American College of Radiology™

The Journey to Safe and Effective AI Adoption

Gilbert W. Beebe Symposium 2025

Mike Tilkin, CIO and EVP Technology
American College of Radiology
March 13, 2025

acr.org

Disclosures

- Employee of the American College of Radiology

Agenda

Background
Recent Provider Perspectives
Navigating the Evolving AI Market
Deploying AI in Production
Notes on LLMs and VLMs

American College of Radiology™
Data Science Institute

100
years

41k
members

39k
facilities

4

Serve patients and society by empowering members to advance the practice, science, and professions of radiological care

ACR Data Science Institute

Established 2017

Advance data science as core to clinically relevant, safe, and effective radiologic care

ACR Data Science Institute Established 2017

Christoph Wald, MD, PhD, MBA, FACR
ACR Commission on Informatics Chair

Keith J. Dreyer, DO, PhD, FACR
ACR DSI Chief Science Officer

Bernardo Bizzo, MD, PhD
ACR DSI Associate Chief Science Officer

Woojin Kim, MD
ACR DSI Chief Medical Officer

Mike Tilkin
ACR Executive Vice President and
Chief Information Officer

Laura Coombs, PhD
ACR Vice President of Informatics

Chris Treml
ACR DSI Director of Operations

Bibb Allen Jr., MD, FACR
Senior Advisor

DSI Oversight Board

- Christoph Wald
- Bibb Allen
- Keith Dreyer
- Amy Kotsenas
- Jackie Bello
- Jim Brink
- Howard Fleishon
- Geraldine McGinty
- Laura Coombs
- Mike Tilkin
- Etta Pisano
- Bill Thorwarth

AI Panel Chairs

6

- Luther B. Adair
- Bhavika Patel
- Robert Rosenberg
- Carlo De Cecco
- Jeremy Durack
- Jay W. Patti
- Alexander M. Norbash
- Christopher T. Whitlow
- Sumit Niogi
- John Tsiouris
- Reid F. Thompson
- Steven Blumer
- Carol Wu
- Alexander J. Towbin
- Adam B. Prater

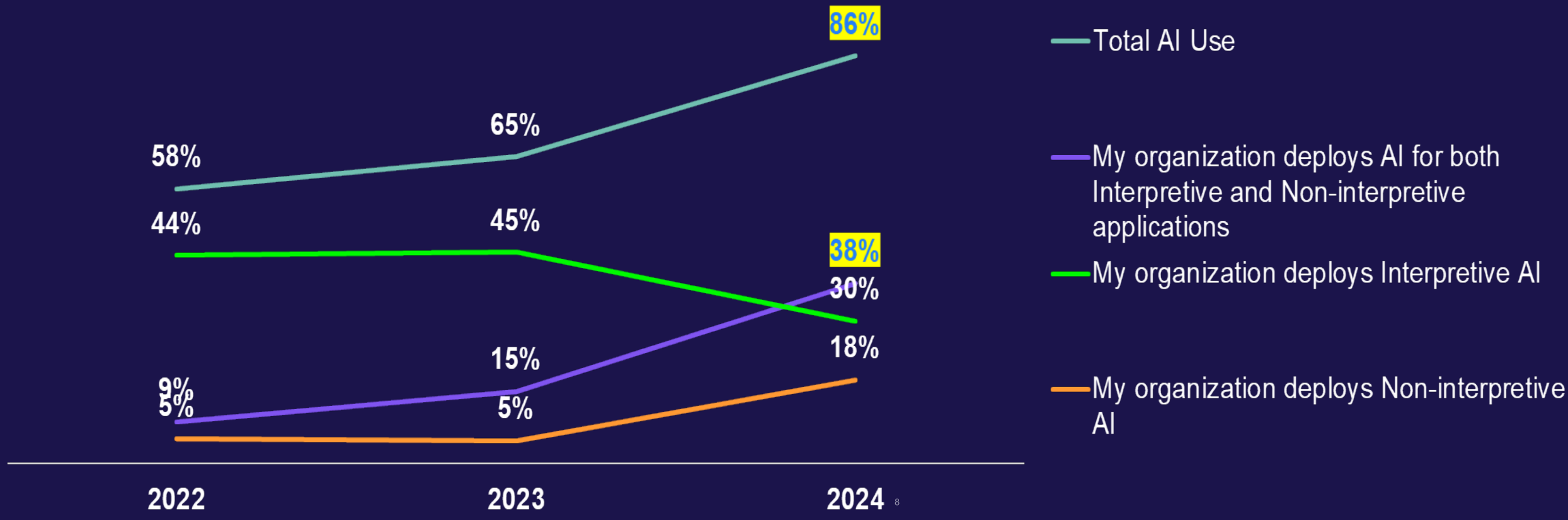
Senior Scientists

- Sheela Agarwal
- Tarik Alkasab
- Katherine P. Andriole
- Bernardo C. Bizzo
- Garryi Choy
- Curtiland Deville. Jr.
- Adam E. Flanders
- J. Raymond Geis
- C. Matthew Hawkins
- Jayashree Kalpathy-Cramer
- Matthew P. Lungren
- Gregory N. Nicola
- Chris L. Sistrom
- Neil Tenenholtz

Provider Perspectives on AI Market

Results from a 2024 ACR DSI AI Tracking Survey

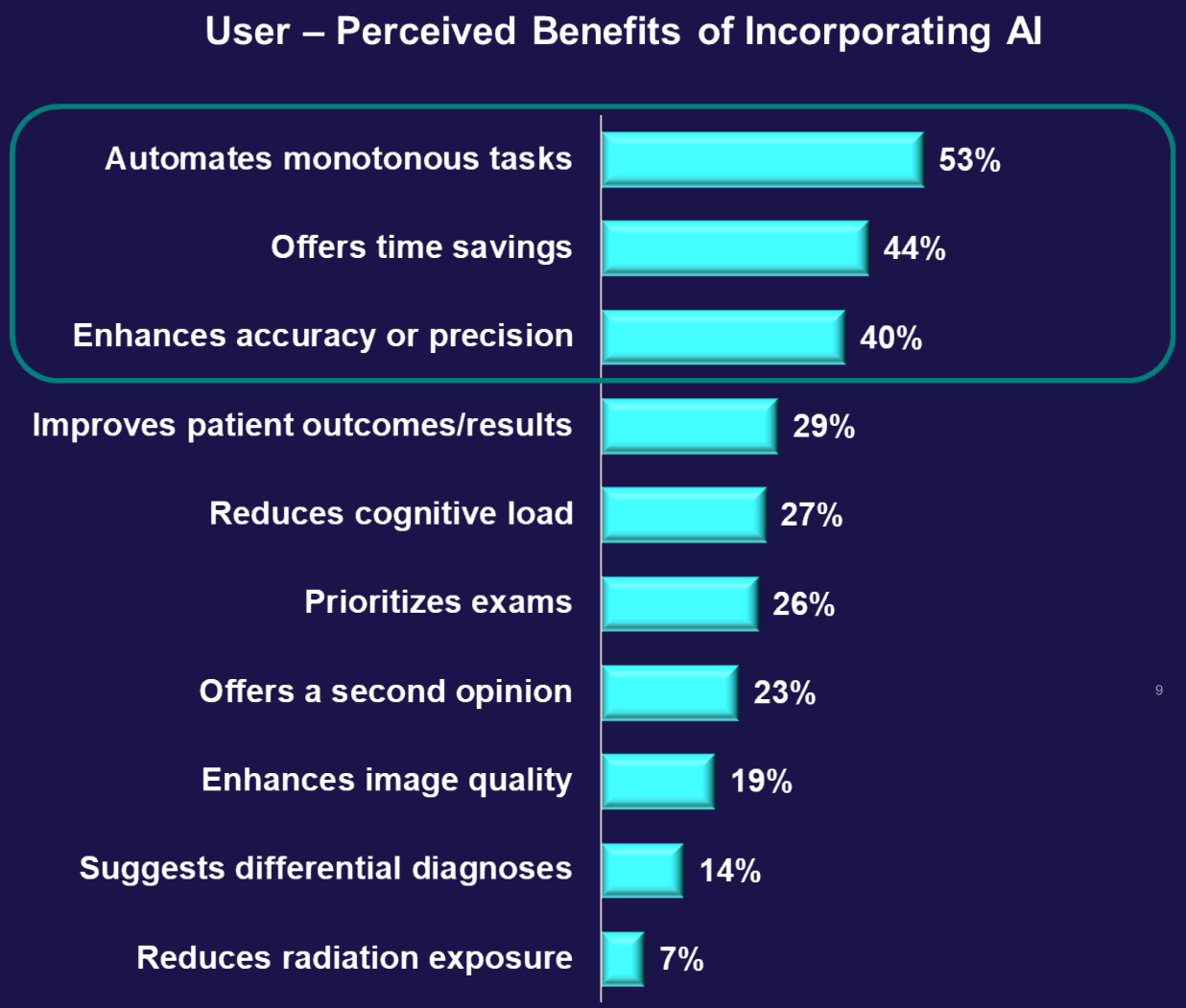
AI Adoption in Radiology Continues to Grow



DSI AI Tracking Survey (ACR publication forthcoming)

- Sample of nearly 1,000 (n=922) Members, representative of the population: age/tenure, practice type, geography, gender, specialty
- Data Collected in Fall/Winter 2024/2025 with margin of error (+/-) of 3.2% @ 95% confidence level

Most appealing Attributes of AI: Address volume and efficiency challenges allowing radiologists to focus on higher-impact activities



The primary value of AI is its role in mitigating the radiology volume crisis – automating repetitive tasks (53%) and saving time (44%) directly address systemic work pressures.

