Clinical trials in Diagnostic AI

A/Prof Tan Cher Heng

MBBS, FRCR, MBA

Group Chief, Research and Innovation, NHG

Assistant Dean, Lee Kong Chian School of Medicine

Senior Consultant, Diagnostic Radiology, TTSH

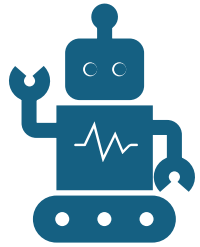
Agenda

Landscape for AI-based diagnostics:

What exists today?

Clinical adoption of medical AI:

What's holding us back?



Clinical trials in AI-based diagnostics:

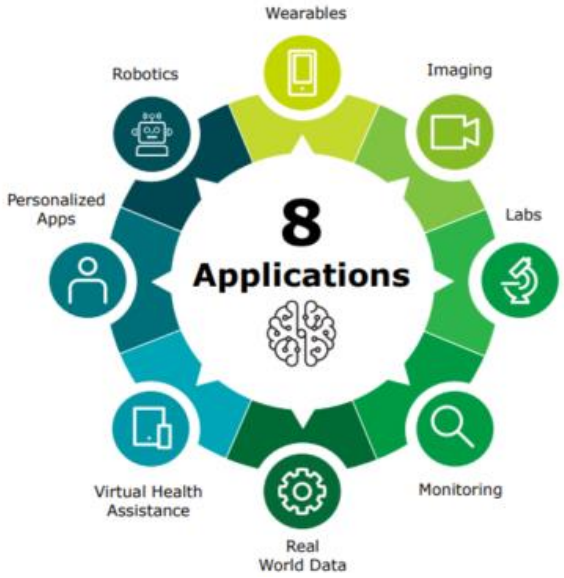
What do we need to consider?

AI is expected to have a major impact on healthcare delivery.

Deloitte.

MedTech Europe
from diagnosis to cure

AI can have a significant socio-economic impact on European health systems



400,000 lives saved yearly

That's the population of a medium-sized city, or **almost two thirds of Luxembourg**[1]

200 billion Euros in annual savings

(including opportunity costs)

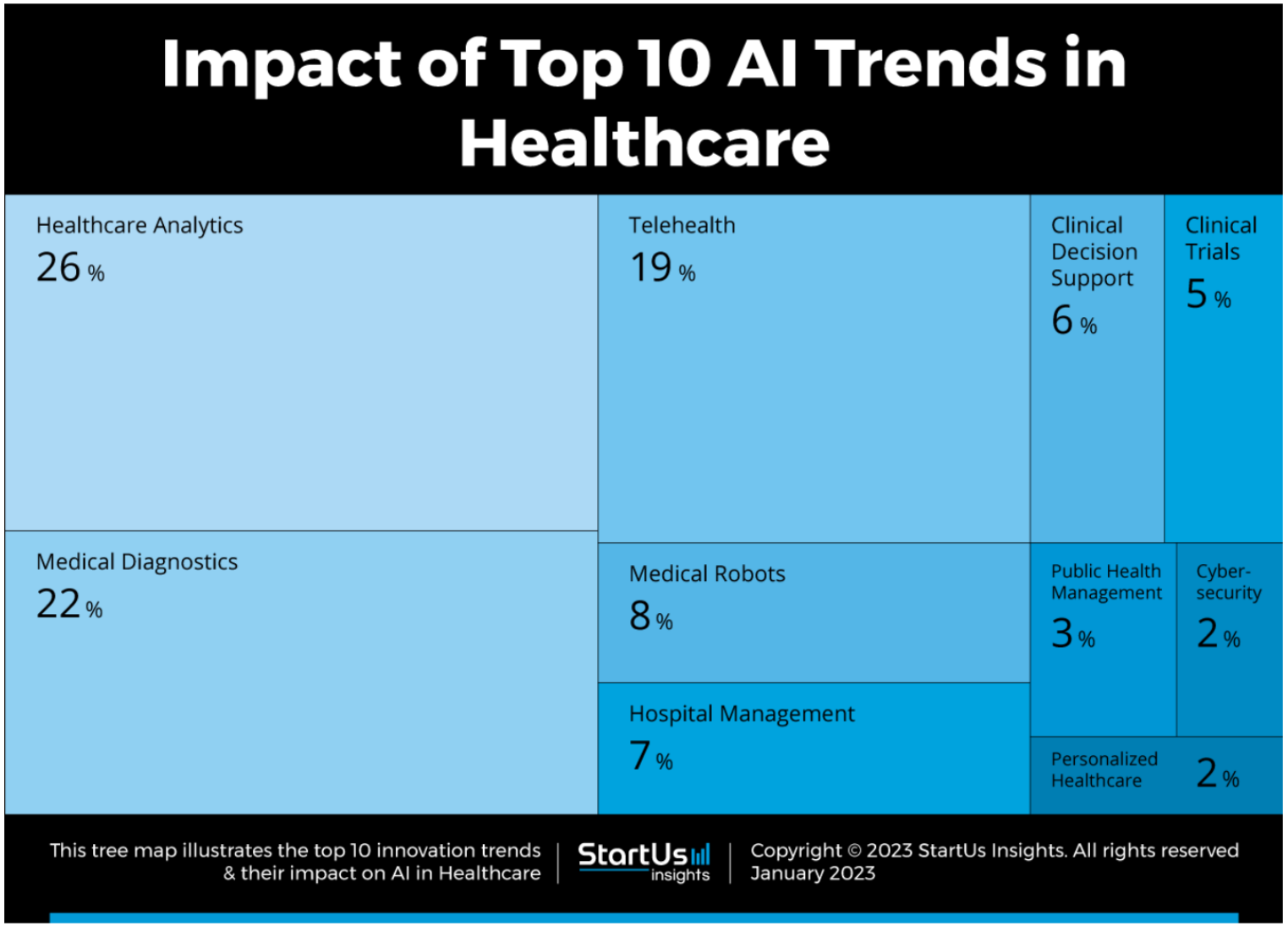
Which is approximately **12% of the total European healthcare expenditures** in 2018[2]

1.8 billion hours freed up every year

That's the equivalent of having **500.000 additional full time health care professionals**[3]

https://www.medtecheurope.org/resource-library/the-socio-economic-impact-of-ai-in-healthcare-addressing-barriers-to-adoption-for-new-healthcare-technologies-in-europe/

Analytics and diagnostics make up more than ½ of AI trends in healthcare.



Today, medical diagnostics are dominated by three clinical specialties.

