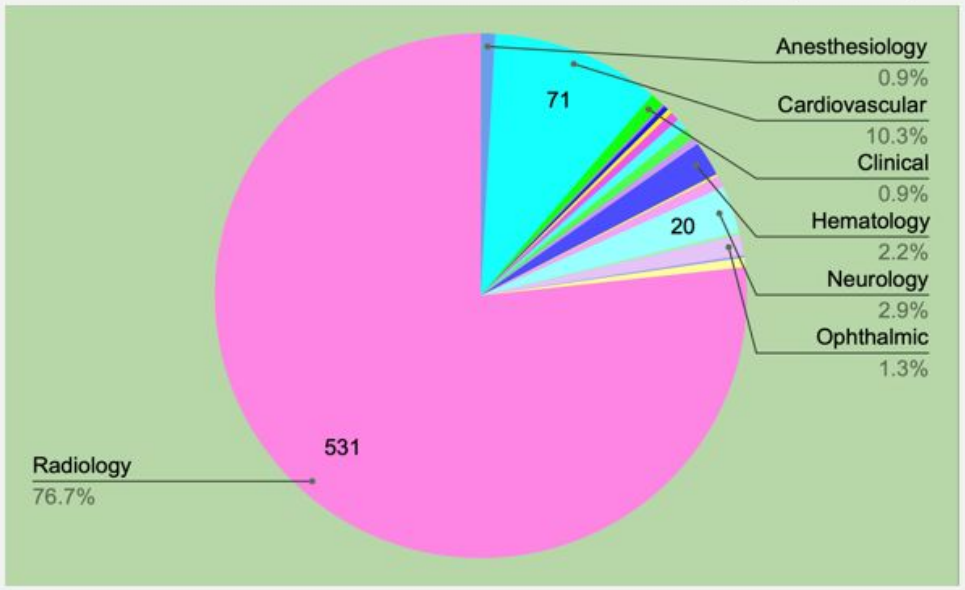
Implementation of AI in clinical practice

Juan Enrique Gutiérrez, MD
Bayer, EMEA Medical Affairs Digital Radiology Lead

MA-PFM-RAD-SK-0004-1; 3/2024

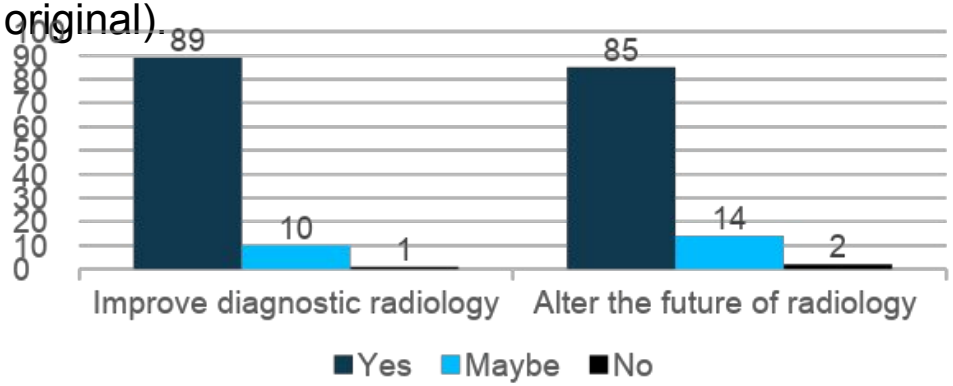
Available AI technologies and expectations



Based on FDA 2023 List Authorized AI-Enabled Medical Devices
Image source Margaretta Colangelo Copyright © 2023

https://www.linkedin.com/pulse/fda-publishes-new-list-692-authorized-ai-enabled-margaretta-colangelo/?trackingId=Dr%2FZGoDqR5Suy01PFLdjLA%3D%3D [Accessed 09-11-2023]

Expected impact of AI. Answered from 1041 respondents from 54 countries (adapted from original).



How can AI help diagnostic radiology? (n = 1029)[a]

Second reader	**829 (78%)**
Workflow optimization	**803 (77%)**
Partial replacement	**493 (47%)**
Full replacement	**11 (1%)**

[a]: multiple answers possible

Huisman M, Ranschaert E, Parker W, Mastrodicasa D, Koci M, Pinto de Santos D, Coppola F, Morozov S, Zins M, Bohyn C, Koç U, Wu J, Veean S, Fleischmann D, Leiner T, Willemink MJ. An international survey on AI in radiology in 1041 radiologists and radiology residents part 2: expectations, hurdles to implementation, and education. Eur Radiol. 2021 Nov;31(11):8797-8806.