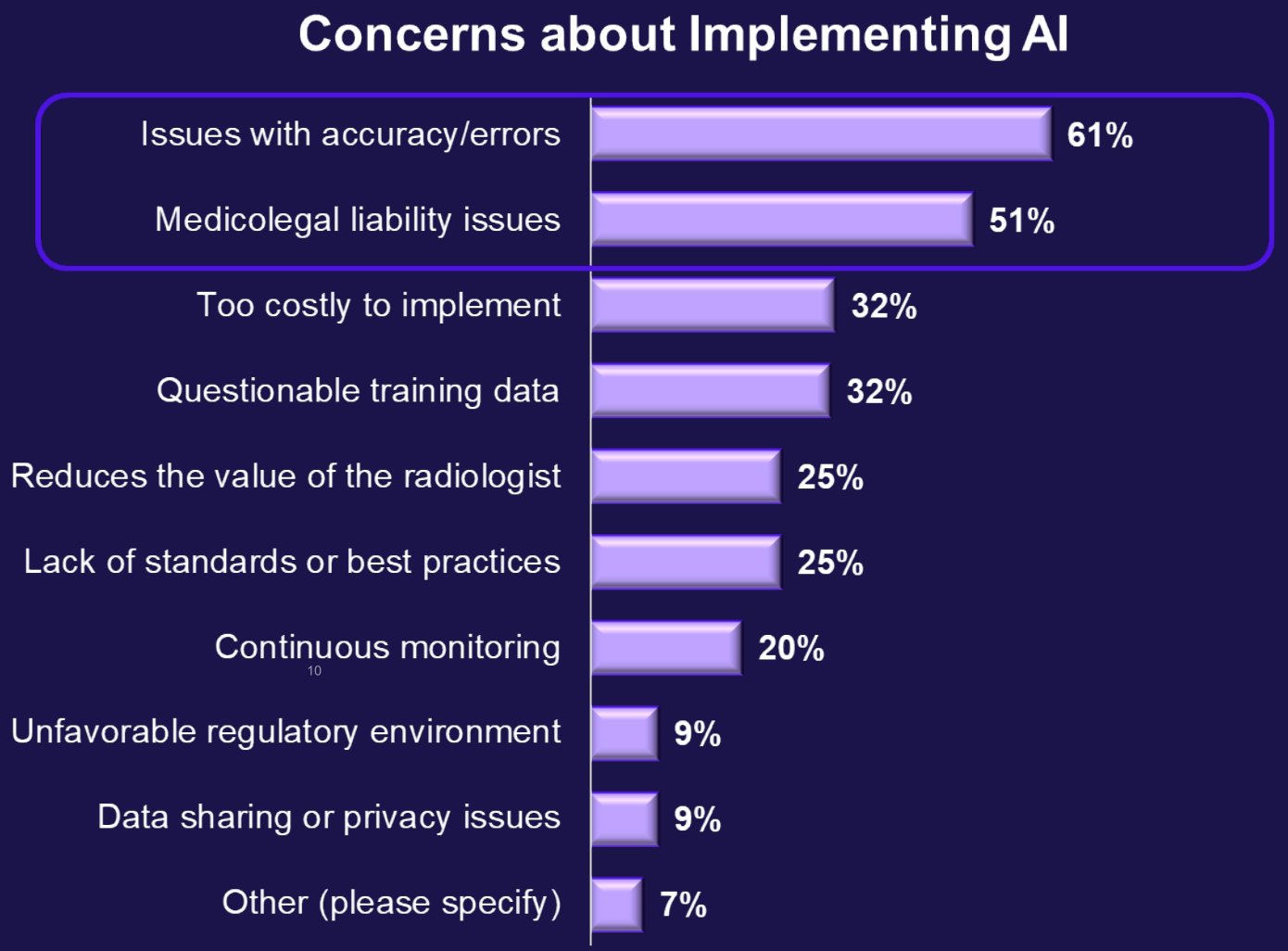
d

9

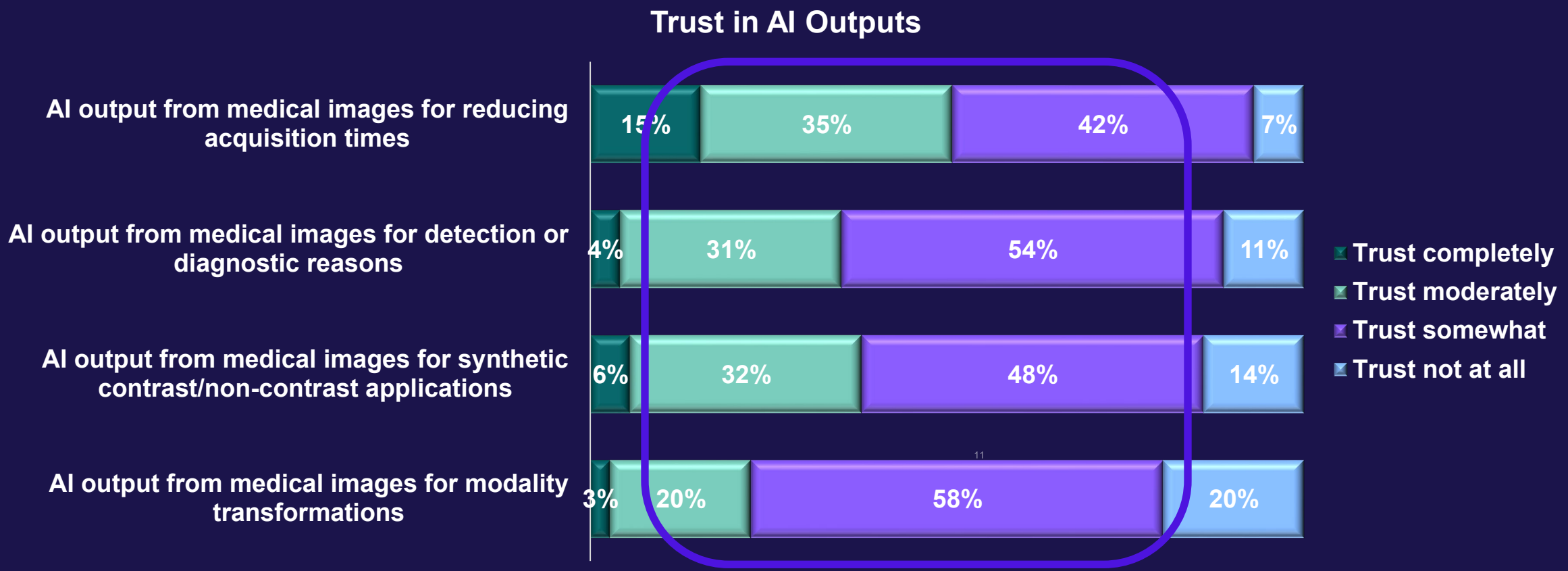
2024 DSI Tracking Study

Accuracy and Medicolegal liability lead AI Implementation concerns

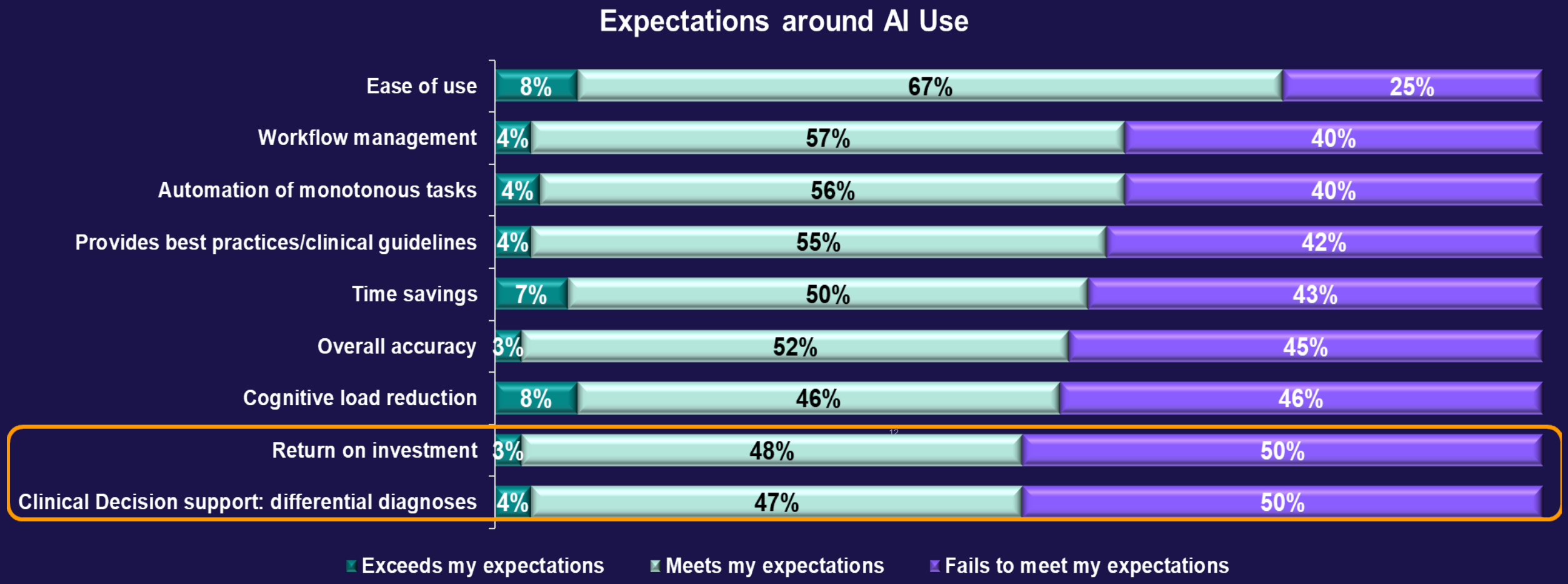
- Accuracy and liability concerns highlight the need for solutions that:
  - Ensure that AI/ML is accurate and in compliance with safety protocols
  - Provide risk mitigation strategies for those practices adopting AI
  - Highlight those products and services that can be trusted and mitigate medicolegal liability



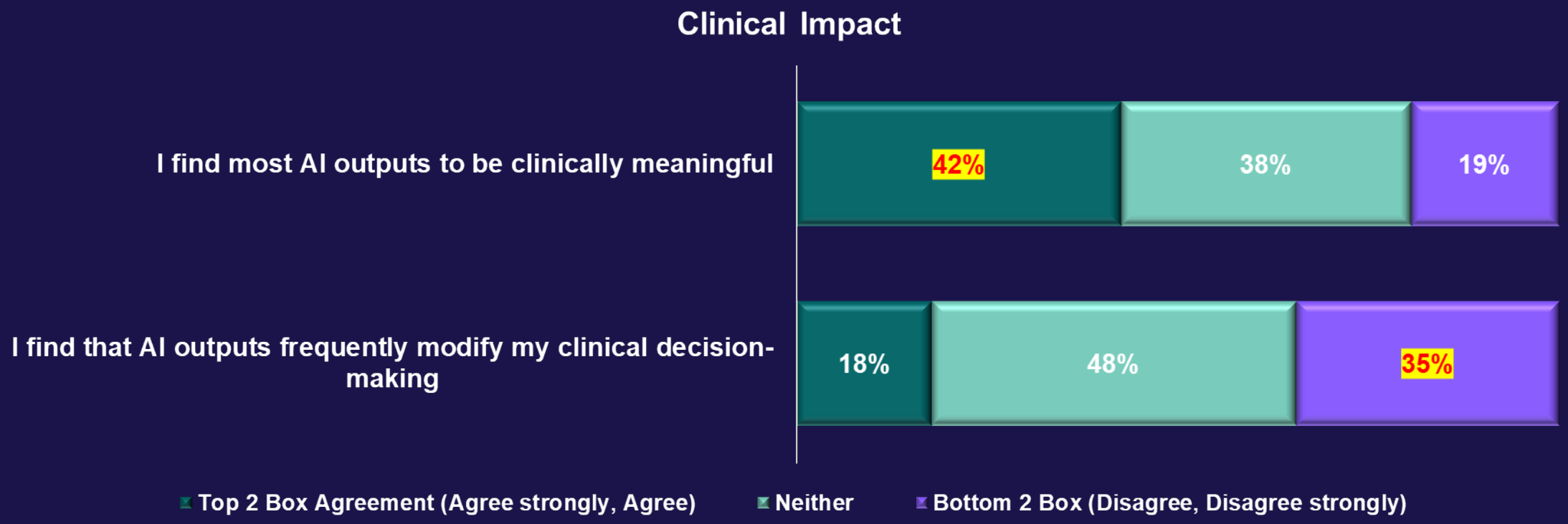
Most Trust AI outputs *moderately or somewhat*



AI Delivers on core expectations; further development needed on ROI and Decision Support to encourage and expand use



AI Demonstrates *clinical value;* however, shows limited current influence on *clinical decision-making*



2024 DSI Tracking Study

Adoption and the AI Marketplace

Technology Diffusion and the AI the Market

Navigating the Early Hype

- Identifying the gaps and Research Imperatives

ARTICLE IN PRESS

ORIGINAL ARTICLE

A Road Map for Translational Research on Artificial Intelligence in Medical Imaging: From the 2018 National Institutes of Health/RSNA/ACR/The Academy Workshop

Bibb Allen Jr, MD[a], Steven E. Seltzer, MD[b,c], Curtis P. Langlotz, MD, PhD[d], Keith P. Dreyer, DO, PhD[e], Ronald M. Summers, MD, PhD[f], Nicholas Petrick, PhD[g], Danica Marinac-Dabic, MD, PhD, MMSC[h], Marisa Cruz, MD[i], Tarik K. Alkasab, MD, PhD[e], Robert J. Hanisch, PhD[j], Wendy J. Nilsen, PhD[k], Judy Burleson, BSW, MHSA[l], Kevin Lyman, BS[m], Krishna Kandarpa, MD, PhD[n]

Abstract

Advances in machine learning in medical imaging are occurring at a rapid pace in research laboratories both at academic institutions and in industry. Important artificial intelligence (AI) tools for diagnostic imaging include algorithms for disease detection and classification, image optimization, radiation reduction, and workflow enhancement. Although advances in foundational research are occurring rapidly, translation to routine clinical practice has been slower. In August 2018, the National Institutes of Health assembled multiple relevant stakeholders at a public meeting to discuss the current state of knowledge, infrastructure gaps, and challenges to wider implementation. The conclusions of that meeting are summarized in two publications that identify and prioritize initiatives to accelerate foundational and

Navigating the Early Hype

- Identifying the gaps and Research Imperatives
- Promoting Standards and Interoperability

August 29, 2018

ACR DSI Urges AI Standardization, Interoperability and Reportability at NIH Workshop

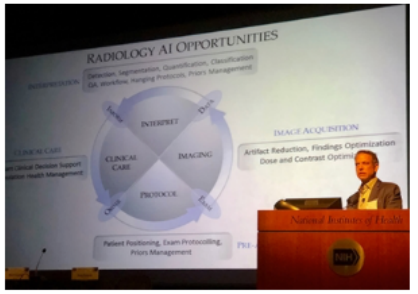
Share Recommend Bookmark

American College of Radiology Data Science Institute™ (ACR DSI) Chief Medical Officer Bibb Allen, MD, FACR, co-chaired the recent National Institute of Biomedical Imaging and Bioengineering (NIBIB) Workshop on Artificial Intelligence (AI) in Medical Imaging. Proceedings from the workshop will be published as a research roadmap for health care and scientific professionals.