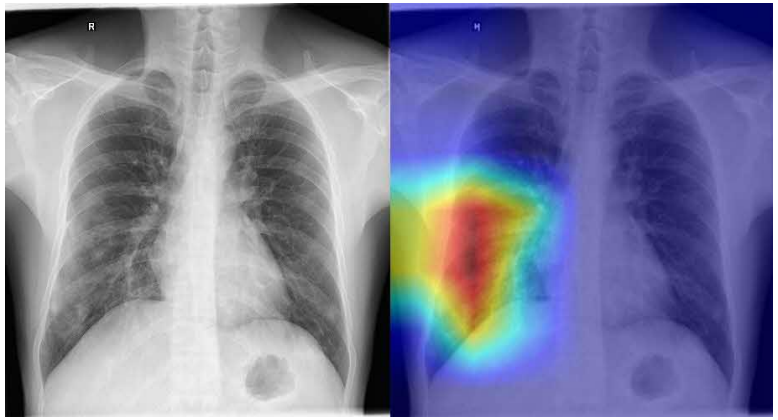
Diagnostic Radiology

Laboratory Medicine

Pathology

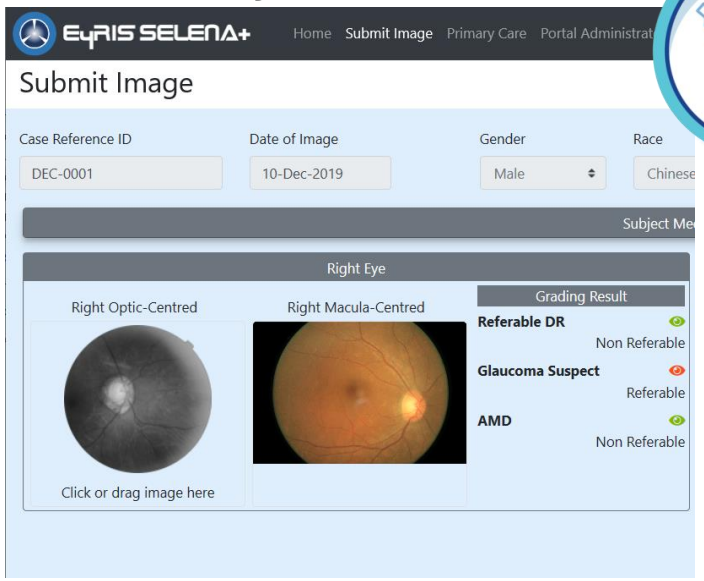
AI excels in pattern recognition

Image-based algorithms are low-hanging fruits for (narrow) AI

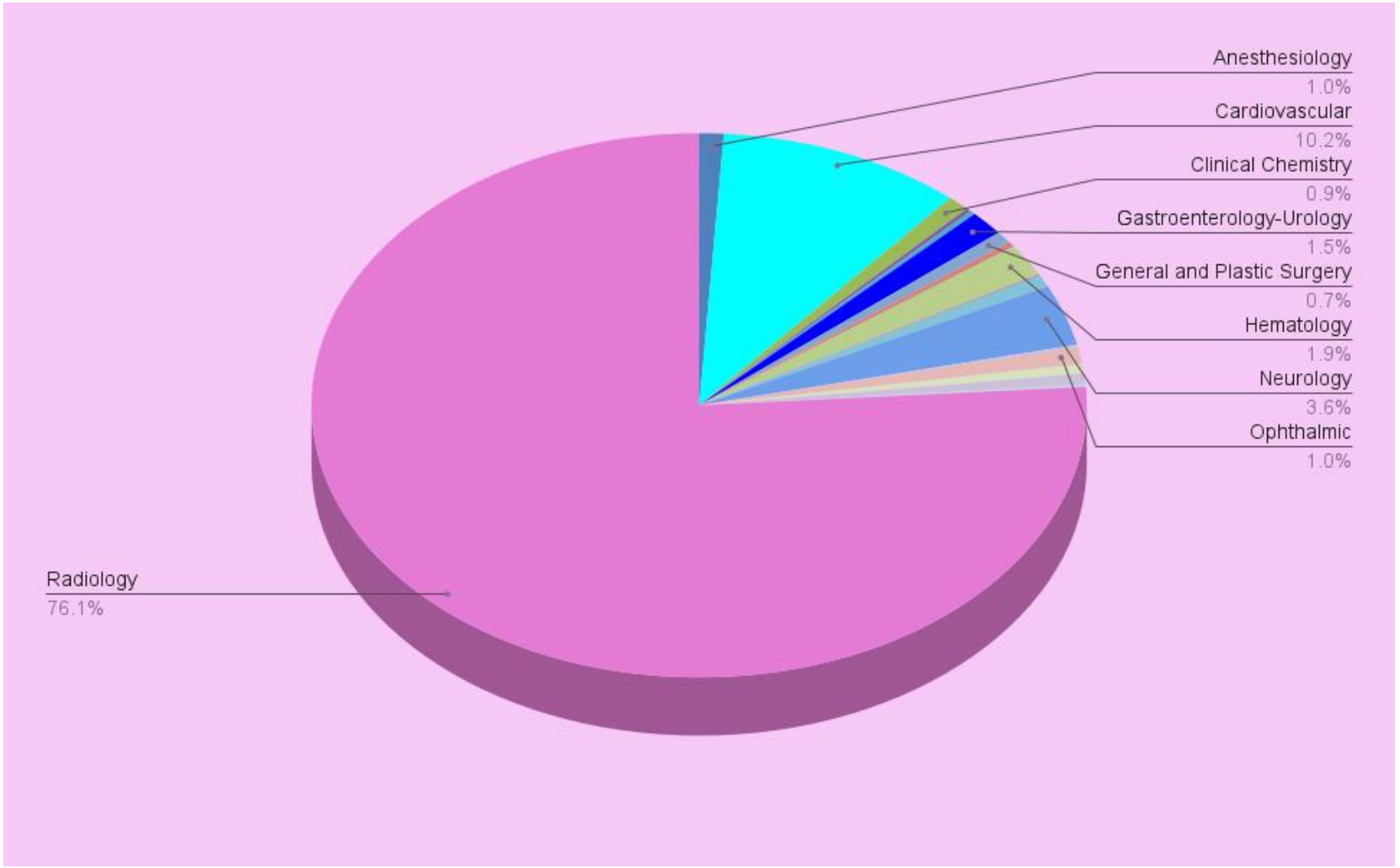


Radilogic: Chest X-ray COVID-19 detection

Selena+:
Diabetic eye disease detection

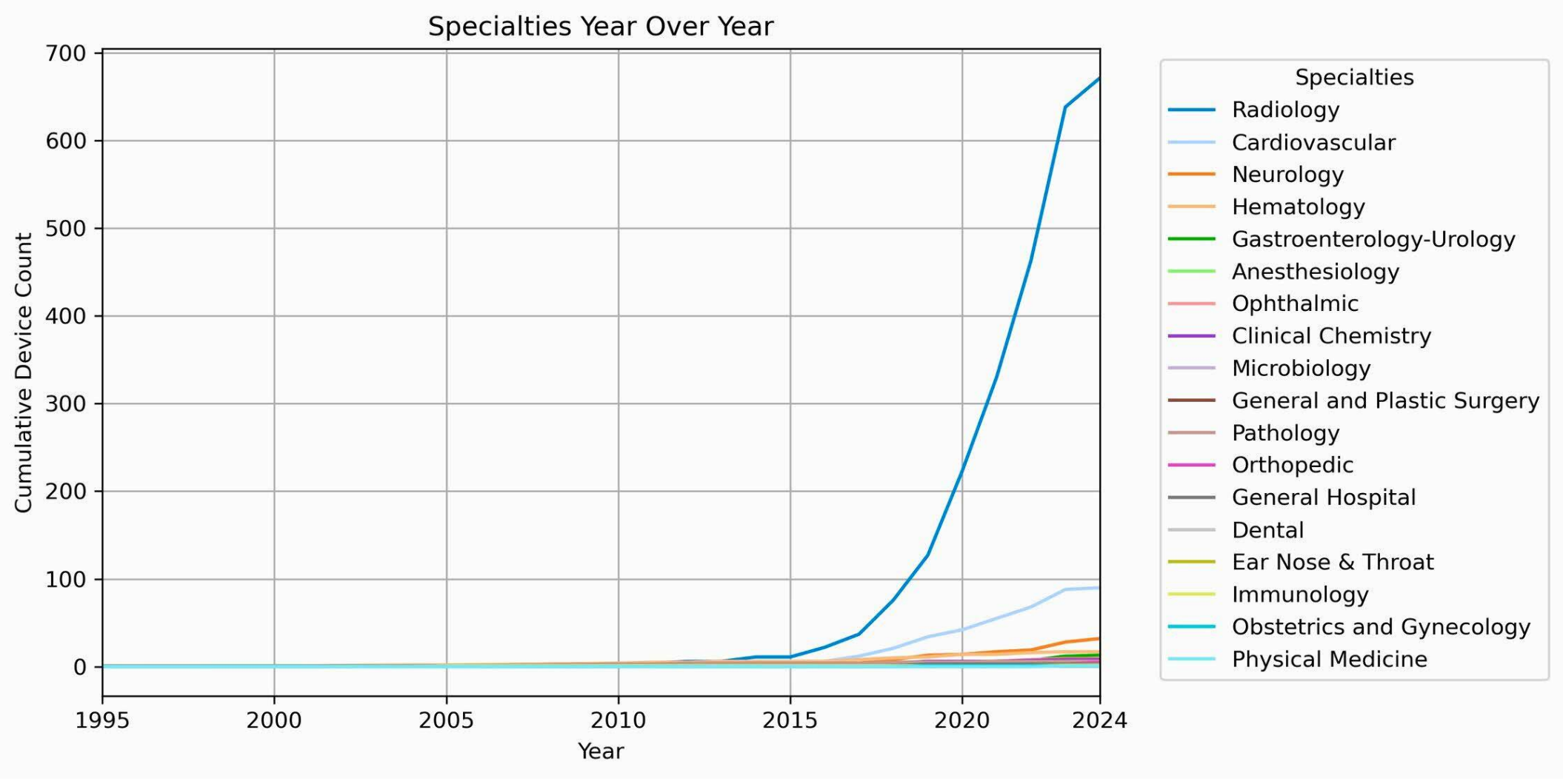


Blade: Cancerous blood cell analysis



https://www.linkedin.com/pulse/882-fda-authorized-ai-enabled-medical-devices-margaretta-colangelo-lofrf/