AI deployment

ORIGINAL ARTICLE ■ *Data Science*

Check for updates

2020 ACR Data Science Institute Artificial Intelligence Survey

Editor's Choice

Bibb Allen, MD[a], Sheela Agarwal, MD[b], Laura Coombs, PhD[c], Christoph Wald, MD[d], Keith Dreyer, DO, PhD[e]

1427 radiologists

493 (~ 34%) using AI in their practice.

934 (~ 66%) not using AI in their practice.

Bibb Allen, Sheela Agarwal, Laura Coombs. 2020 ACR Data Science Institute Artificial Intelligence Survey, Journal of the American College of Radiology, Volume 18, Issue 8, 2021, Pages 1153-1159,.

STATEMENT Open Access

Current practical experience with artificial intelligence in clinical radiology: a survey of the European Society of Radiology

Check for updates

690 radiologists (44 countries)

276 (~ 40%) practical clinical experience with AI.

412 (~ 60%) no clinical experience with AI.

European Society of Radiology (ESR). Current practical experience with artificial intelligence in clinical radiology: a survey of the European Society of Radiology. Insights Imaging 13, 107 (2022).

AI adoption barriers

Pre-implementation

- Data quality/annotation
- Lack of evidence
- Benchmarking
- Economics

Implementation

- Scalability
- Workflow disruption
- Interoperability

Post-Implementation

- Monitoring
- Maintenance

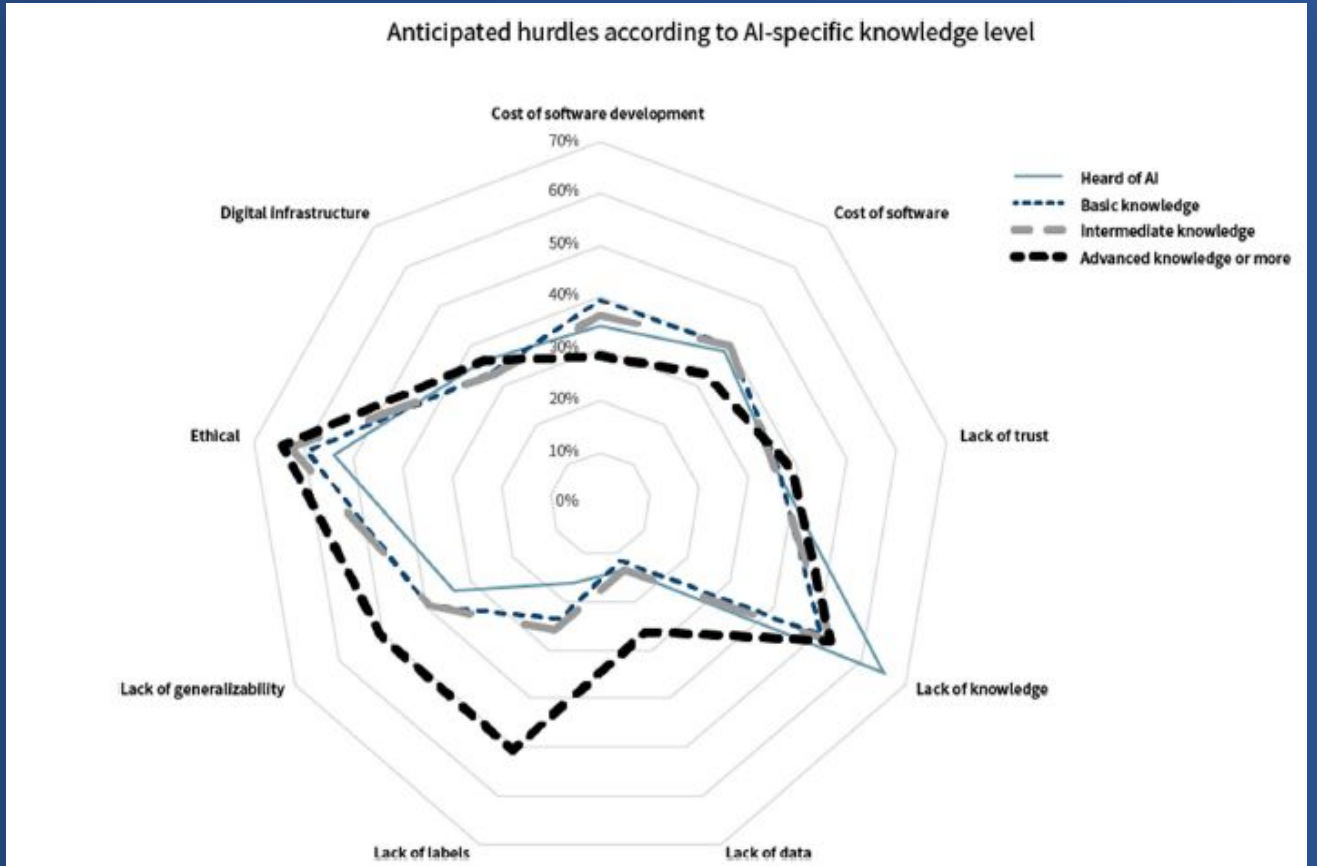


Fig. 2 Anticipated hurdles as indicated by respondents according to AI-specific knowledge level

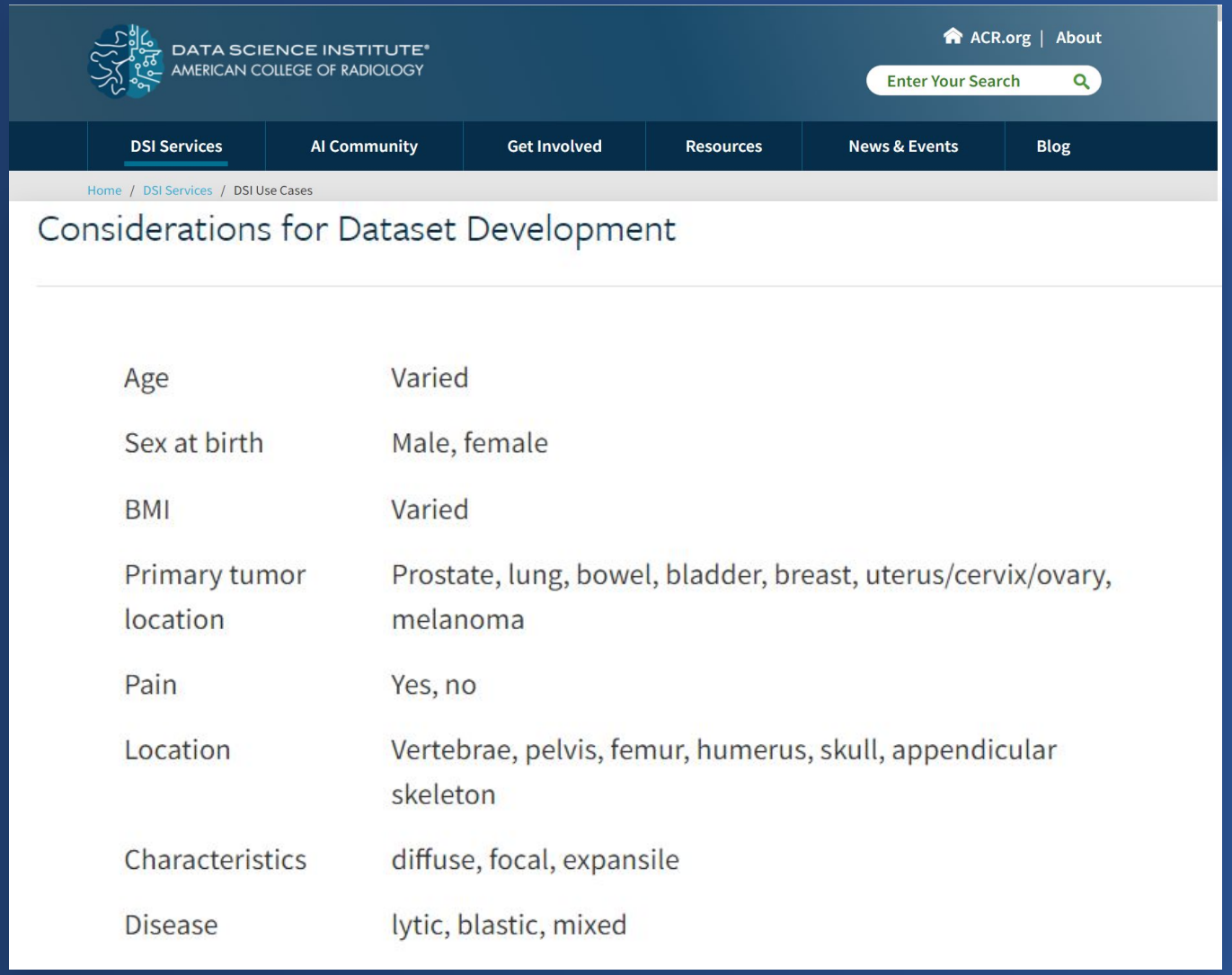
Huisman M, Ranschaert E, Parker W, Mastrodicasa D, Koci M, Pinto de Santos D, Coppola F, Morozov S, Zins M, Bohyn C, Koç U, Wu J, Veean S, Fleischmann D, Leiner T, Willemink MJ. An international survey on AI in radiology in 1041 radiologists and radiology residents part 2: expectations, hurdles to implementation, and education. Eur Radiol. 2021 Nov;31(11):8797-8806.