ACR DSI Chief Medical Officer Bibb Allen, MD, FACR

ACR DSI presenters stressed the need for standardized AI use cases to not only ensure that the relevant clinical questions are answered, but to promote interoperability and reportability for ongoing quality improvement — including that:

- The AI market is dependent upon both the development of AI algorithms and integration with current digital solutions (PACS, reporting systems, etc.)
- The DSI is working with various clinical experts, academic radiology departments and other radiology professional organizations to create standardized use cases for developing AI algorithms that will solve the most important clinical problems radiologists encounter
- Compared to algorithms developed by industry developers or at single academic institutions, AI solutions that follow DSI AI use case standards will be able to be readily integrated into clinical practice



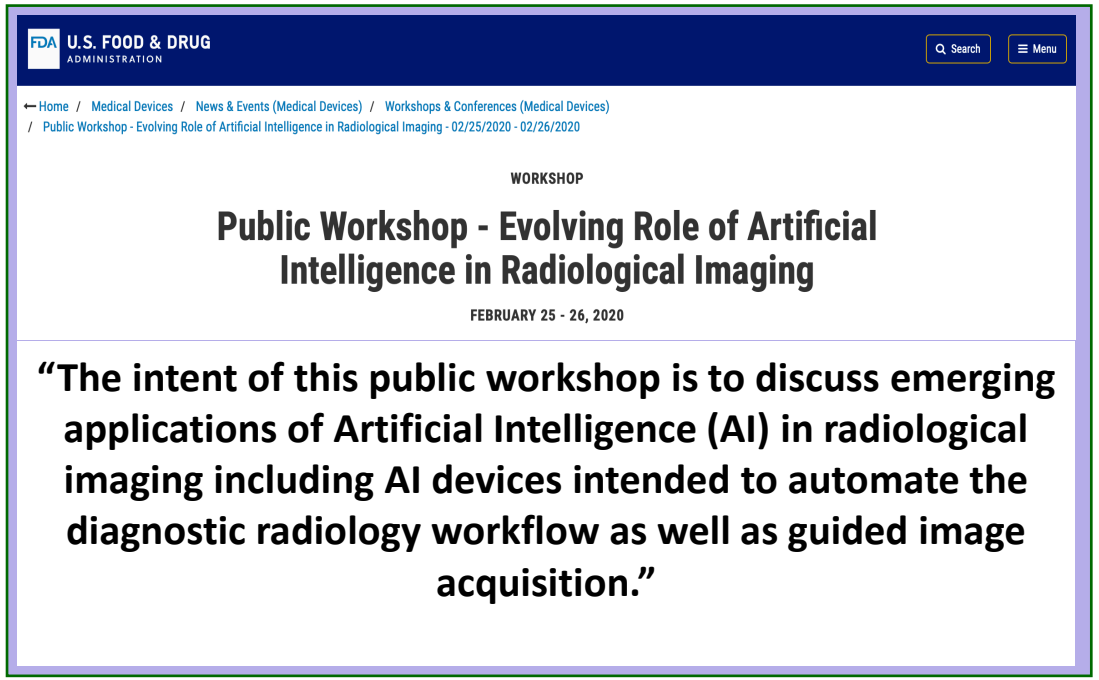
ACR DSI Chief Science Officer Keith Dreyer, DO, PhD, FACR

ACR DSI chief science officer Keith Dreyer, DO, PhD, FACR, specifically outlined the Lung-RADS® Assist: Advanced Radiology Guidance, Reporting and Monitoring use case that was recently chosen as a pilot project by the FDA-funded National Evaluation System for Health Technology Coordinating Center (NESTcc). Dreyer outlined that the Lung-RADS® standardization provided a basis for this standardized use case that will:

- Utilize existing ACR resources with data from multiple institutions to demonstrate the ability to validate algorithm performance prior to FDA clearance
- Facilitate interoperability between reporting and AI developers to generate standardized data in a real-world setting
- Capture real-world data in a national registry to monitor the performance of AI algorithms in clinical practice and enable both facility-level and developer reporting to ensure the algorithm performs as expected in clinical practice

Navigating the Early Hype

- Identifying the gaps and Research Imperatives
- Promoting Standards and Interoperability
- Working with Regulators on the Limits of AI (e.g., Autonomous AI)



Navigating the Early Hype

- Identifying the gaps and Research Imperatives
- Promoting Standards and Interoperability
- Working with Regulators on the Limits of AI (e.g., Autonomous AI)
- Educating and Convening

2018 Economics of Artificial Intelligence

2019 Data Access in Healthcare – Implications for the Artificial Intelligence Ecosystem

2020 Monitoring and Evaluating AI: Challenges and Practical Implications

2021 AI in Clinical Practice: Choosing the Right Tools for your Practice and Patients

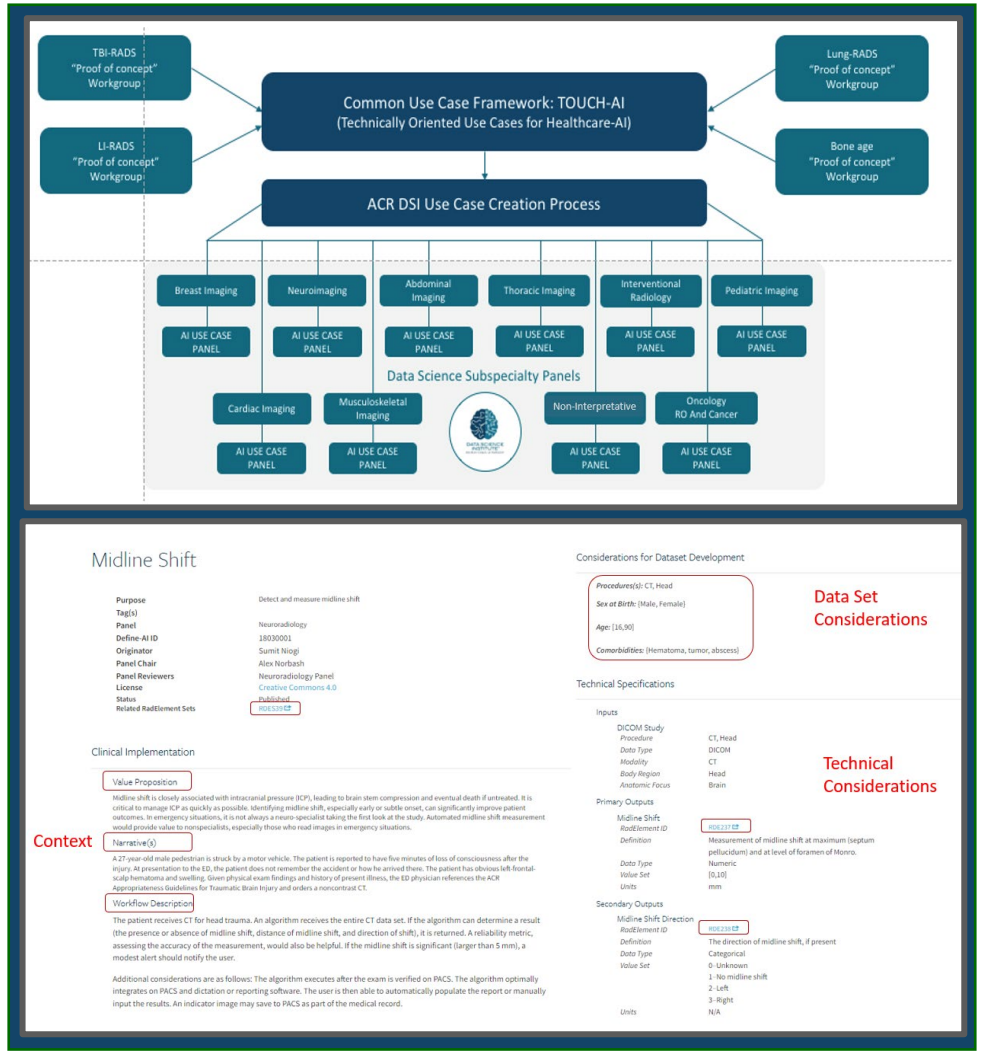
2022 Using AI in Clinical Practice: A Practical Guide for Radiologists

2023 Quantifying the Value of AI: Investment, Costs, Reimbursement and Benefits

2024 From Data to Diagnosis with Tested, Transparent, and Trustworthy AI

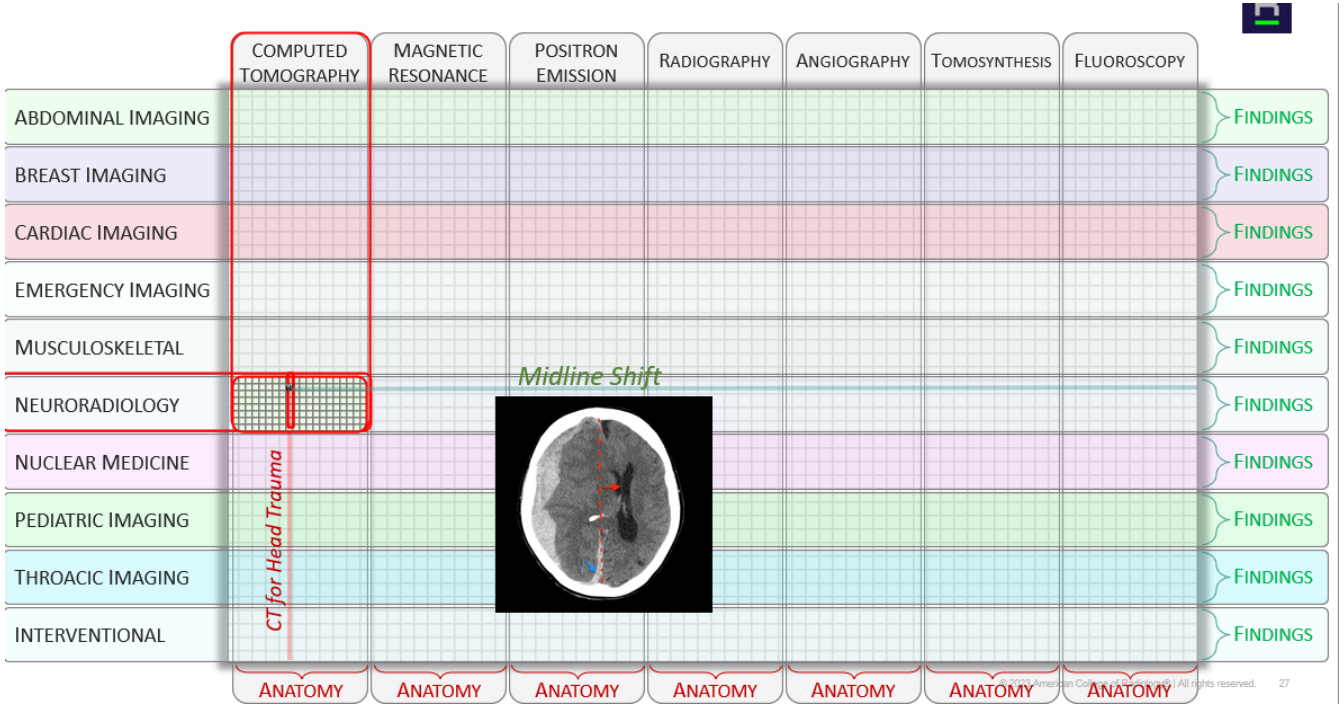
Navigating the Early Hype

- Identifying the gaps and Research Imperatives
- Promoting Standards and Interoperability
- Working with Regulators on the Limits of AI (e.g., Autonomous AI)
- Educating and Convening
- Defining Use Cases (DEFINE-AI)



Navigating the Early Hype

- Identifying the gaps and Research Imperatives
- Promoting Standards and Interoperability
- Working with Regulators on the Limits of AI (e.g., Autonomous AI)
- Educating and Convening
- Defining Use Cases (DEFINE-AI)



TOTAL FDA CLEARED SaMD PRODUCTS: 351

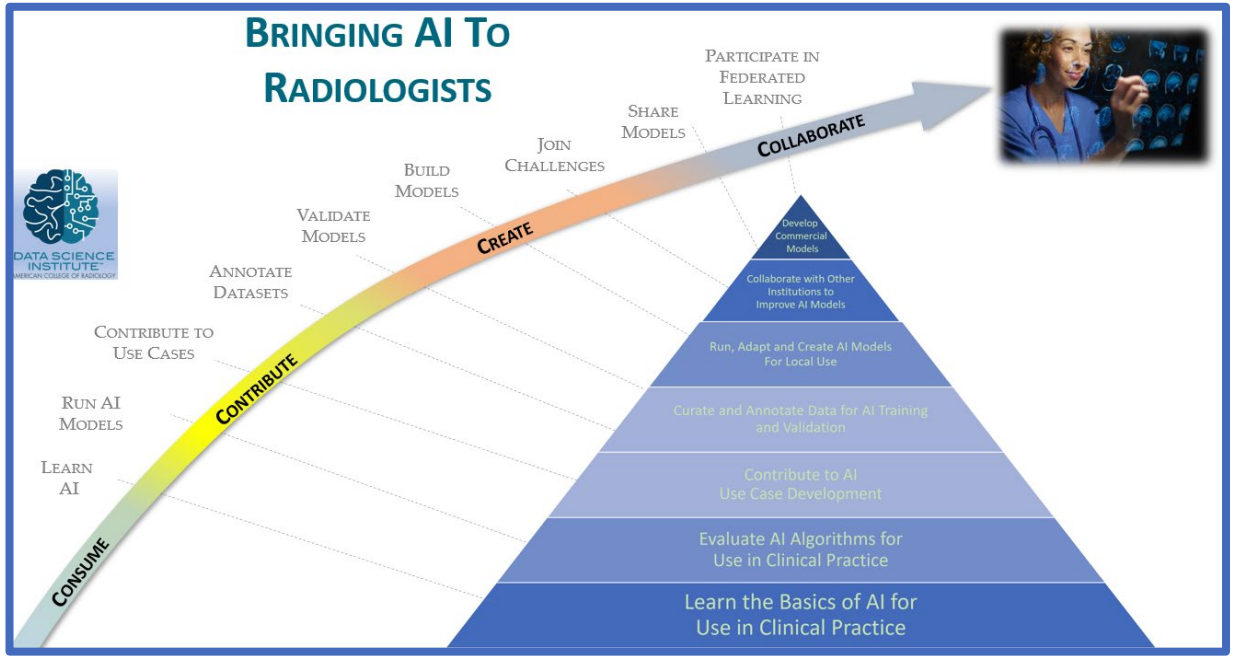
☑ Pediatric
☑ Adult

Manufacturer: (All)