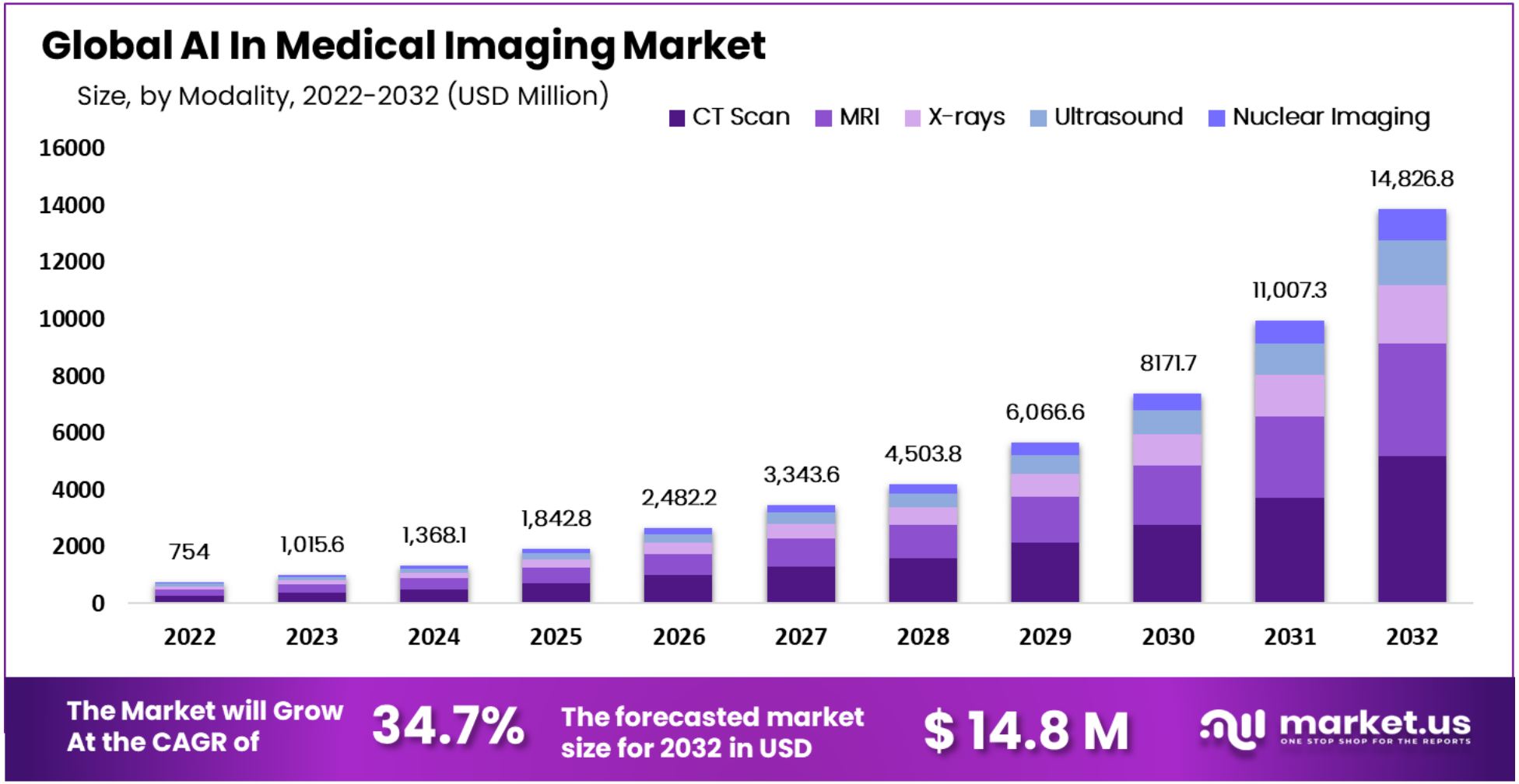
Radiology accounted for > 2/3 of FDA approved AI-MD solutions in 2023.



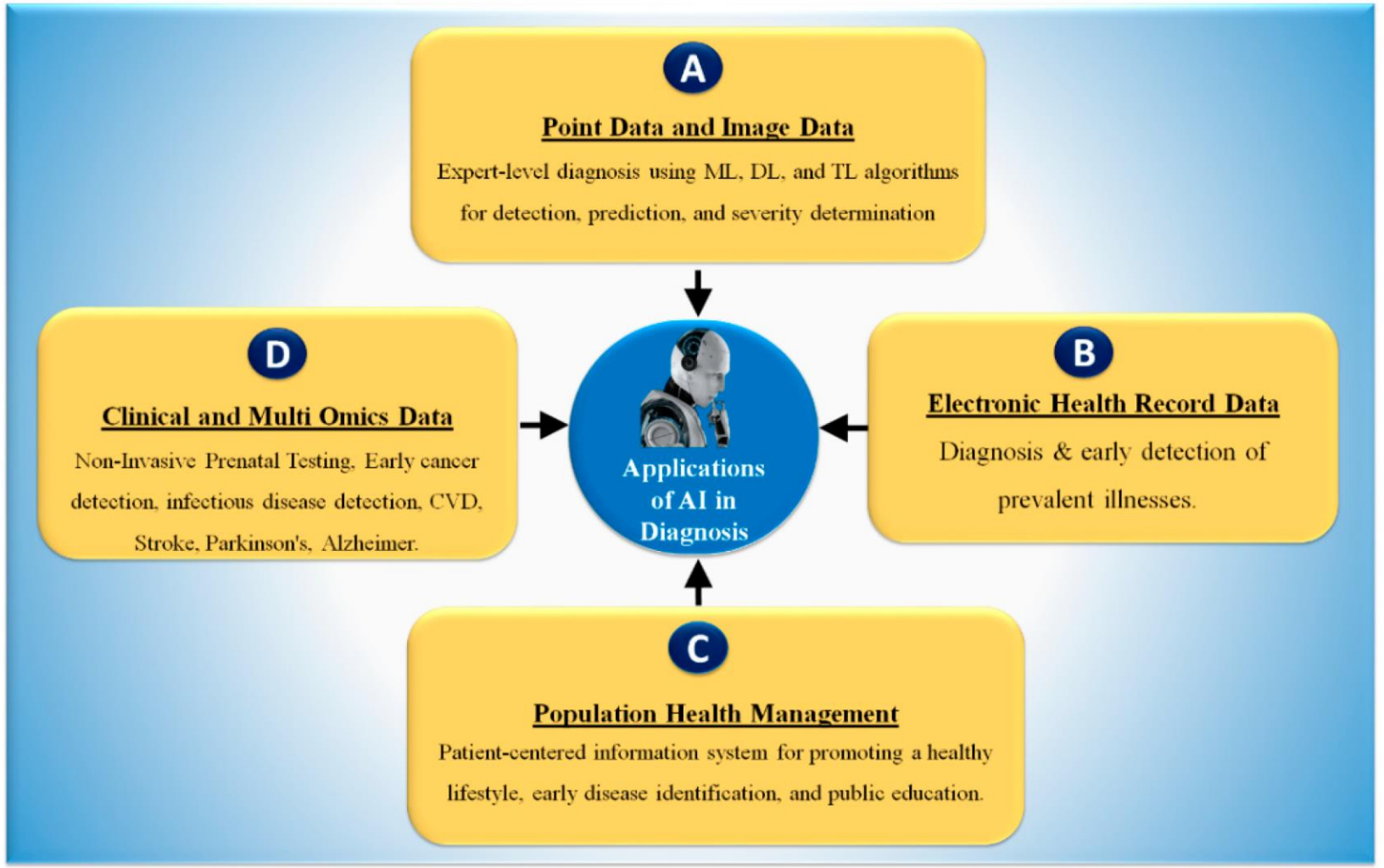
Radiology leads the pack of rapidly increasing specialty-based solutions.

https://www.linkedin.com/pulse/882-fda-authorized-ai-enabled-medical-devices-margaretta-colangelo-lofrf/

The projected growth of AI in the medical imaging market is exponentially increasing.



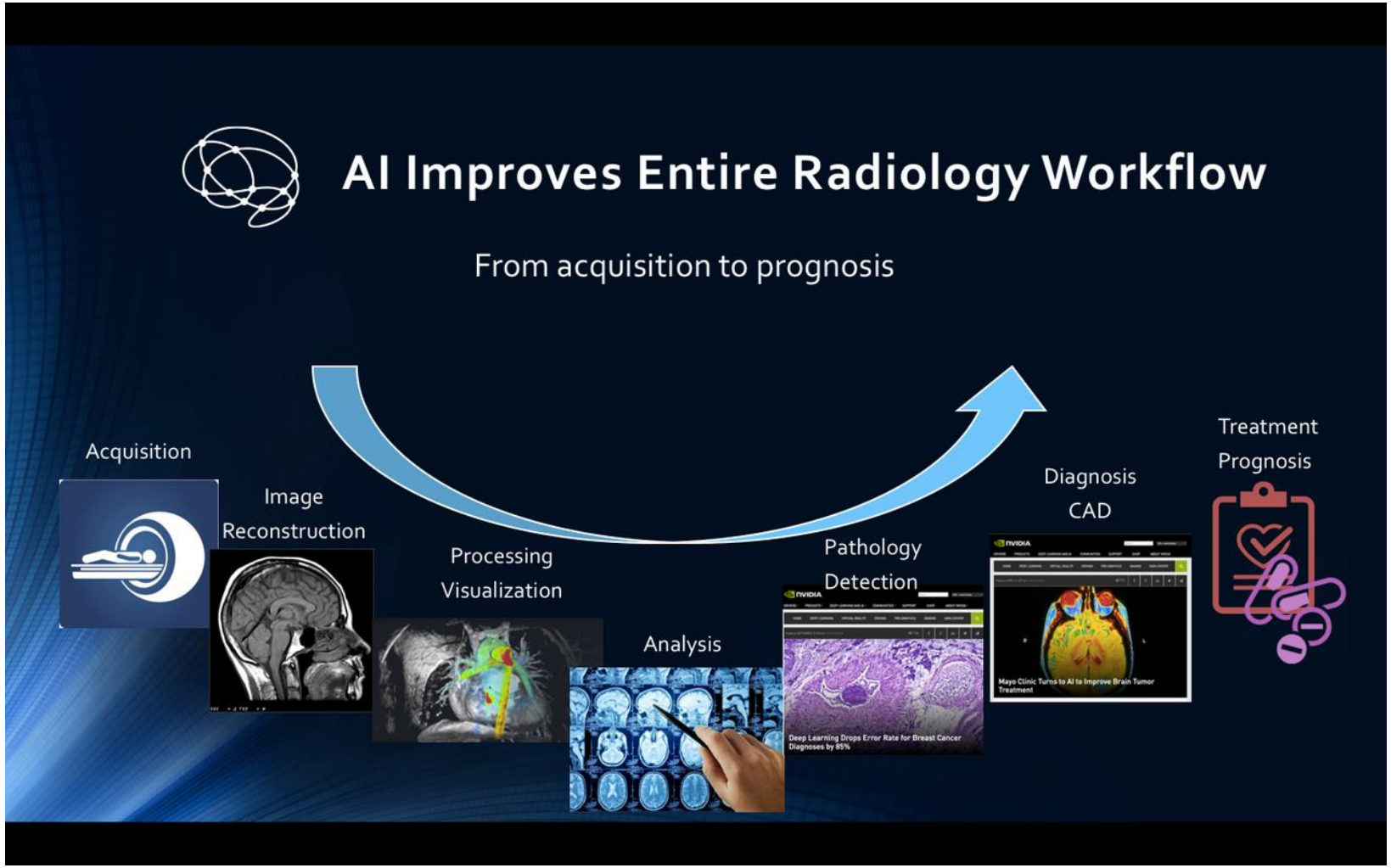
https://market.us/report/ai-in-medical-imaging-market/



https://www.mdpi.com/2227-9032/10/12/2493

The holy grail: Diagnostic AI incorporating multimodal data sources.