Clinical need

AI use case

ACR DSI use cases:

- Overview
- Clinical implementation
- Dataset considerations
- Technical specifications
- Future developments



https://www.acrdsi.org/DSI-Services/Define-AI/Use-Cases/Bony-Metastatic-Disease-Detection. [Accessed: 27-10-2023]

AI governance framework

Imaging AI governing body

Interdisciplinary team

Objectives:

- Guidelines for data access and system integrations.
- Select AI tools based on their clinical value and impact on patient care.
- Assess the safety and effectiveness of AI algorithms.

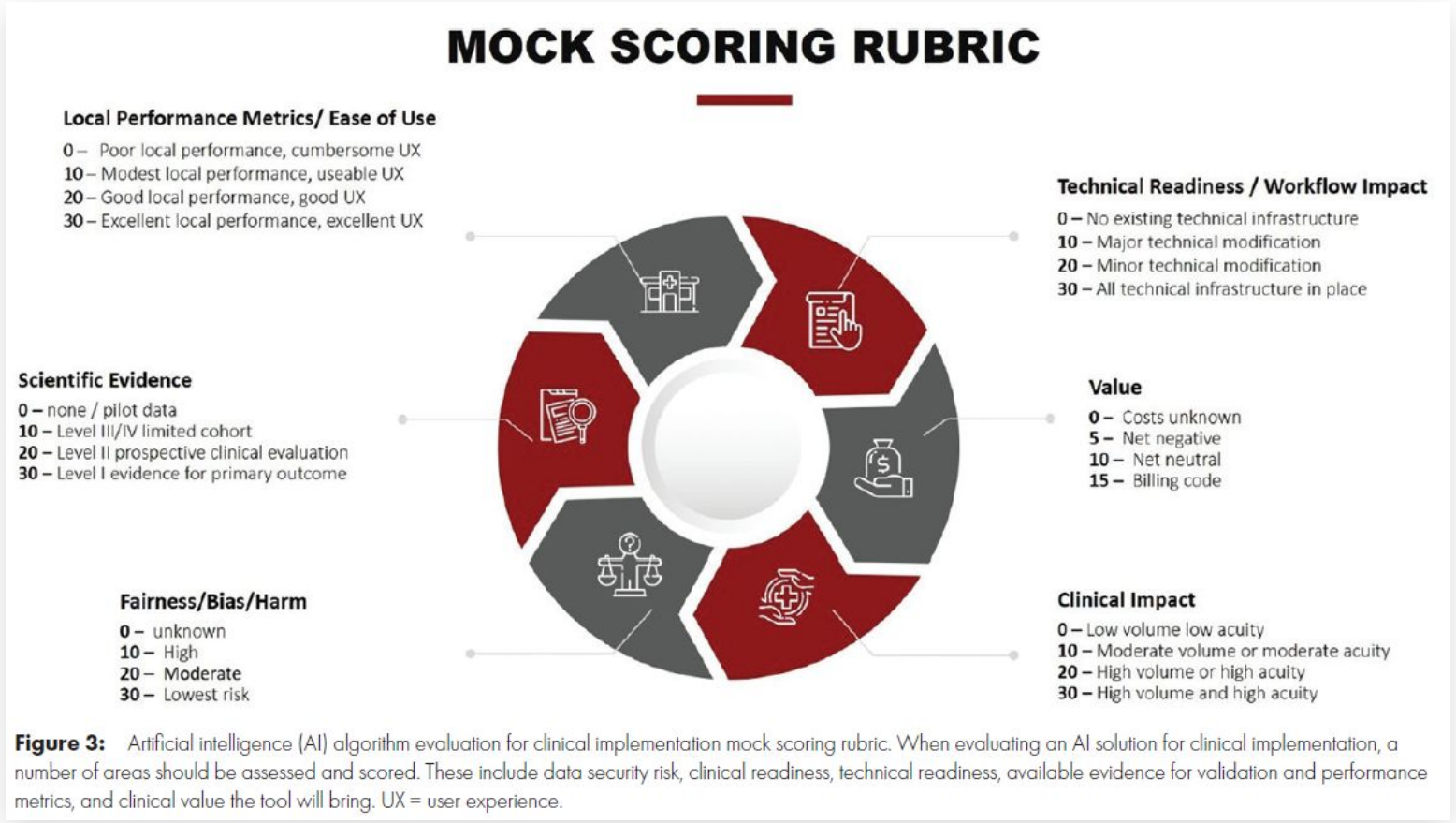


Figure 3: Artificial intelligence (AI) algorithm evaluation for clinical implementation mock scoring rubric. When evaluating an AI solution for clinical implementation, a number of areas should be assessed and scored. These include data security risk, clinical readiness, technical readiness, available evidence for validation and performance metrics, and clinical value the tool will bring. UX = user experience.

Daye D, Wiggins WF, Lungren MP, Alkasab T, Kottler N, Allen B, Roth CJ, Bizzo BC, Durniak K, Brink JA, Larson DB, Dreyer KJ, Langlotz CP. Implementation of Clinical Artificial Intelligence in Radiology: Who Decides and How? Radiology. 2022 Dec;305(3):555-563.

AI governance framework

AI evaluation for clinical implementation

- Data security risks
- Local performance metrics
- Scientific evidence

Table 1 | Overview of existing and upcoming AI reporting guidelines

AI reporting guidelines

Name	Stage	Study design	Comment
TRIPOD-AI	Preclinical development	**Prediction model evaluation**	Extension of TRIPOD. Used to report prediction models (diagnostic or prognostic) development, validation and updates. Focuses on model performance.
STARD-AI	Preclinical development and offline validation	**Diagnostic accuracy studies**	Extension of STARD. Used to report diagnostic accuracy studies, either at development stage or as an offline validation in clinical settings. Focuses on diagnostic accuracy.
