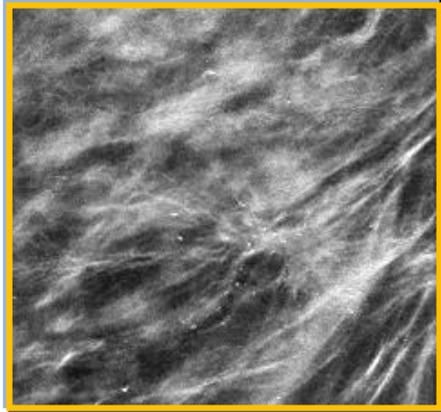
Thank you!

Oliver Díaz

University of Barcelona

Artificial Intelligence in Medicine Lab (BCM-AIM)

oliver.diaz@ub.edu

Still curious about this case?

Invasive Ductal Carcinoma.
Grade II. Luminal B subtype.

48yo woman. Asymptomatic. Heterogeneously dense breast (ACR C).
Courtesy of Dr. Javier Del Riego – Hospital Universitari Parc Taulí.