Particularly with generative AI, there are also immense opportunities for entire workflow optimization.



Will AI replace radiologists?

“We should stop training radiologists now. It’s just completely obvious that within five years, deep learning is going to do better than radiologists.”

- Geoffrey Hinton (Turing Award recipient; "the Godfather of AI“) in 2016

Why Won't AI Replace Radiologists?

Early failures in model performance and data privacy spooked providers.

AI + ML

This article is more than **1 year old**

Machine learning the hard way: IBM Watson's fatal misdiagnosis

The doctor won't see you now

Rupert Goodwins

Mon 31 Jan 2022 // 10:29 UTC

OPINION It started in Jeopardy and ended in loss. IBM's flagship AI Watson Health has been sold to venture capitalists for an undisclosed sum thought to be around a billion dollars, or a quarter of what the division cost IBM in acquisitions alone since it was spun off in 2015.

https://www.theregister.com/2022/01/31/machine_learning_the_hard_way/

DeepMind faces legal action over NHS data use

1 October 2021

https://www.healthcareitnews.com/news/emea/google-and-deepmind-face-legal-claim-unauthorised-use-nhs-medical-records

GETTY IMAGES

A legal case has been launched on behalf of more than a million people whose confidential medical records were obtained by Google.

Real-world clinical performance may not be as good as lab-based performance.

Radiology

ORIGINAL RESEARCH • THORACIC IMAGING

Commercially Available Chest Radiograph AI Tools for Detecting Airspace Disease, Pneumothorax, and Pleural Effusion

Louis Lind Plesner, MD • Felix C. Müller, MD, PhD • Mathias W. Brejnebøl, MD • Lene C. Laustrup, MD • Finn Rasmussen, MD, DMSc • Olav W. Nielsen, MD, PhD • Mikael Boesen, MD, PhD • Michael Brun Andersen, MD, PhD**

Conclusion: Current-generation AI tools showed moderate to high sensitivity for detecting airspace disease, pneumothorax, and pleural effusion on chest radiographs. However, they produced more false-positive findings than radiology reports, and their performance decreased for smaller-sized target findings and when multiple findings were present.

This may negatively impact on productivity.

npj | digital medicine

www.nature.com/npjdigitalmed

ARTICLE OPEN

Check for updates

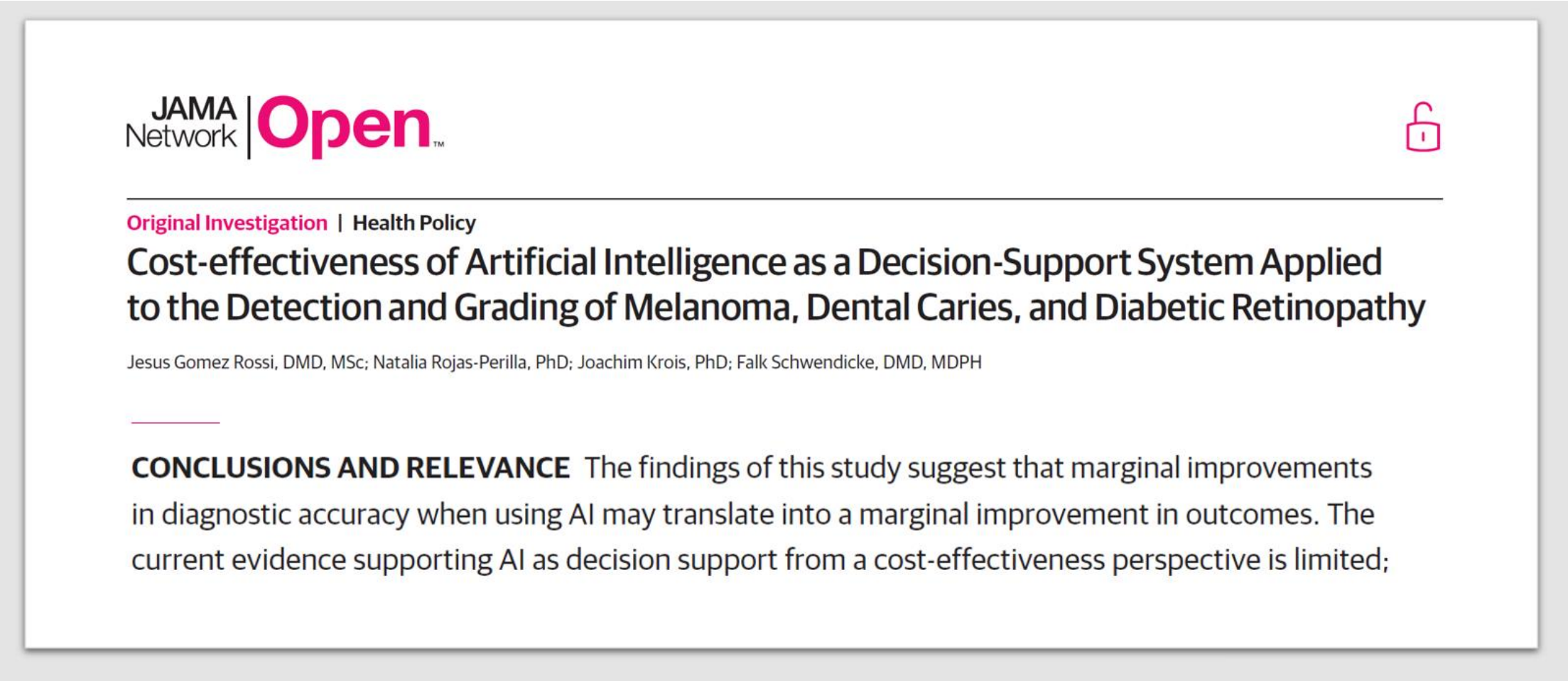
The impact of artificial intelligence on the reading times of radiologists for chest radiographs

Hyun Joo Shin[1,2], Kyunghwa Han[3], Leeha Ryu[4] and Eun-Kyung Kim[1,2]✉

Whether the utilization of artificial intelligence (AI) during the interpretation of chest radiographs (CXRs) would affect the radiologists' workload is of particular interest. Therefore, this prospective observational study aimed to observe how AI affected the reading times of radiologists in the daily interpretation of CXRs. Radiologists who agreed to have the reading times of their CXR interpretations collected from September to December 2021 were recruited. Reading time was defined as the duration in seconds from opening CXRs to transcribing the image by the same radiologist. As commercial AI software was integrated for all CXRs, the radiologists could refer to AI results for 2 months (AI-aided period). During the other 2 months, the radiologists were automatically blinded to the AI results (AI-unaided period). A total of 11 radiologists participated, and 18,680 CXRs were included. Total reading times were significantly shortened with AI use, compared to no use (13.3 s vs. 14.8 s, $p < 0.001$). When there was no abnormality detected by AI, reading times were shorter with AI use (mean 10.8 s vs. 13.1 s, $p < 0.001$). However, if any abnormality was detected by AI, reading times did not differ according to AI use (mean 18.6 s vs. 18.4 s, $p = 0.452$). Reading times increased as abnormality scores increased, and a more significant increase was observed with AI use (coefficient 0.09 vs. 0.06, $p < 0.001$). Therefore, the reading times of CXRs among radiologists were influenced by the availability of AI. Overall reading times shortened when radiologists referred to AI; however, abnormalities detected by AI could lengthen reading times.

npj Digital Medicine (2023)6:82 ; https://doi.org/10.1038/s41746-023-00829-4

Even if there were improved accuracies, its actual value has not been clearly shown.



JAMA Network | Open

Original Investigation | Health Policy

Cost-effectiveness of Artificial Intelligence as a Decision-Support System Applied to the Detection and Grading of Melanoma, Dental Caries, and Diabetic Retinopathy

Jesus Gomez Rossi, DMD, MSc; Natalia Rojas-Perilla, PhD; Joachim Krois, PhD; Falk Schwendicke, DMD, MDPH

CONCLUSIONS AND RELEVANCE The findings of this study suggest that marginal improvements in diagnostic accuracy when using AI may translate into a marginal improvement in outcomes. The current evidence supporting AI as decision support from a cost-effectiveness perspective is limited;

Real-world evidence raise question on credibility of use of AI.