DECIDE-AI	**Early live clinical evaluation**	Various (prospective cohort studies, non-randomized controlled trials, …)[a] with additional features, such as modification of intervention, analysis of pre-specified subgroups or learning curve analysis.	Stand-alone guideline. Used to report the early evaluation of AI systems as an intervention in live clinical settings (small-scale, formative evaluation), independently of the study design and AI system modality (diagnostic, prognostic, therapeutic). Focuses on clinical utility, safety and human factors.
SPIRIT-AI	Comparative prospective evaluation	**Randomized controlled trials (protocol)**	Extension of SPIRIT. Used to report the protocols of randomized controlled trials evaluating AI systems as interventions.
CONSORT-AI	Comparative prospective evaluation	**Randomized controlled trials**	Extension of CONSORT. Used to report randomized controlled trials evaluating AI systems as interventions (large-scale, summative evaluation), independently of the AI system modality (diagnostic, prognostic, therapeutic). Focuses on effectiveness and safety.

Bold font indicates the primary target of the guidelines, either a specific stage or a specific study design. [a]Although existing reporting guidelines exist for some of these study designs (for example, STROBE for cohort studies), none covers all the core aspects of AI system early-stage evaluation, and none would fit all possible study designs; DECIDE-AI was, therefore, developed as a new stand-alone reporting guideline for these early, live, clinical AI studies.

Vasey, B., Nagendran, M., Campbell, B. et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. Nat Med 28, 924–933 (2022).

Maintenance and monitoring

Decrease in model performance may occur over time after real-world implementation.

Parameters to monitor:

- Sensitivity, Specificity, PPV.
- Examination metadata.
- Machine metadata.