Subspecialty	CT	MR	PET	US	3D-Angio	XRAY/MAM
Abdominal	21	24	1	7	2	3
Cardiac	41	19	2	17	1	6
Chest	47	8	1	6	1	38
Musculoskeletal	19	9	1	3	1	16
Neuroradiology	76	48	4	1	1	15
Breast Imaging	8	6	1	6	1	29
Pediatric	3	7		2		12

Evolving our Understanding of AI

- Democratizing AI



Evolving our Understanding of AI

- Democratizing AI
- Tools to Foster Engagement & Literacy

Evolving our Understanding of AI

- Democratizing AI
- Tools to Foster Engagement & Literacy
- Research and Data Collection

500+ Clinical Projects Conducted

100+ Active Clinical` Projects

150+ Clinical Research Staff

20,000+ Sites Using TRIAD

50+ Years of Research

TMIST Breast Cancer Screening Trial
ECOG-ACRIN

CANDID-4-AI
CANcer Data and Image Database for AI

MIDRC
MEDICAL IMAGING AND DATA RESOURCE CENTER

NIH
National Institute of Biomedical Imaging and Bioengineering

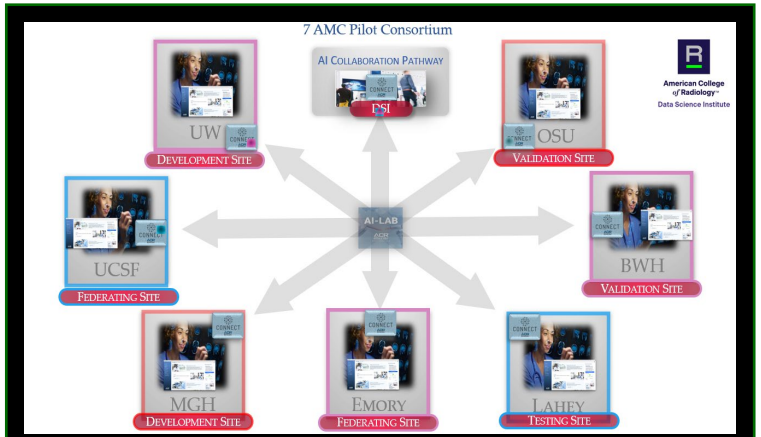
American College of Radiology

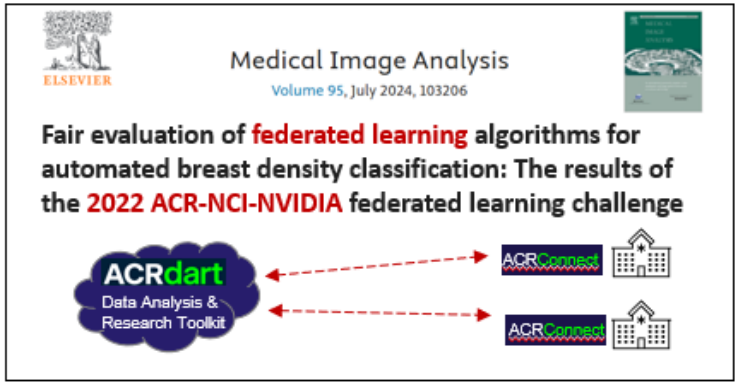
AAPM

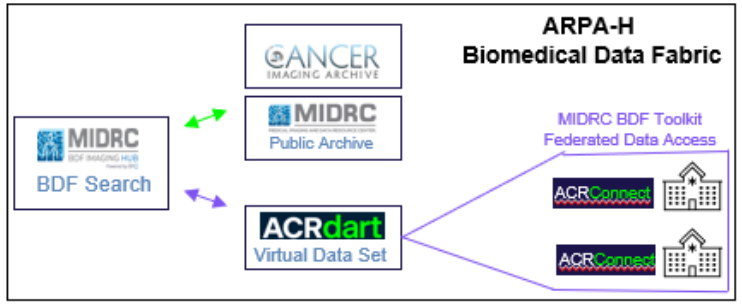
RSNA

Evolving our Understanding of AI

- Democratizing AI
- Tools to Foster Engagement & Literacy
- Research and Data Collection
- Federated Data Access and Learning

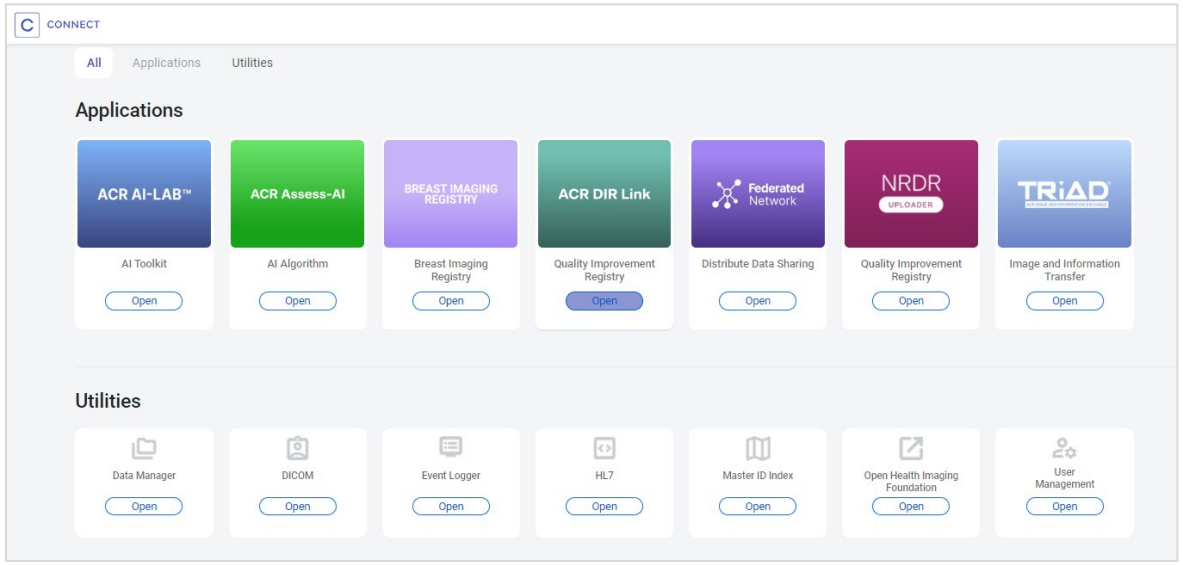


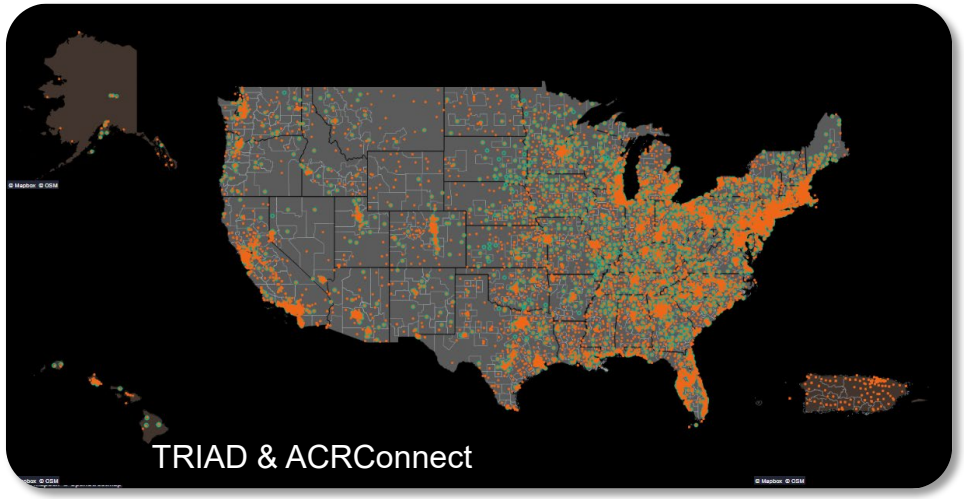




Evolving our Understanding of AI

- Democratizing AI
- Tools to Foster Engagement & Literacy
- Research and Data Collection
- Federated Data Access and Learning





Evolving our Understanding of AI

- Democratizing AI
- Tools to Foster Engagement & Literacy
- Research and Data Collection
- Federated Data Access and Learning
- Facilitating the Discussion with Regulators, Payers, and Industry

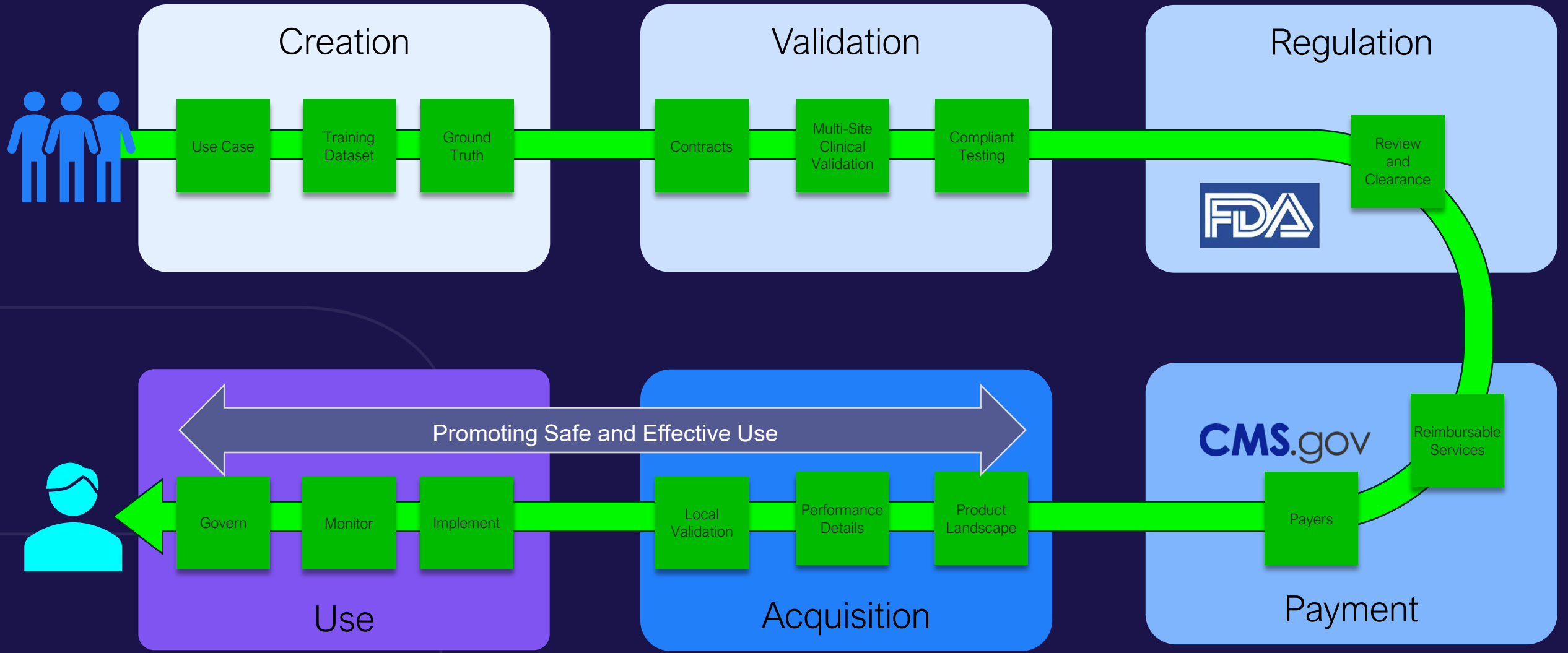
Evolving our Understanding of AI

- Democratizing AI
- Tools to Foster Engagement & Literacy
- Research and Data Collection
- Federated Data Access and Learning
- Facilitating the Discussion with Regulators, Payers, and Industry
- Real-world Testing