JAMA Internal Medicine | Original Investigation

External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients

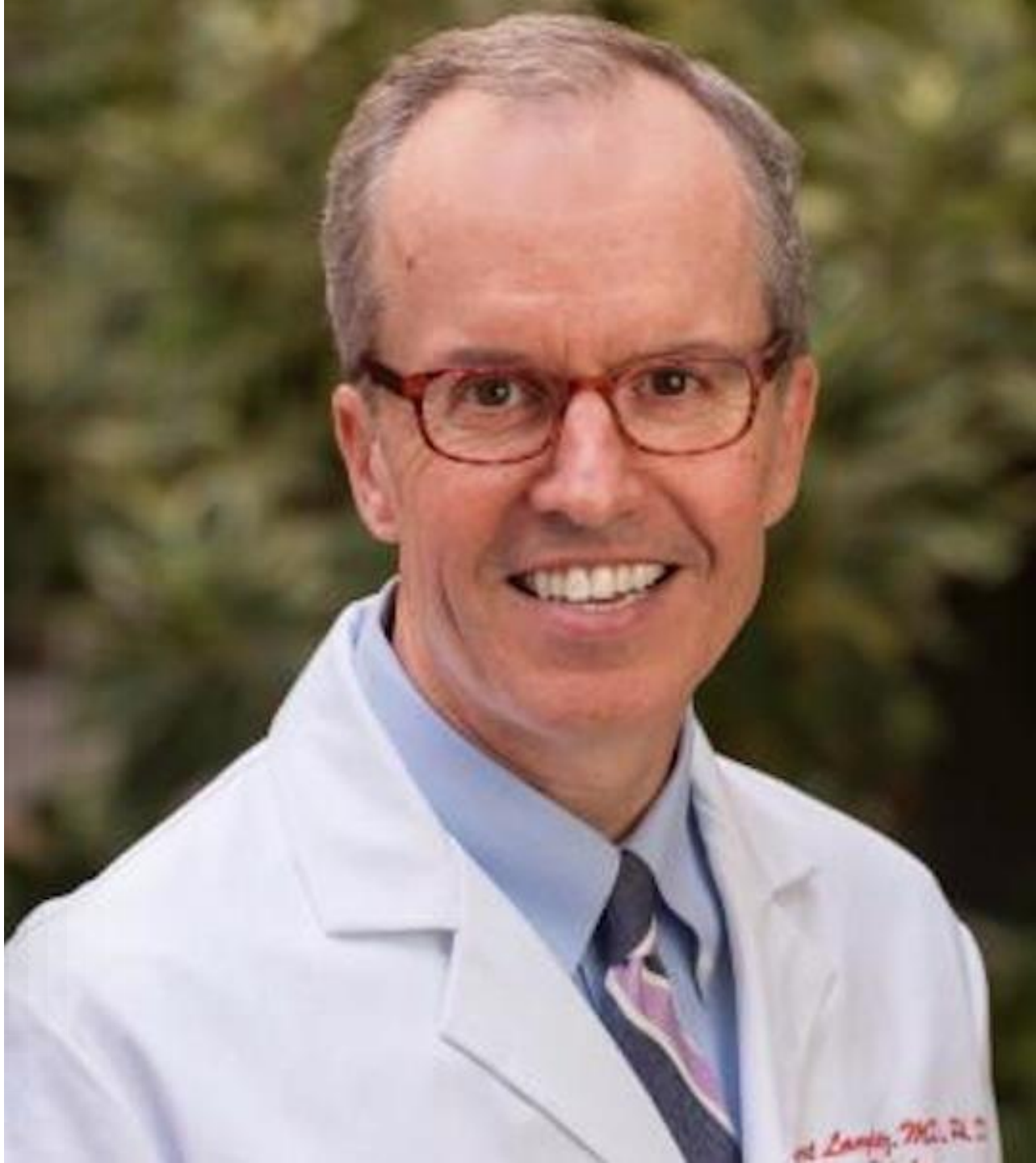
Andrew Wong, MD; Erkin Otles, MEng; John P. Donnelly, PhD; Andrew Krumm, PhD; Jeffrey McCullough, PhD; Olivia DeTroyer-Cooley, BSE; Justin Pestrue, MEcon; Marie Phillips, BA; Judy Konye, MSN, RN; Carleen Penoza, MHSA, RN; Muhammad Ghous, MBBS; Karandeep Singh, MD, MMSc

CONCLUSIONS AND RELEVANCE This external validation cohort study suggests that the ESM has poor discrimination and calibration in predicting the onset of sepsis. The widespread adoption of the ESM despite its poor performance raises fundamental concerns about sepsis management on a national level.

Will AI ever replace radiologists?

“I say the answer is no—but radiologists who use AI will replace radiologists who don’t.”

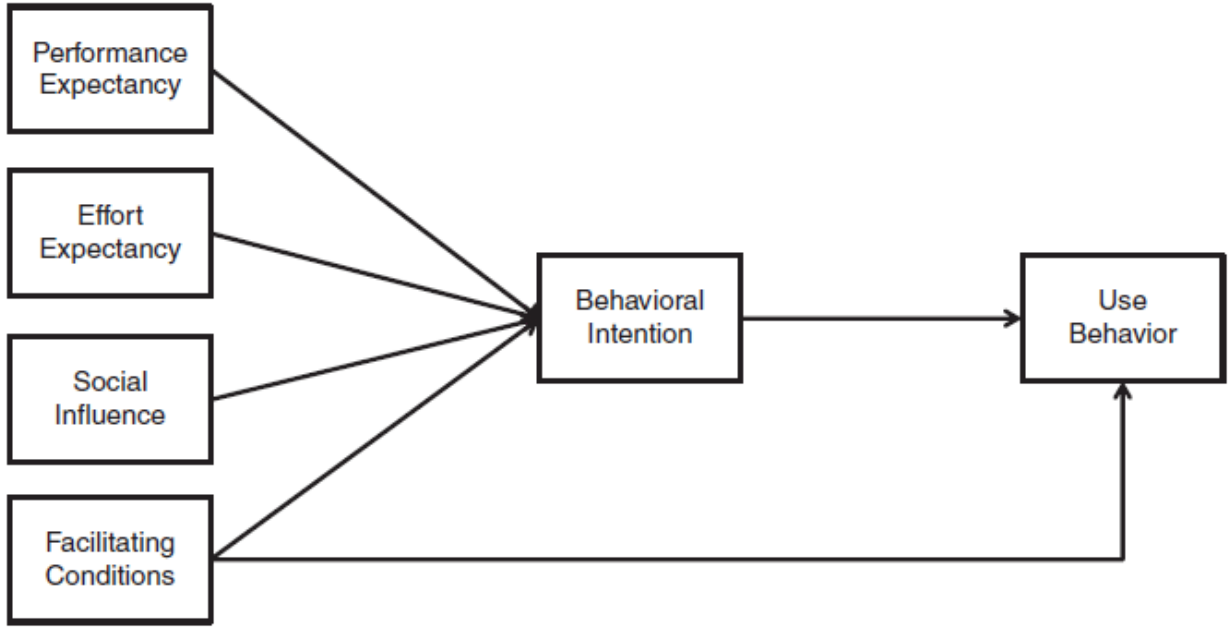
- Curtiz Langlotz, Director of Center for AI in Medicine & Imaging, Stanford University



Model performance is the most critical aspect for clinical adoption of AI.

Unified Theory of Acceptance and Use of Technology (UTAUT)

- Performance expectancy (Trust)
- Effort expectancy (Literacy)
- Social Influence
- Facilitating conditions (User interface)
- …