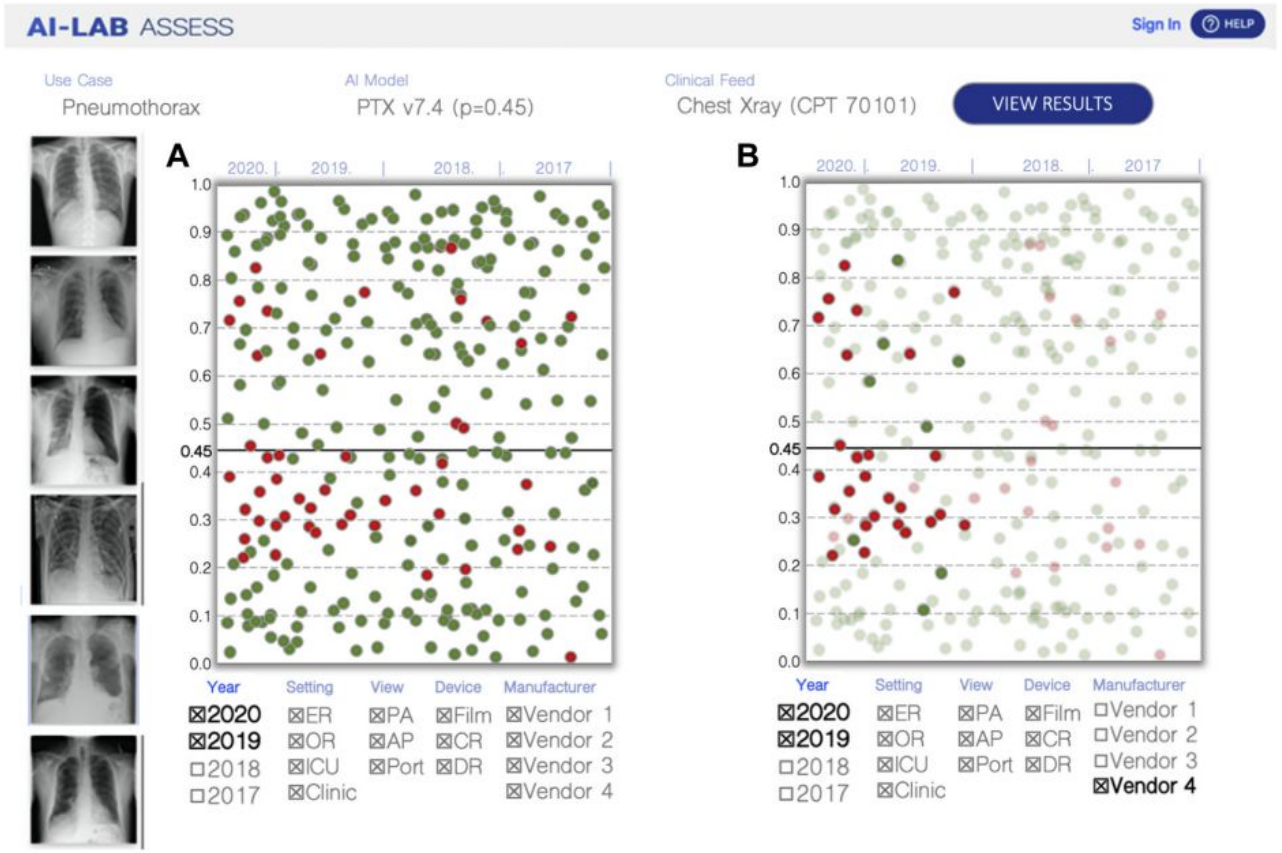
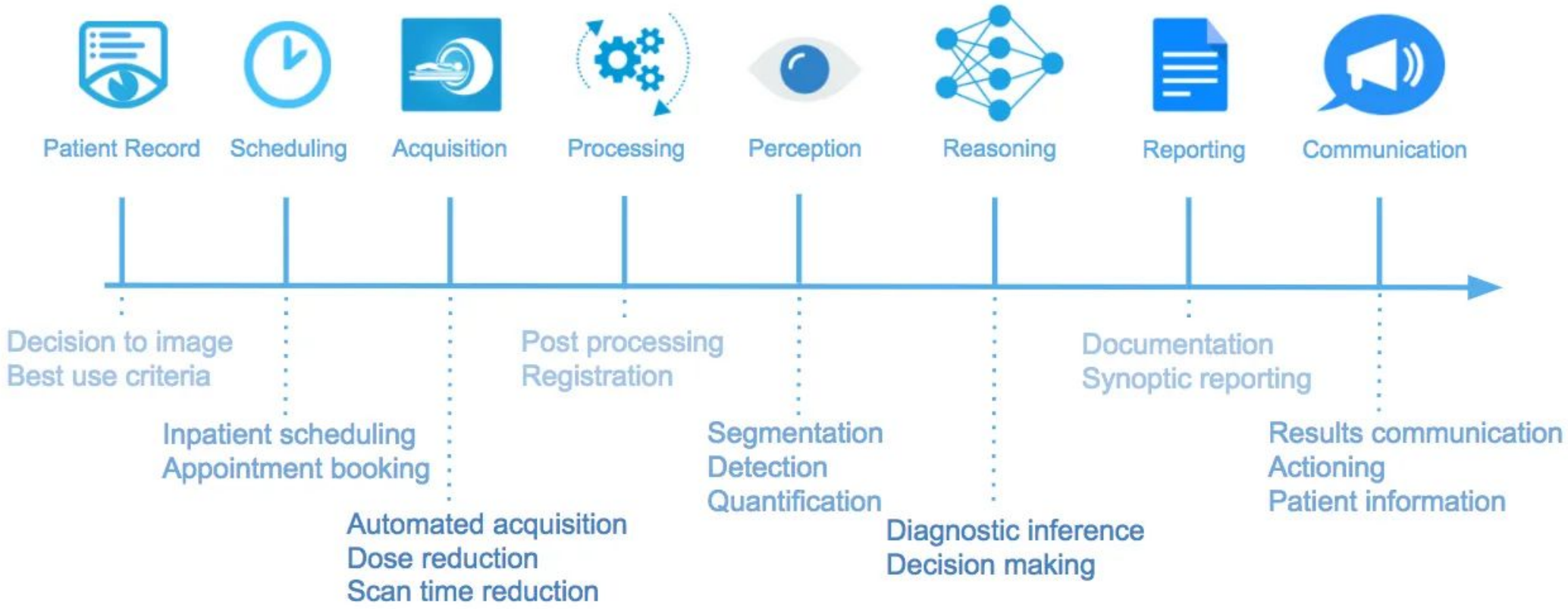