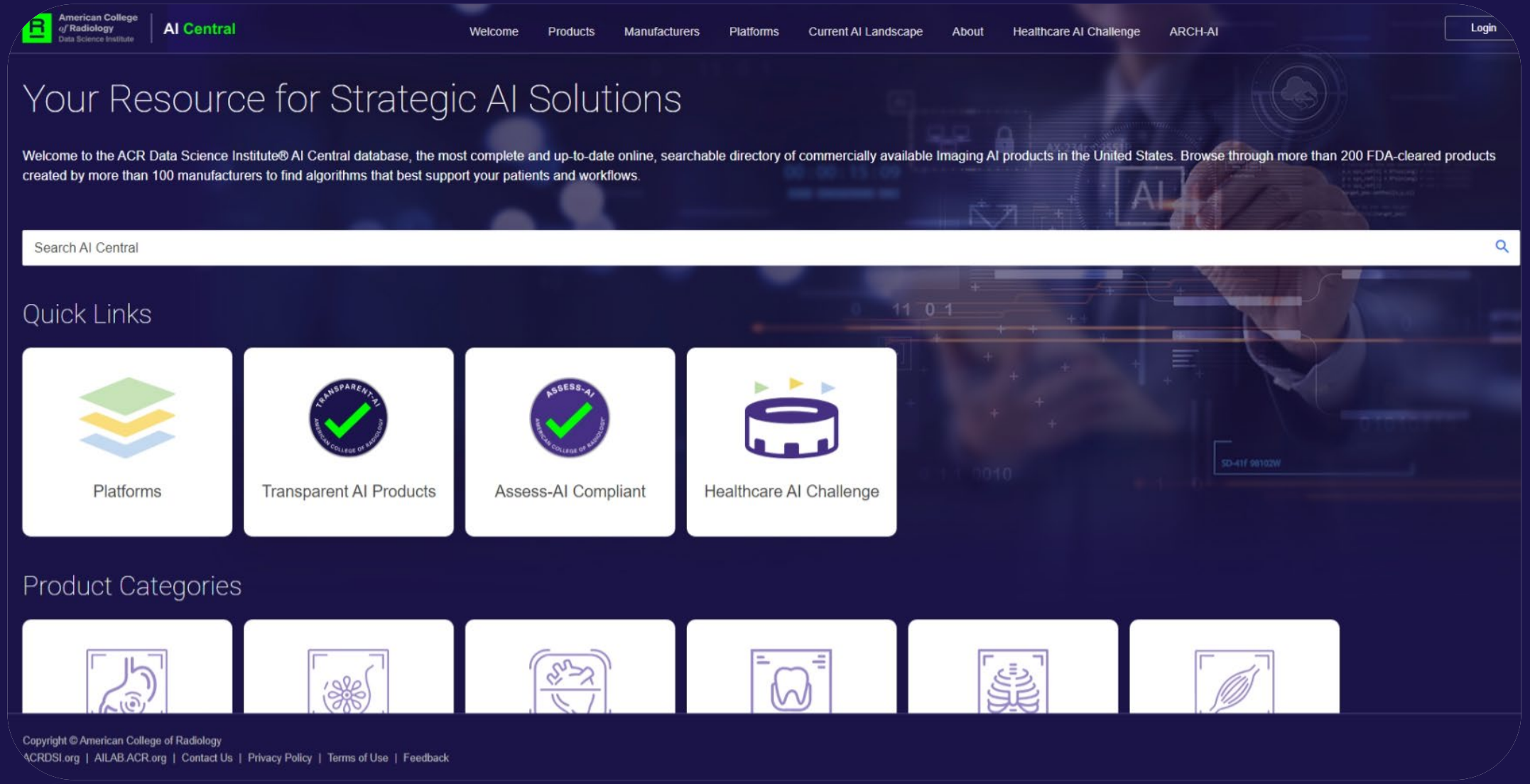
Scaling Adoption

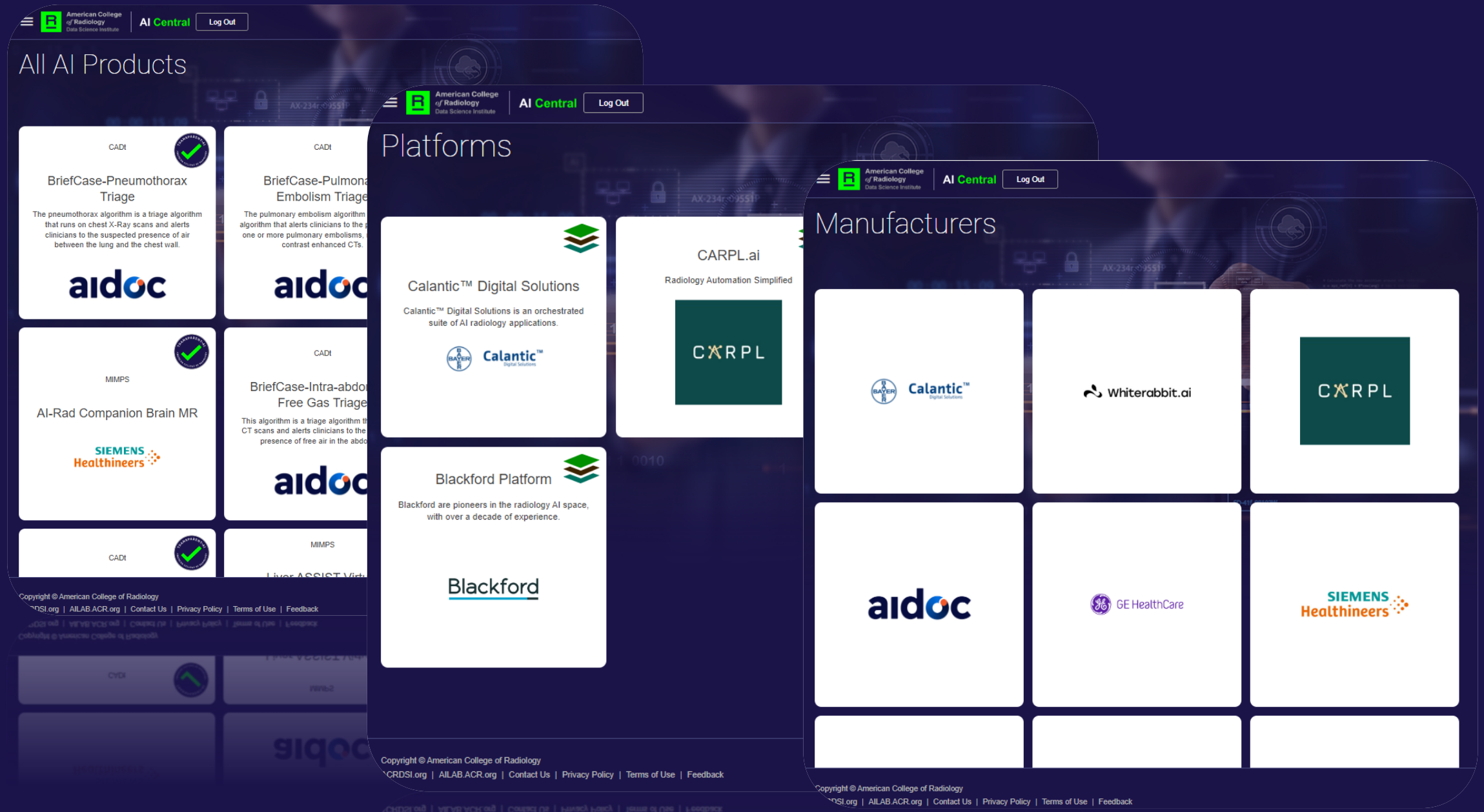
A framework to Promote Safe and Effective Use of Clinical AI