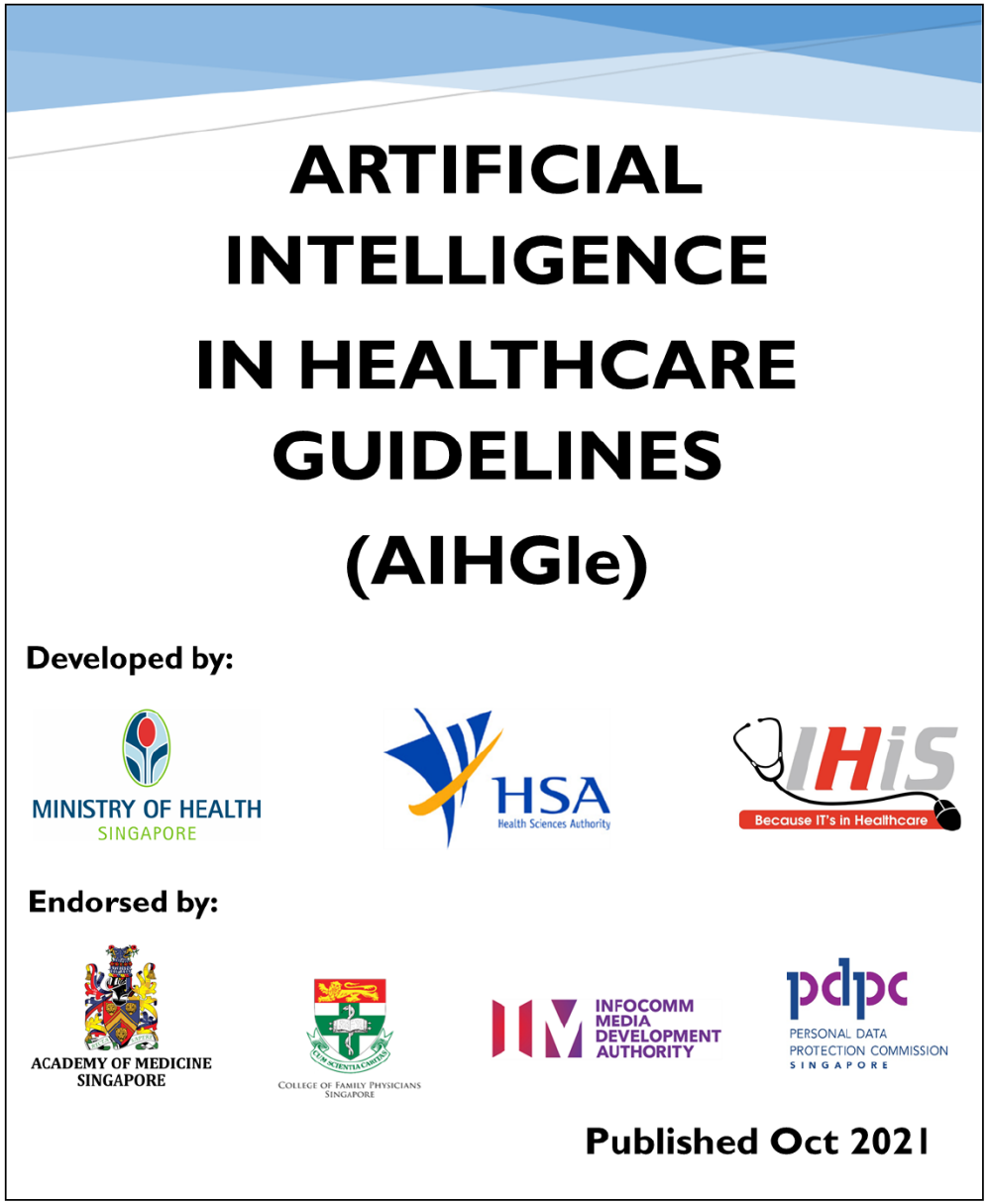
"…medical performance expectancy was found to have the largest impact."

-Cornelissen J, et al. JMIR Form Res. 2022 Jun; 6(6): e33368.

The greatest barriers in the local healthcare system are integration and cost concerns.

https://insights.daffodilsw.com/blog/six-major-barriers-to-ai-adoption-their-practical-solutions

Clinical trials in AI-based diagnostics: What do we need to consider?



4.10 Evaluation and Monitoring of AI-MD

Validation

4.10.1 **Developers should periodically evaluate and validate their AI-MD's performance to ensure it minimally meets the clinical practice baseline** (*see Section 4.4*), and verify the accuracy, and reproducibility of the AI-MD's algorithmic decisions.

4.10.2 The clinical performance of an AI-MD involves more than just technical measures of its algorithm's performance (e.g. Area under the Curve (AUC) – Receiving Operating Characteristic Curve (ROC) or Precision-Recall Curve (PRC), True Positive Rate, Positive/Negative Predictive Value, Cohen's Kappa Score, etc.). Developers should work with implementers to ensure that the actual clinical outcomes[14] of the AI-MD (i.e. impact on patients when the AI-MD is introduced to their care) are measured and assessed.

4.10.3 **Table 3** sets out a suggested stepwise AI-MD validation approach, and the types of risks it assesses. Developers should compare results from each step with the current clinical practice baseline.

For AI-MD, trials need to look beyond safety and efficacy…

Fig. 4. Characteristics of trustworthy AI systems. Valid & Reliable is a necessary condition of trustworthiness and is shown as the base for other trustworthiness characteristics. Accountable & Transparent is shown as a vertical box because it relates to all other characteristics.

"Artificial Intelligence Risk Management Framework (AI RMF 1.0)", NIST National Institute of Standards and Technology (USA Department of Commerce) 2023

...and beyond model performance.

How medical AI devices are evaluated: limitations and recommendations from an analysis of FDA approvals

A comprehensive overview of medical AI devices approved by the US Food and Drug Administration sheds new light on limitations of the evaluation process that can mask vulnerabilities of devices when they are deployed on patients.

- **Generalisability:**
  - Evaluating the performance of AI devices in multiple clinical sites is important for ensuring that the algorithms perform well across representative populations.
- **Reproducibility:**
  - Encouraging prospective studies with comparison to standard of care reduces the risk of harmful overfitting and more accurately captures true clinical outcomes.
- **Real-world evidence:**
  - Post-market surveillance of AI devices is also needed for understanding and measurement of unintended outcomes and biases that are not detected in prospective, multi-center trials.
  - For example, most computer-aided detection diagnostic devices are intended to be decision-support tools rather than primary diagnostic tools. A prospective randomized study may reveal that clinicians are misusing this tool for primary diagnosis and that outcomes are different from what would be expected if the tool were used for decision support

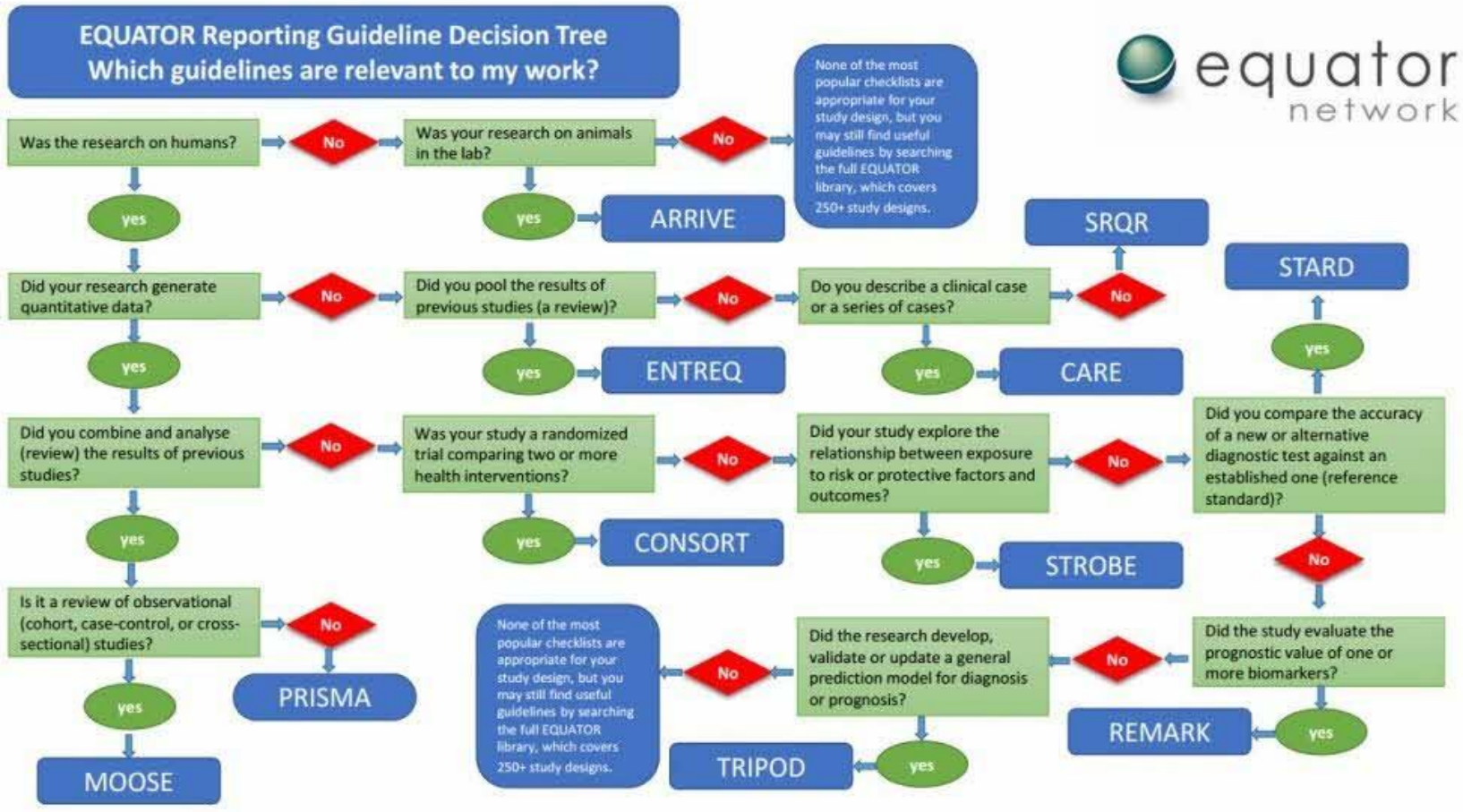
- Wu et al, Nature Medicine, 2021

What should trials demonstrate?