Fig 5. (A) Real-world performance monitoring allows identification of diminished model performance. (B) By isolating machine specific metadata, the failure was attributed to a single machine not a global failure of the model [42]. AP = anteroposterior; CTP = Current Procedural Terminology; ER = emergency room; ICU = intensive care unit; OR = operating room; PA = posteroanterior; PTX = pneumothorax.

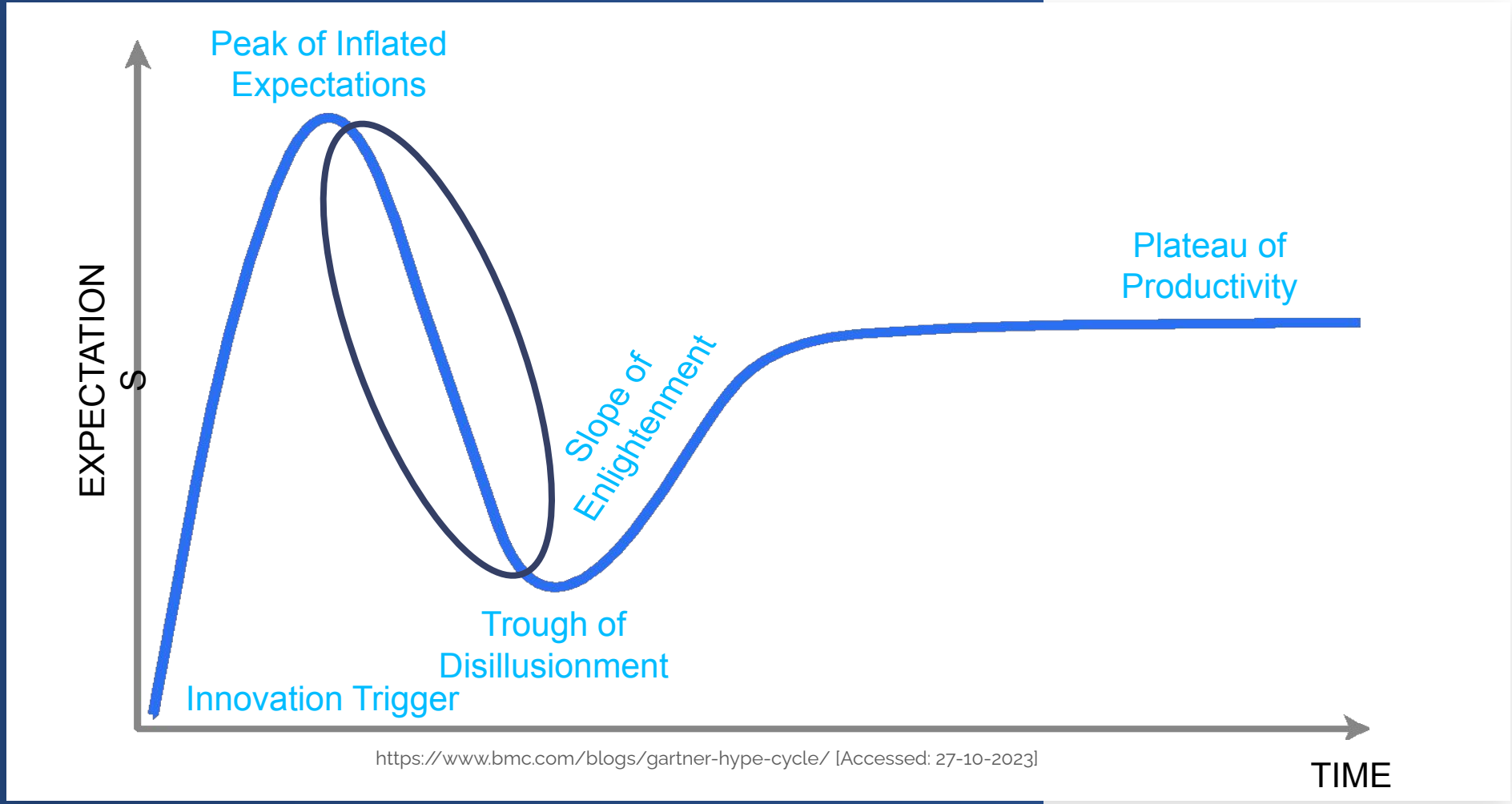
Allen B, Dreyer K, Stibolt R Jr, Agarwal S, Coombs L, Treml C, Elkholy M, Brink L, Wald C. Evaluation and Real-World Performance Monitoring of Artificial Intelligence Models in Clinical Practice: Try It, Buy It, Check It. J Am Coll Radiol. 2021 Nov;18(11):1489-1496.