AI Life Cycle

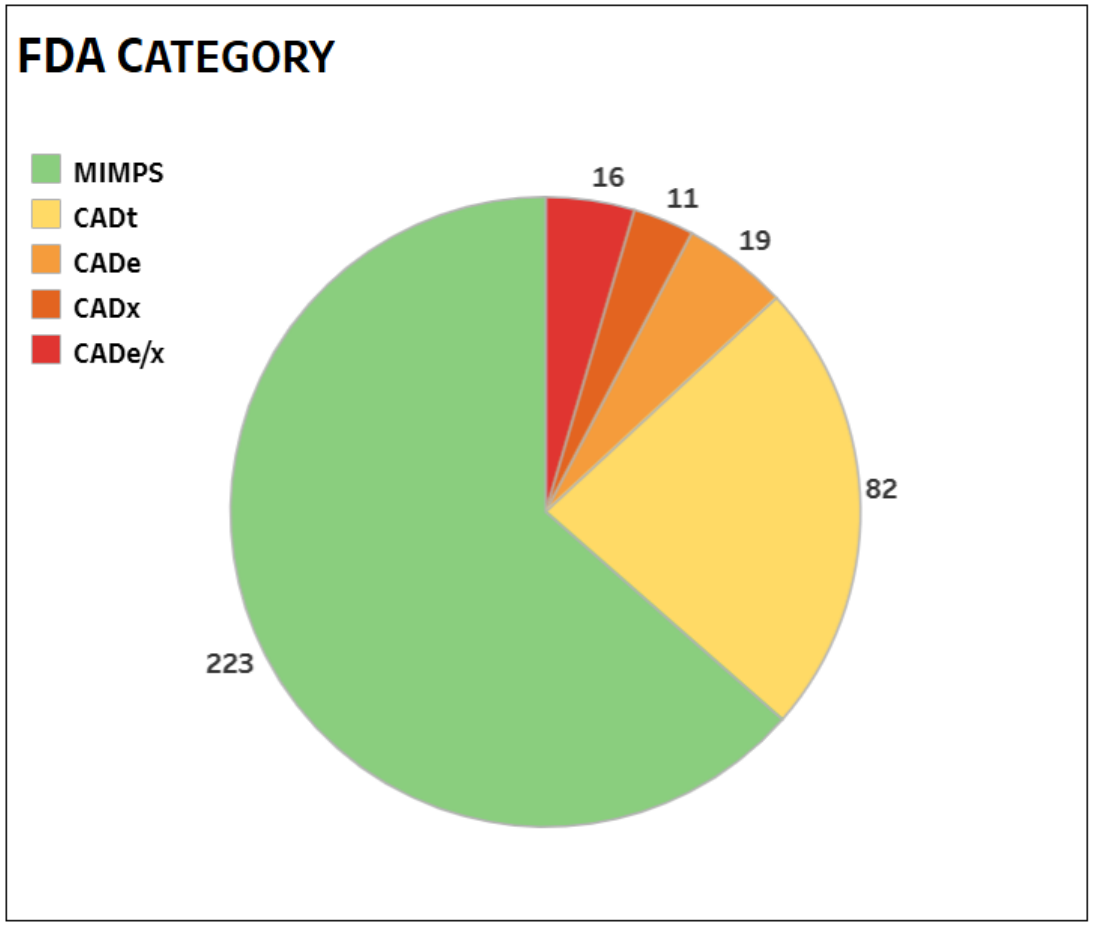


AI Central





SaMD Product Drilldown



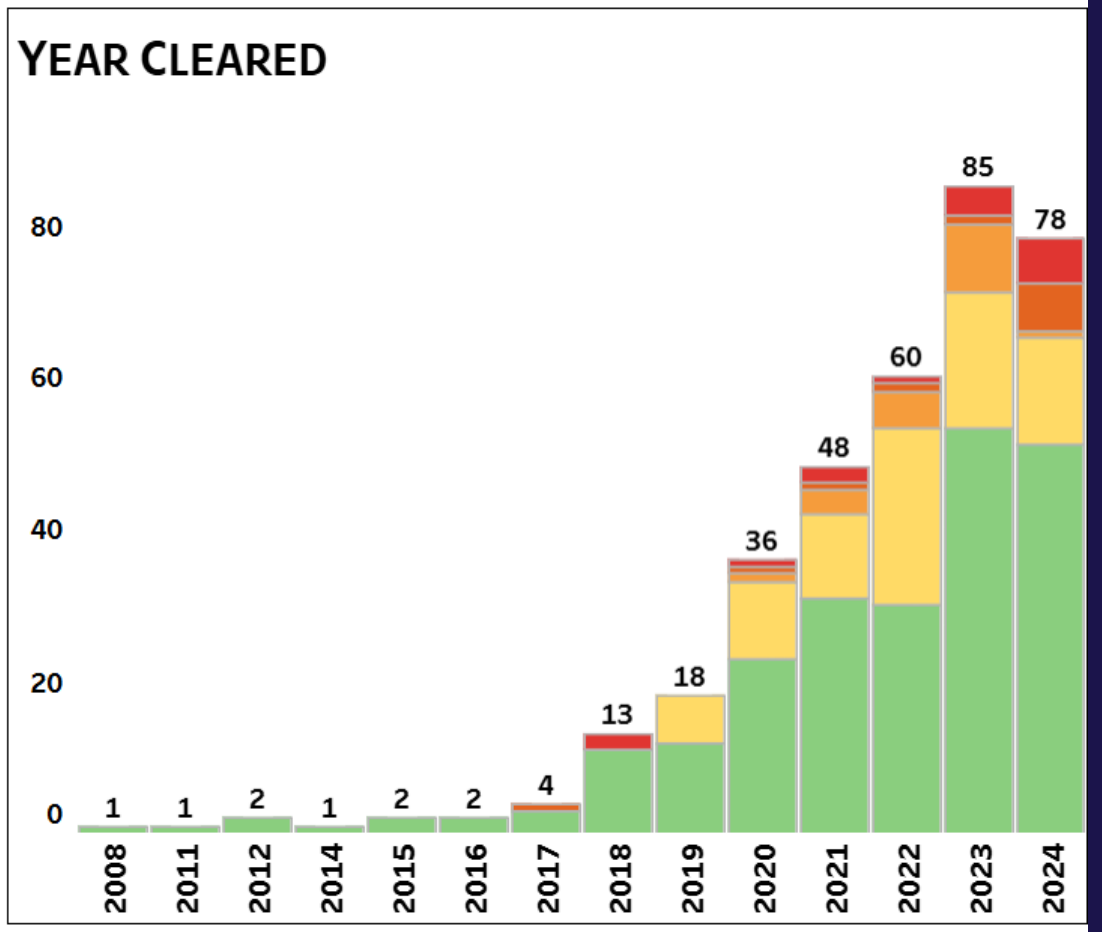


Image Processing And Analysis Tools MIMPS

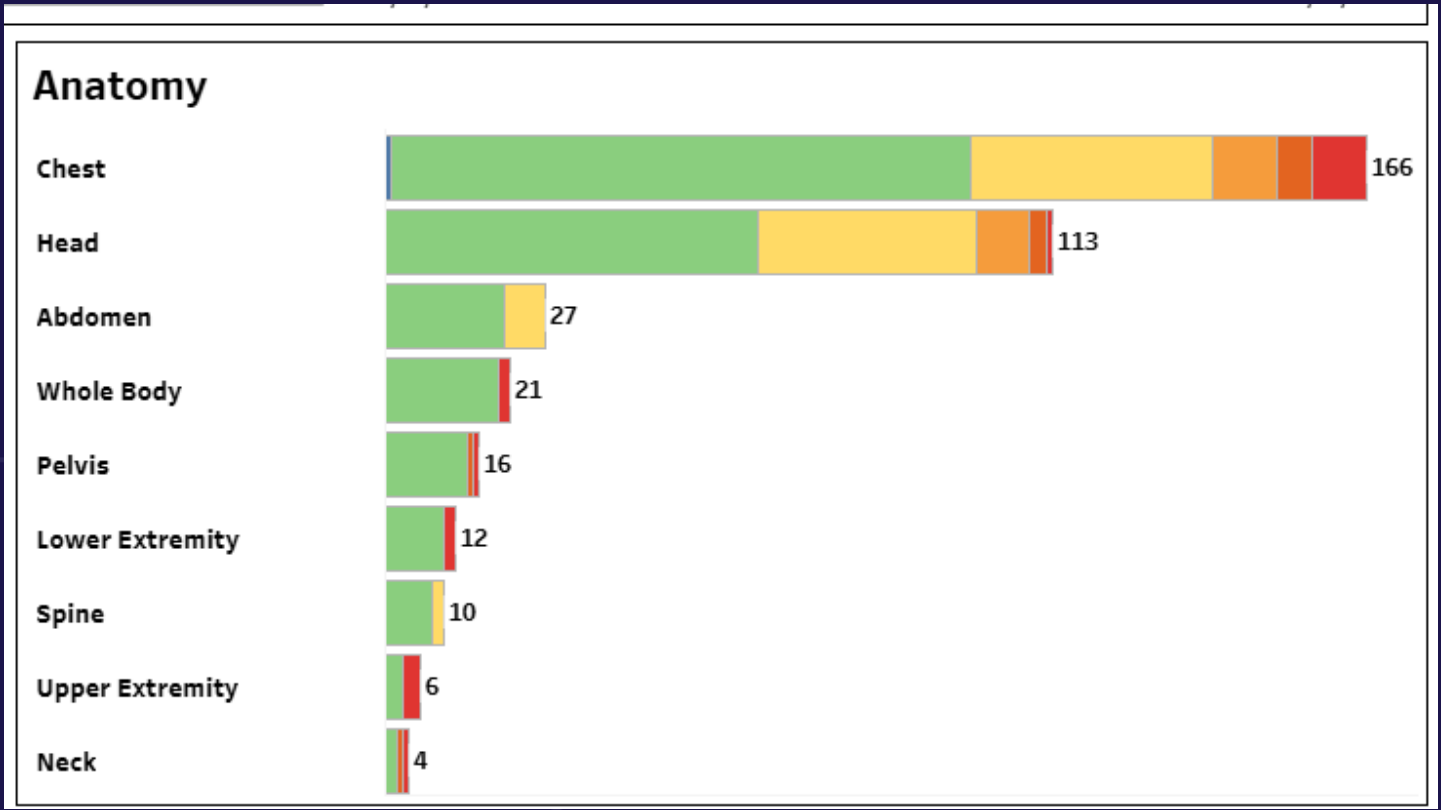
Computer Aided Triage CADt

Computer Aided Detection CADe

Computer Aided Diagnosis CADx

Computer Aided Detection & Diagnosis CADe/x

SaMD Product Drilldown



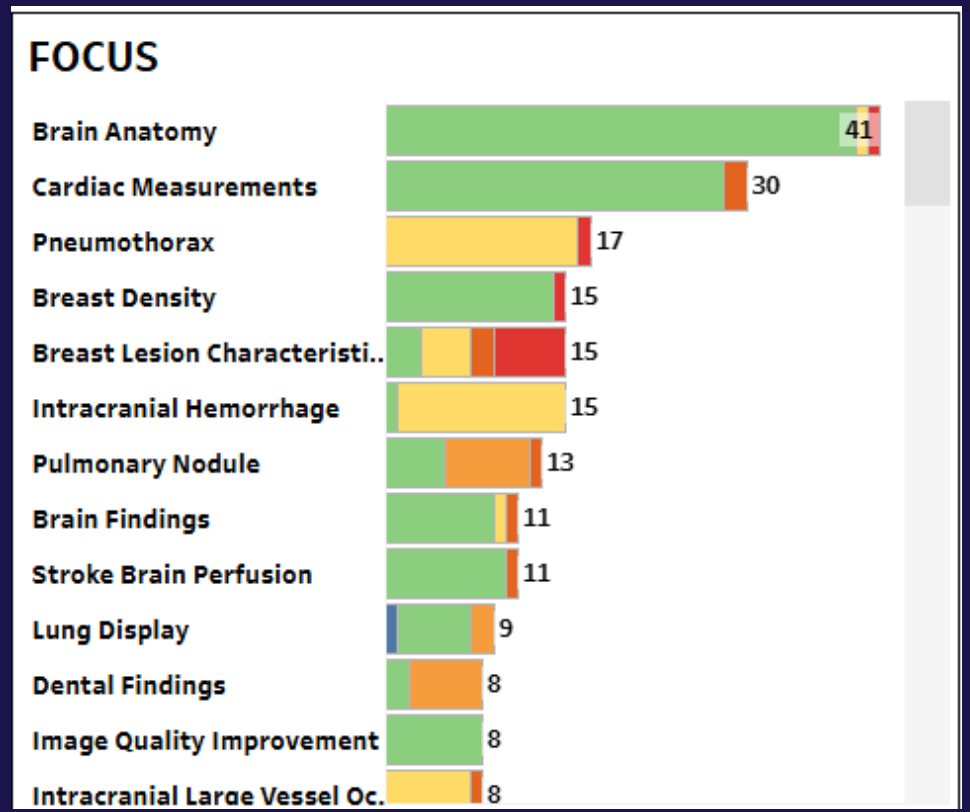


Image Processing And Analysis Tools MIMPS

Computer Aided Triage CADt

Computer Aided Detection CADe

Computer Aided Diagnosis CADx

Computer Aided Detection & Diagnosis CADe/x

Promoting Transparency

Transparent-AI Data Elements

Appendix

Type	Element Name	ID	Description	Required
Model Identifier	Company name	MI.1	Name of company.	Yes
	Product name	MI.2	Name of FDA cleared product.	Yes
	Product version	MI.3	Version of the Algorithm used for inference.	Yes
	Product clearance date	MI.5	Date the product was cleared by FDA.	Yes
	Product FDA submission number	MI.6	Unique FDA ID that is referred as submission number.	Yes
	Additional information	MI.7	Free text field.	No
Model Type	Product category	MT.1	Indicate CAD type the product is cleared for.	Yes
	Additional information	MT.2	Free text field.	No
Model Characteristics	Data inclusion criteria	MC.1	Indicate the inclusion criteria for the product to run and provide output.	Yes
	Data exclusion criteria	MC.2	Indicate the exclusion criteria where the product will not run.	Yes
	Instructions for use	MC.3	Indicate if your company will be able to provide instructions for use document.	Yes
	Additional information	MC.4	Free text field.	No
Indications for Use	Clinical output	IU.1	What is the finding(s) that the product is intended to output	Yes
	Intended user	IU.2	Indicate the intended user of the product.	Yes
	Patient age	IU.3	Age group the product cleared is cleared for.	Yes
	Body area	IU.4	What target body part is the product intended for use.	Yes
	Modality	IU.5	Target modality for the product.	Yes
	Additional information	IU.6	Free text field.	No
Model Performance (performance testing and/or reader study)	Study type	MP.1	Indicate Reader Study or Stand-Alone Performance.	No
	Reference standard (ground truth)	MP.2	What was the reference standard (ground truth) based on?	No
	Number of readers	MP.3	If reader study was conducted, how many readers read each case?	No
	Number of cases	MP.4	How many image studies were used to evaluate model performance?	No
	Number of positive and negative cases	MP.5	Indicate number of positive and negative cases.	No
	Age range	MP.6	Indicate age range for data set.	No
	Sex ratio	MP.7	Indicate sex ratio of data set.	Yes
	Race/ethnicity ratio	MP.8	Indicate the race/ethnicity breakdown of data set.	Yes
	Geographic breakdown	MP.9	Indicate the geographic breakdown of data set.	Yes
	Manufacturer list	MP.10	Indicate the manufacturers in data set.	Yes
	Scanner list	MP.11	Indicate the scanners in data set.	Yes
	Number of sites	MP.12	How many sites were used.	No
	Percent of cases with and without finding	MP.13	Indicate the percent of positive and negative cases.	No
	Model accuracy	MP.14	Indicate the model accuracy.	No
	Model sensitivity	MP.15	Indicate the model sensitivity.	Yes
	Model specificity	MP.16	Indicate the model specificity.	Yes
	Additional information	MP.17	Free text field.	No
Details on Training Data Sets	Model dataset — Number of cases	TD.1	Indicate number of cases the model was trained on.	No
	Age range the model was trained on as well as age distribution	TD.2	Indicate age range the model was trained on as well as the age distribution.	No
	Sex ratio model was trained on	TD.3	Indicate sex ratio the model was trained on.	No
	Positive and negative cases	TD.4	Indicate number of positive and negative cases the model was trained on.	No
	Race/ethnicity ratio	TD.5	Indicate the race/ethnicity breakdown the model was trained on.	No
	Geographic breakdown	TD.6	Indicate the geographic breakdown the model was trained on.	No
	Manufacturers	TD.7	Indicate the manufacturers the model was trained on.	No
	Scanners	TD.8	Indicate the scanners that the model was trained on.	No
	Additional information	TD.9	Free text field.	No
Model Limitations, Warnings and Precautions	Supported scanners/ manufacturer	ML.1	Indicate if algorithm will only work on specific scanner or manufacturer.	No

Assess-AI

Assess-AI

Monitoring performance in clinical practice by capturing real clinical data.