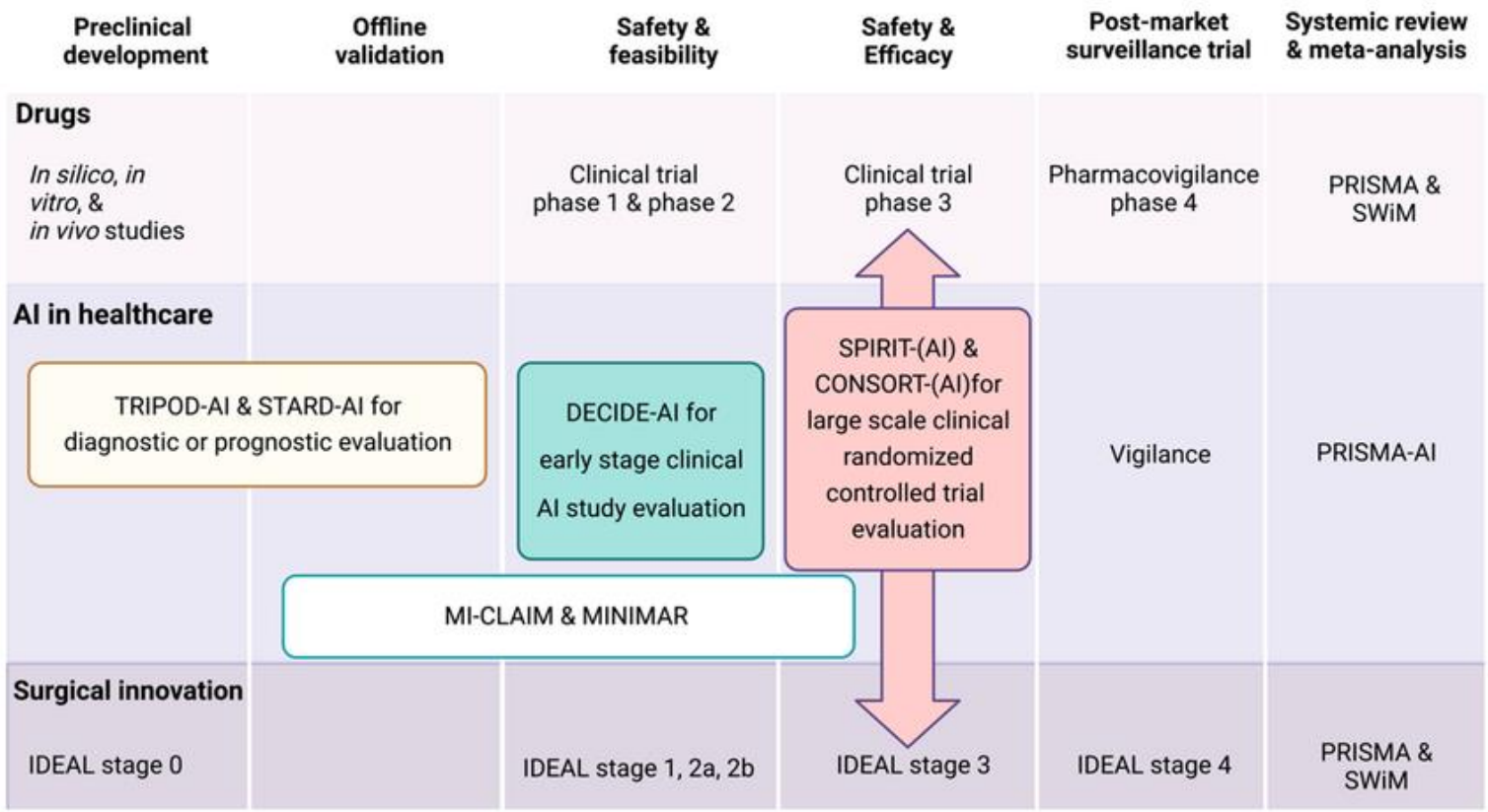
Key factor	Main considerations	Opportunities
Performance expectancy	• Clinical performance (accuracy, sensitivity, etc) • Technical performance (robustness, explainability, generalisability, etc)	• Randomised controlled clinical trials or equivalent • Standardised technical evaluation of AI solutions
Effort expectancy	• AI know-how • User interface / User experience	• Clinician education and training • User-centric interface design
Social influence	• Peer user experience • Patient acceptance	• Change management • Patient education
Facilitating conditions	• Clinical workflow integration • Digital infrastructure • Cost-effectiveness	• Care and process re-design • Data-sharing platforms • Value assessment

Tan CH, et al. SMJ 2024.

Taking reference from the EQUATOR (Enhancing the QUAlity and Transparency Of health Research)…



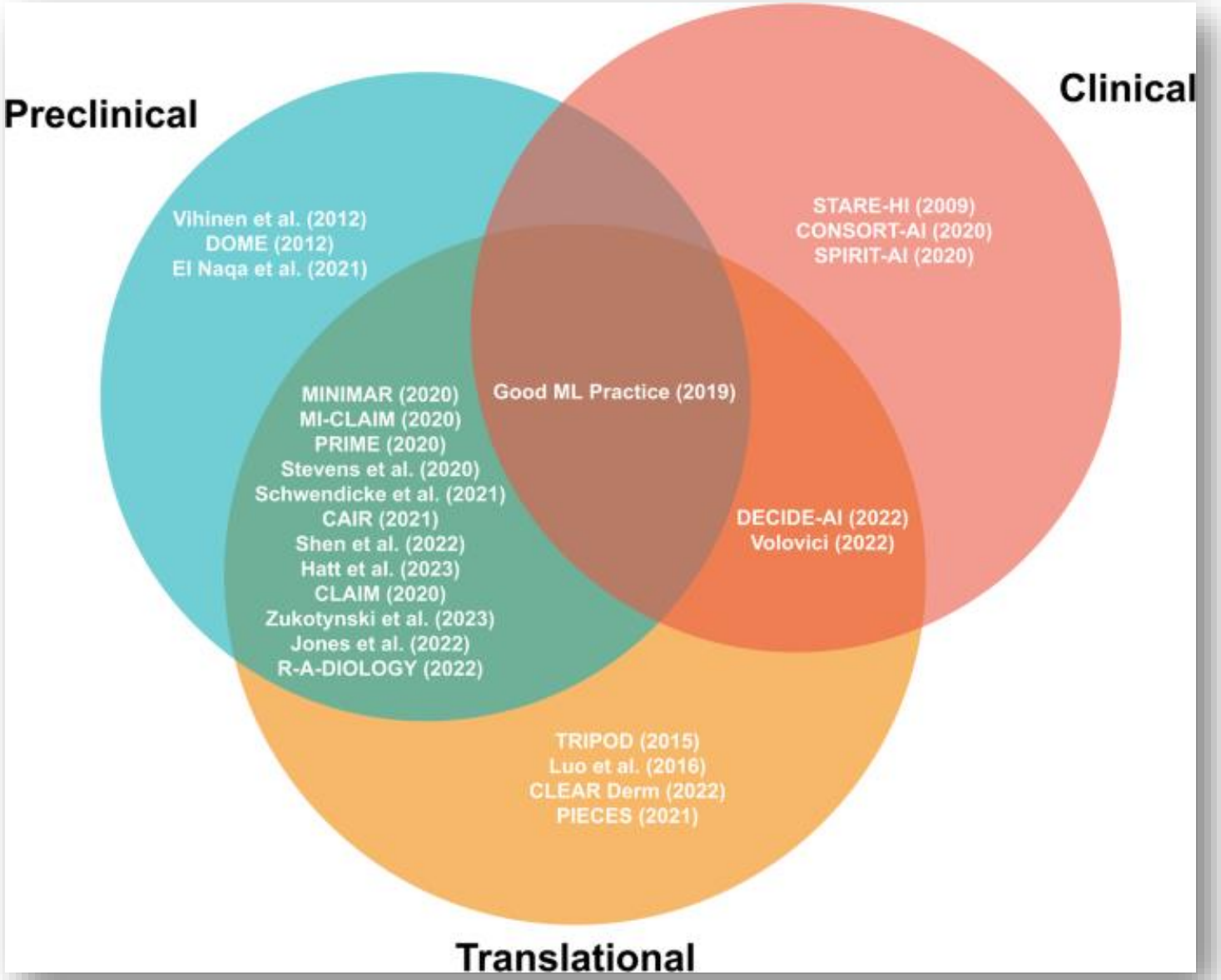
...parallel standards for reporting of trials of AI solutions have been defined according to phase of development and intended use.



Salybekov, et al. Artificial Intelligence Reporting Guidelines' Adherence in Nephrology for Improved Research and Clinical Outcomes. Biomedicines 2024

MINIMAR (MINimum Information for Medical AI Reporting)

There are now >20 guidelines for reporting of studies for AI in healthcare



Kolbinger, F.R., et al. Reporting guidelines in medical artificial intelligence: a systematic review and meta-analysis. Commun Med 4, 71 (2024).

But what's the catch?

Radiology: Artificial Intelligence

EDITORIAL

AI Reporting Guidelines: How to Select the Best One for Your Research

Michail E. Klontzas, MD, PhD • Anthony A. Gatti, PhD • Ali S. Tejani, MD • Charles E. Kahn, Jr, MD, MS

Item 22. Detailed description of model. If novel model architecture is used, provide a complete and detailed structure of the model, including inputs, outputs, and all intermediate layers, in sufficient detail that another investigator could exactly reconstruct the network. For neural network models, include all details of pooling, normalization, regularization, and activation in the layer descriptions. Model inputs must match the form of the preprocessed data. Model outputs must correspond to the requirements of the stated clinical problem, and for supervised learning should match the form of the reference standard annotations. If a previously published model architecture is employed, cite a reference that meets the preceding standards and fully describe every modification made to the model. Cite a reference for any proprietary model described previously, as well. In some cases, it may be more convenient to provide the structure of the model in code as supplemental data.