Workflow integration



Authorization for use of images @drhughharvey

Gartner Hype Cycle



https://www.bmc.com/blogs/gartner-hype-cycle/ [Accessed: 27-10-2023]

Promise of AI

Articles

Artificial intelligence-supported screen reading versus standard double reading in the Mammography Screening with Artificial Intelligence trial (MASAI): a clinical safety analysis of a randomised, controlled, non-inferiority, single-blinded, screening accuracy study

Kristina Lång, Viktoria Josefsson, Anna-Maria Larsson, Stefan Larsson, Charlotte Högberg, Hanna Sartor, Solveig Hofvind, Ingvar Andersson, Aldana Rosso

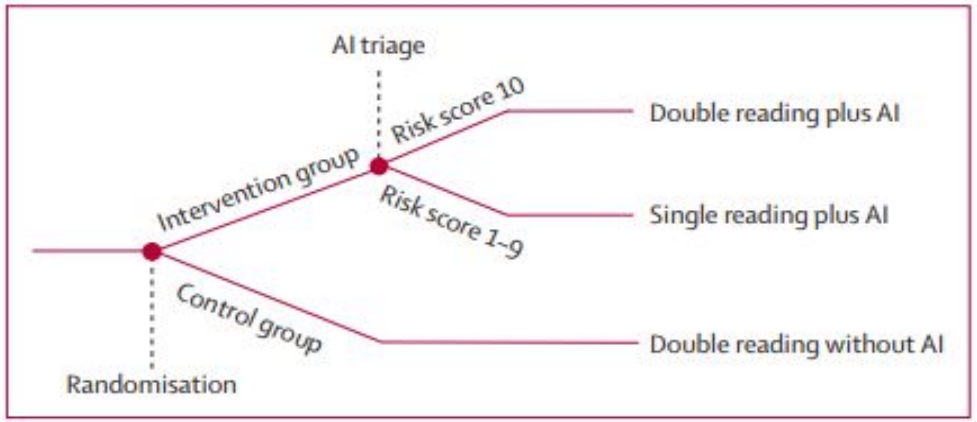


Figure 1: **Overview of trial intervention**
AI=artificial intelligence.

	Intervention group (n=39 996)	Control group (n=40 024)
Early screening performance		
Number of recalls	861	817
Recall rate, %	2·2% (2·0–2·3)	2·0% (1·9–2·2)
Number of screen-detected cancers	244	203
Cancer-detection rate, per 1000 participants screened	6·1 (5·4–6·9)	5·1 (4·4–5·8)
False positive rate, %	1·5% (1·4–1·7)	1·5% (1·4–1·7)
Positive predictive value of recall, %	28·3% (25·3–31·5)	24·8% (21·9–28·0)
Workload		
Number of screen readings	46 345	83 231
Number of consensus meetings	1584	1576
Consensus meeting rate	4·0% (3·8–4·2)	3·9% (3·8–4·1)

Data are n or point estimate (95% CI).

Table 2: **Early screening performance and workload measures, modified intention-to-treat population**

Lång K, Josefsson V, Larsson AM, Larsson S, Högberg C, Sartor H, Hofvind S, Andersson I, Rosso A. Artificial intelligence-supported screen reading versus standard double reading in the Mammography Screening with Artificial Intelligence trial (MASAI): a clinical safety analysis of a randomised, controlled, non-inferiority, single-blinded, screening accuracy study. Lancet Oncol. 2023 Aug;24(8):936-944.

Thank you!

Questions?

juan.gutierrezalliende@bayer.com