First national registry for monitoring AI Results

Assess-AI National Registry

RSNA | ACR RadElement

RIS

AI Vendor

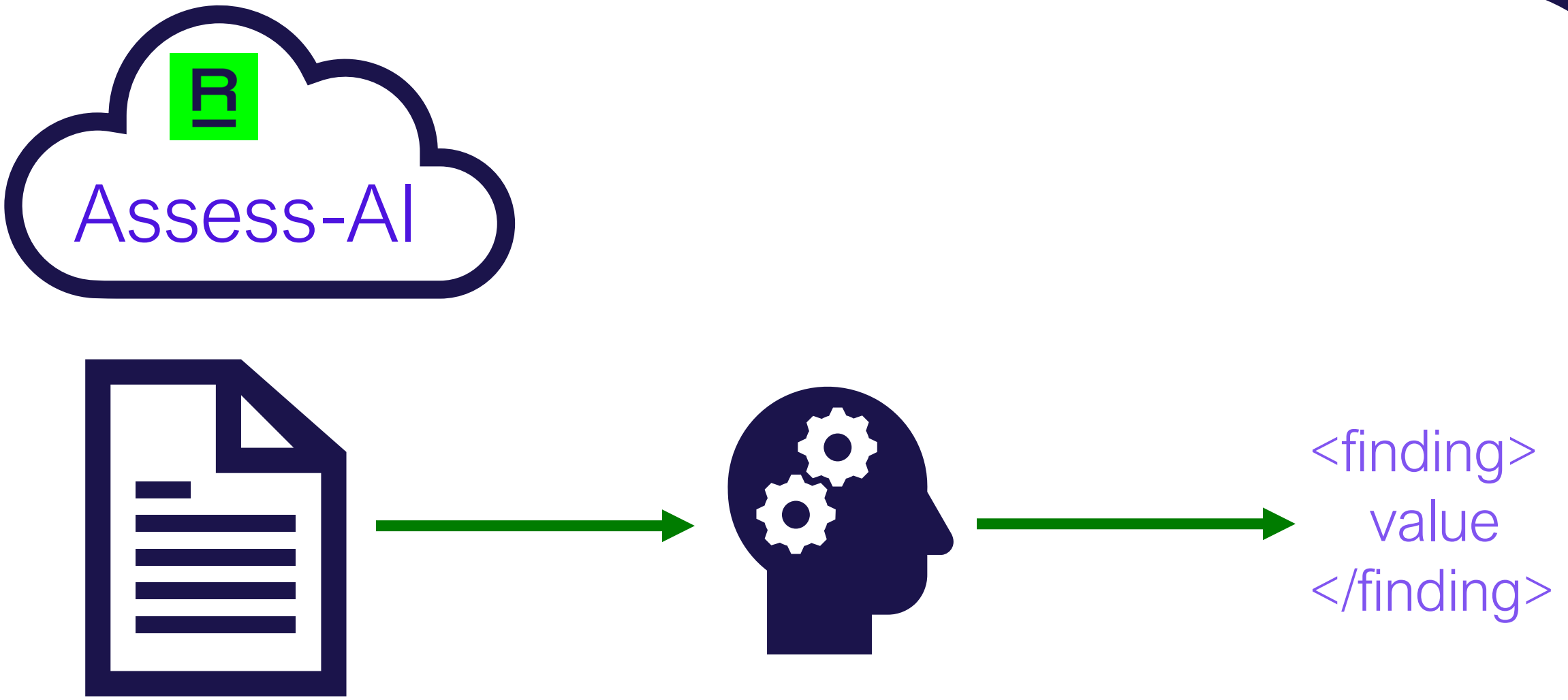
PACS

CONNECT

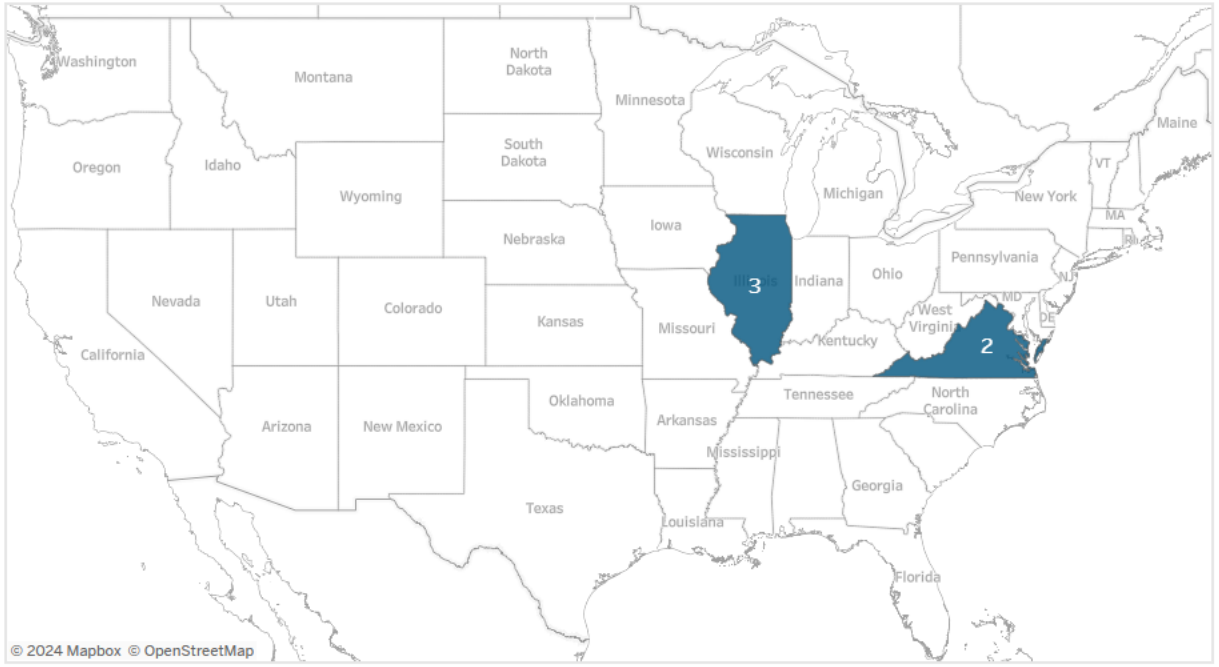
Facility

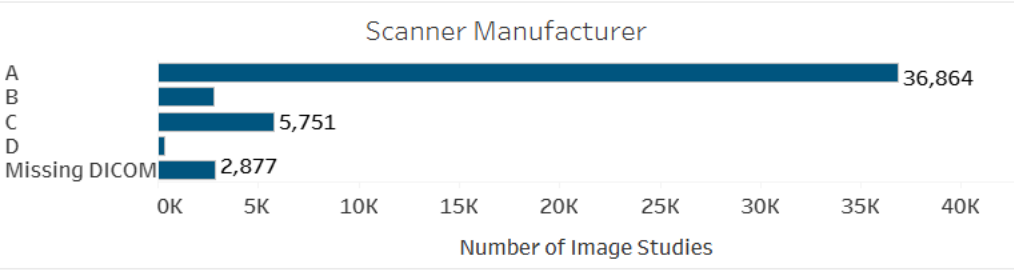
HL7 FHIR

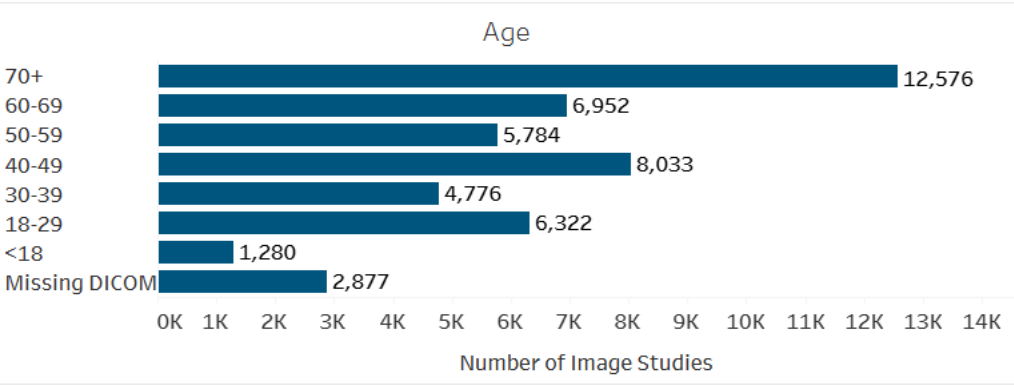
Assess-AI

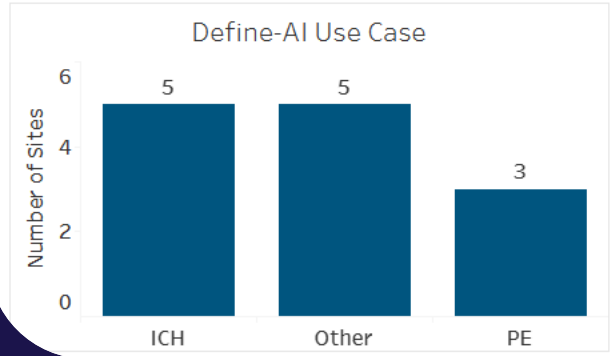


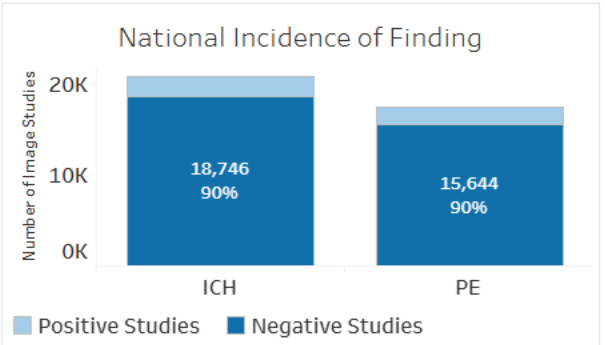
National Assess-AI Registry Dashboard

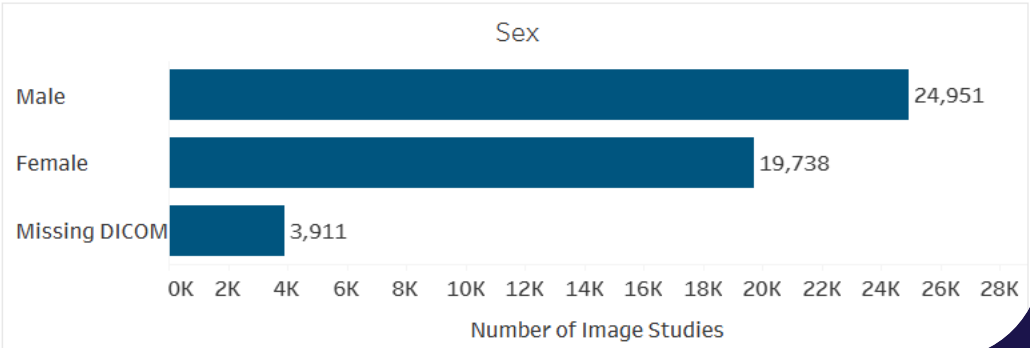


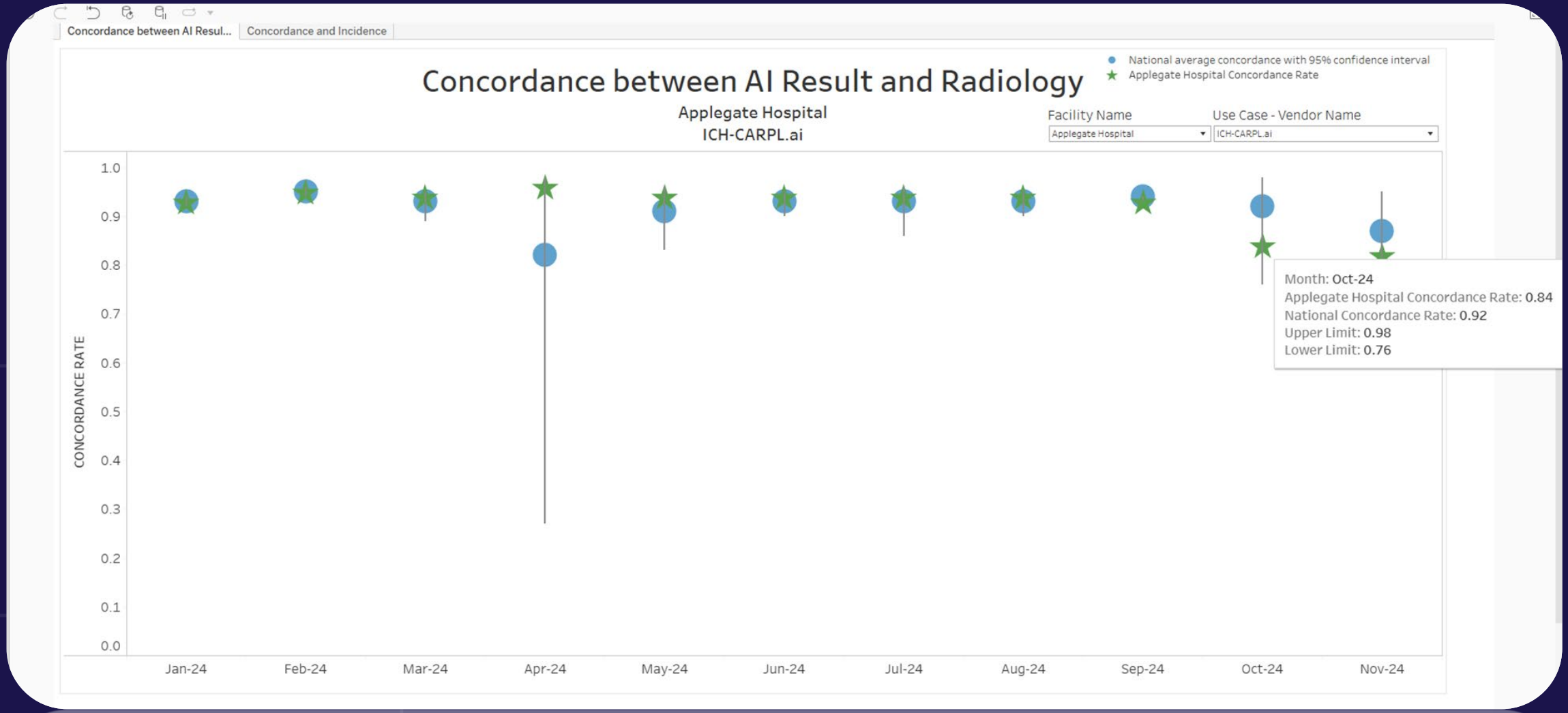












Concordance between AI Resul... | Concordance and Incidence

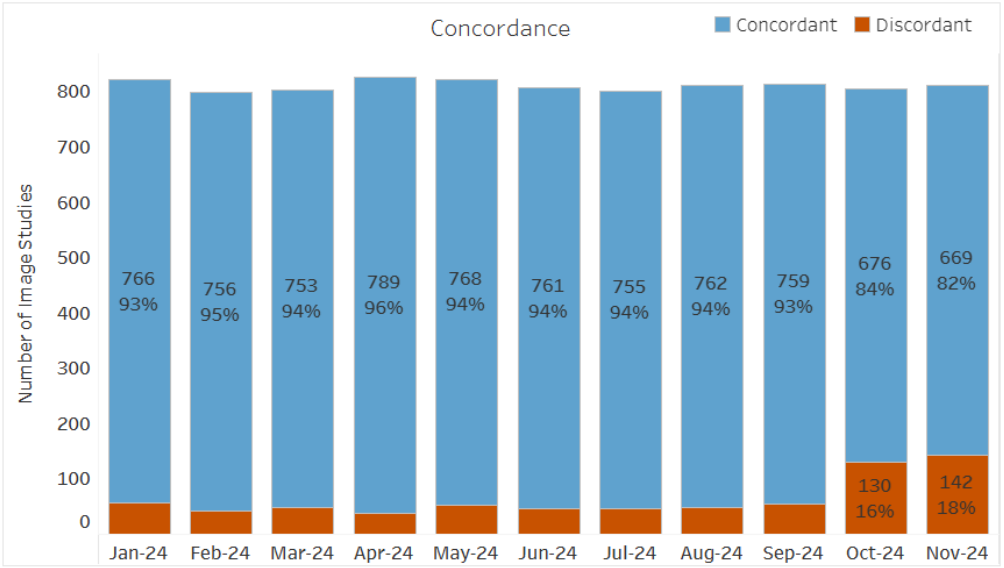
Concordance and Incidence

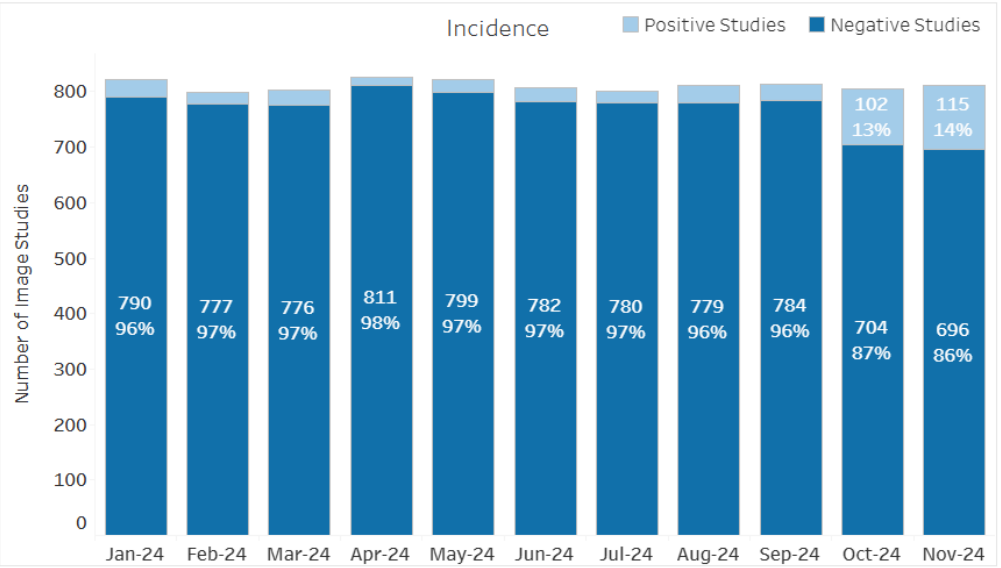
Applegate Hospital

ICH - CARPL.ai

Facility Name: Applegate Hospital

Select Use Case: ICH-CARPL.ai





Accession #	Patient Id	Physician Id	Study Date	Scanner Manufacturer	Scanner Station Name	Scanner Software Version	Age	Sex	Concordance Flag
999955076547	KEKC2GAP6E4T9NT	J0uwdLqD	2024-03-17	GE Healthcare	-	DM_PLATFORM_XRAY_DUCKWEED_4.8.3_2020_FW41.2_65	78	F	Concordant
999881749735	HBWKPW8O31TMHM0	6C0xJLCK	2024-09-14	GE Healthcare	-	DM_PLATFORM_XRAY_DUCKWEED_4.8.3_2020_FW41.2_65	94	M	Discordant
999781316916	PBWLMJI1B9Y6EEV	L1nsbp94	2024-09-02	GE Healthcare	-	DM_PLATFORM_XRAY_DUCKWEED_4.8.3_2020_FW41.2_65	27	M	Concordant
999758289085	-	qWeP98da		-	-	-	-	-	Concordant
999732563414	25ROFTZ4FQUKI5I	CdBQz74o	2024-11-26	GE Healthcare	-	DM_PLATFORM_XRAY_DUCKWEED_4.8.3_2020_FW41.2_65	75	F	Discordant
999448797649	SXKMW0DJZ26JZH8	31y07Ndc	2024-03-15	GE Healthcare	-	DM_PLATFORM_XRAY_DUCKWEED_4.8.3_2020_FW41.2_65	77	M	Concordant