- *Checklist for Artificial Intelligence in Medical Imaging (CLAIM): 2024 Update: https://doi.org/10.1148/ryai.240300*

Is the jury still out on the Epic Sepsis Model?

ORIGINAL CLINICAL REPORT

OPEN

Epic Sepsis Model Inpatient Predictive Analytic Tool: A Validation Study

John Cull, MD
Robert Brevetta, DO
Jeff Gerac, MD
Shanu Kothari, MD
Dawn Blackhurst, DrPH

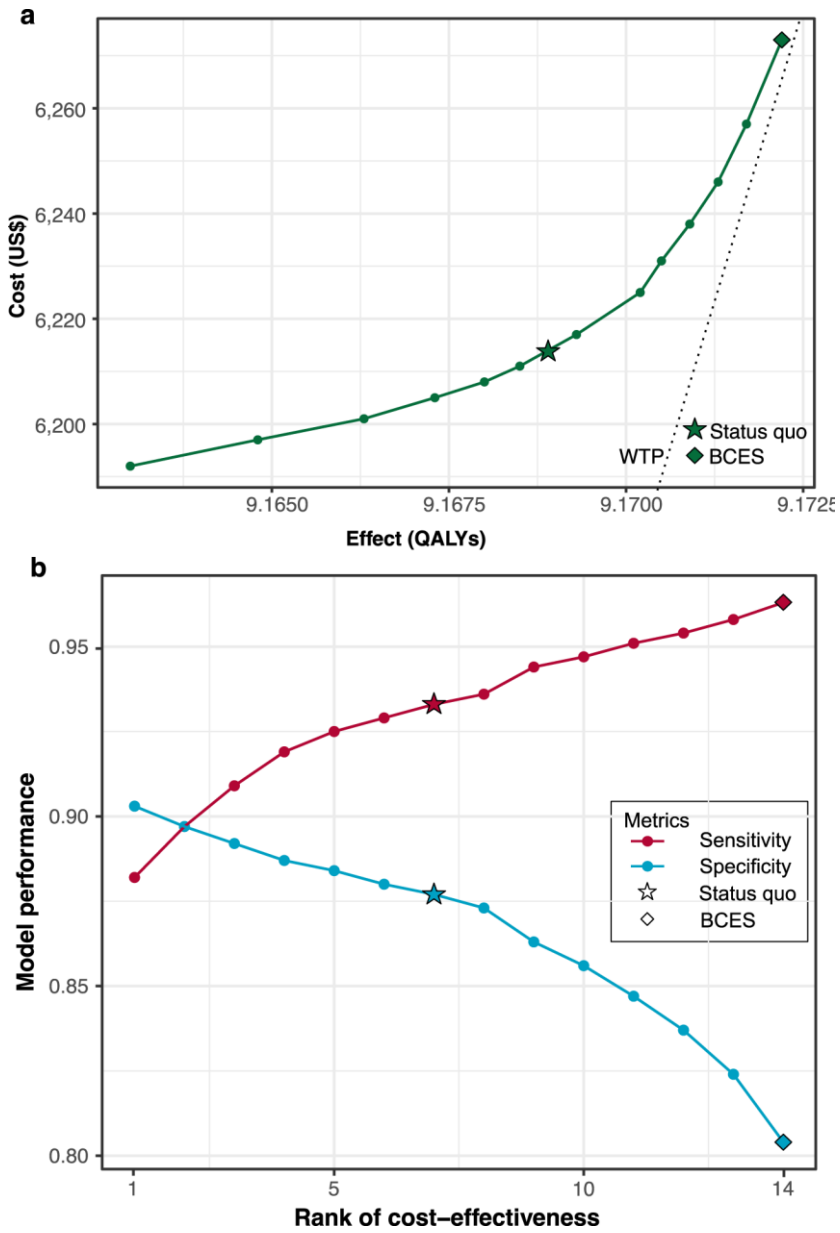
OBJECTIVES: Earlier treatment of sepsis leads to decreased mortality. Epic is an electronic medical record providing a predictive alert system for sepsis, the Epic Sepsis Model (ESM) Inpatient Predictive Analytic Tool. External validation of this system is lacking. This study aims to evaluate the ESM as a sepsis screening tool and determine whether an association exists between ESM alert system implementation and subsequent sepsis-related mortality.

CONCLUSIONS: In this single-center before-and-after study, utilization of the ESM score as a screening test was associated with a 44% reduction in the odds of sepsis-related mortality. Due to wide utilization of Epic, this is a potentially promising tool to improve sepsis mortality in the United States. This study is hypothesis generating, and further work with more rigorous study design is needed.

"There are several limitations of the study by Wong et al (23) that could explain the differing results. First, in the study by Wong et al (23), the **score threshold for alerting the physicians** for the possibility of sepsis was arbitrarily chosen at a score threshold of 6 or higher without validating the scoring system at their institution....Wong et al (23) found that only **7%** of patients with sepsis who were missed by a clinician based on timely administration of antibiotics prior to an ESM score of greater than 6, while **31.2%** of patients in our study with sepsis were not on sepsis-appropriate antibiotics prior to the ESM score of greater than or equal to 5."

Trials need to demonstrate real-world value.

- Model performance is highly dependent on the use case and operating thresholds among other factors.
- It is important to define the implementation strategy including the clinical workflow and payor-provider relationship.
- Ultimately, the cost-effectiveness will be a key determinant of clinical trials for diagnostics AI.



The best cost-effective scenario (rank 14) was identified with a sensitivity of 96.3% and a specificity of 80.4%. Performance of the AI with a large decrease in specificity but a minor increase in sensitivity can still benefit cost-effectiveness.

Wang Y, et al. npj Digit. Med. 7, 43 (2024).

Take home messages

01

Low-hanging fruits for medical AI are in the imaging diagnostics.

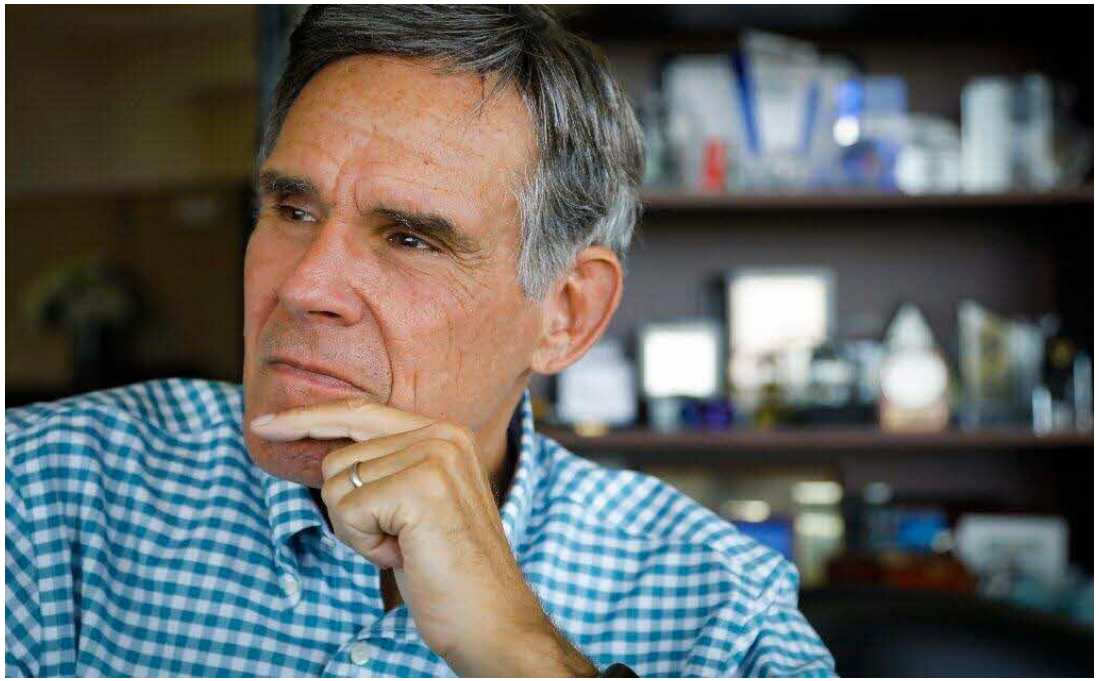
02

We are at the tipping point of adoption of AI into routine clinical care.

03

Clinical trials for AI adoption need to incorporate implementation and demonstrate real-world value.

"The greatest opportunity offered by AI is not reducing errors or workloads, or even curing cancer: it is the *opportunity to restore the precious and time-honored connection and trust*—the human touch—between patients and doctors."



- Eric Topol, Founder and director of the Scripps Research Translational Institute

Thank you

AI has potential to automate or enhance various aspects of clinical trials.

https://www.linkedin.com/pulse/ai-automation-clinical-trials-part-i-manuj-vangipurapu/

Explainable AI is still a distance away from us...