999157174370	F3HUIIJRJ16IC8H2	EBXXEoq7	2024-11-22	GE Healthcare	-	DM_PLATFORM_XRAY_DUCKWEED_4.8.3_2020_FW41.2_65	67	F	Concordant

ARCH-AI

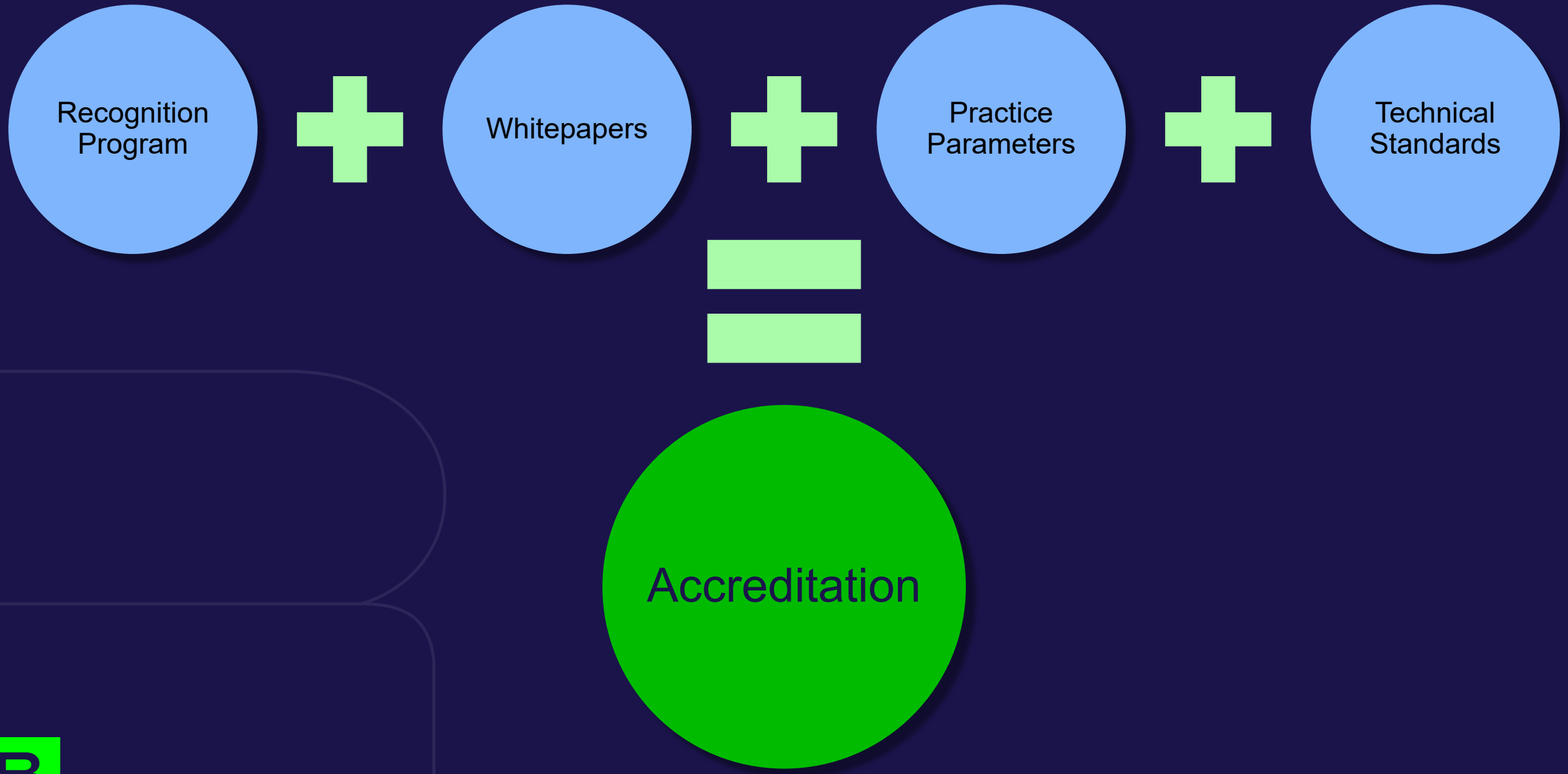
ARCH-AI

First national artificial intelligence quality assurance program for radiology facilities[45]

Designation

Accreditation

Recognition



JACR | Journal of the American College of Radiology™

ACR Login
Submit

Articles Publish Topics About Contact

ORIGINAL ARTICLE · Articles in Press, March 07, 2025

The Road Map for ACR Practice Accreditation for Radiology Artificial Intelligence

David B. Larson, MD, MBA [a] · Mythreyi Bhargavan-Chatfield, PhD [b] · Michael Tilkin, MS [c] · Laura Coombs, PhD [d] · Christoph Wald, MD, PhD, MBA [e]

Affiliations & Notes Article Info

Abstract

As the use of artificial intelligence (AI) continues to grow in radiology, it has become clear that its real-world performance often differs from that demonstrated in premarket testing, underscoring

Criteria

Governance
Selection
Testing

Monitoring

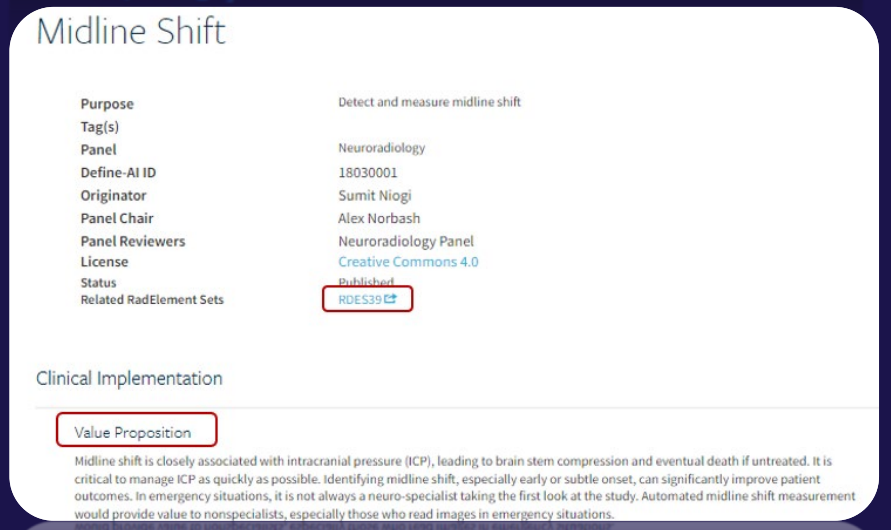
Governance

- Establish an AI governance group to develop policies for evaluating, accepting, monitoring, and sunsetting models
  - Clinical champion (radiologist)
  - Technical lead
  - Administrative personnel
  - Data scientists
  - Informaticists
- Maintain a local inventory of clinically deployed model(s).
  - Model Version
  - Deployment Date
  - Instructions For Use
- Mandatory checkpoint with local Cybersecurity and/or Compliance teams before the deployment

Model Selection

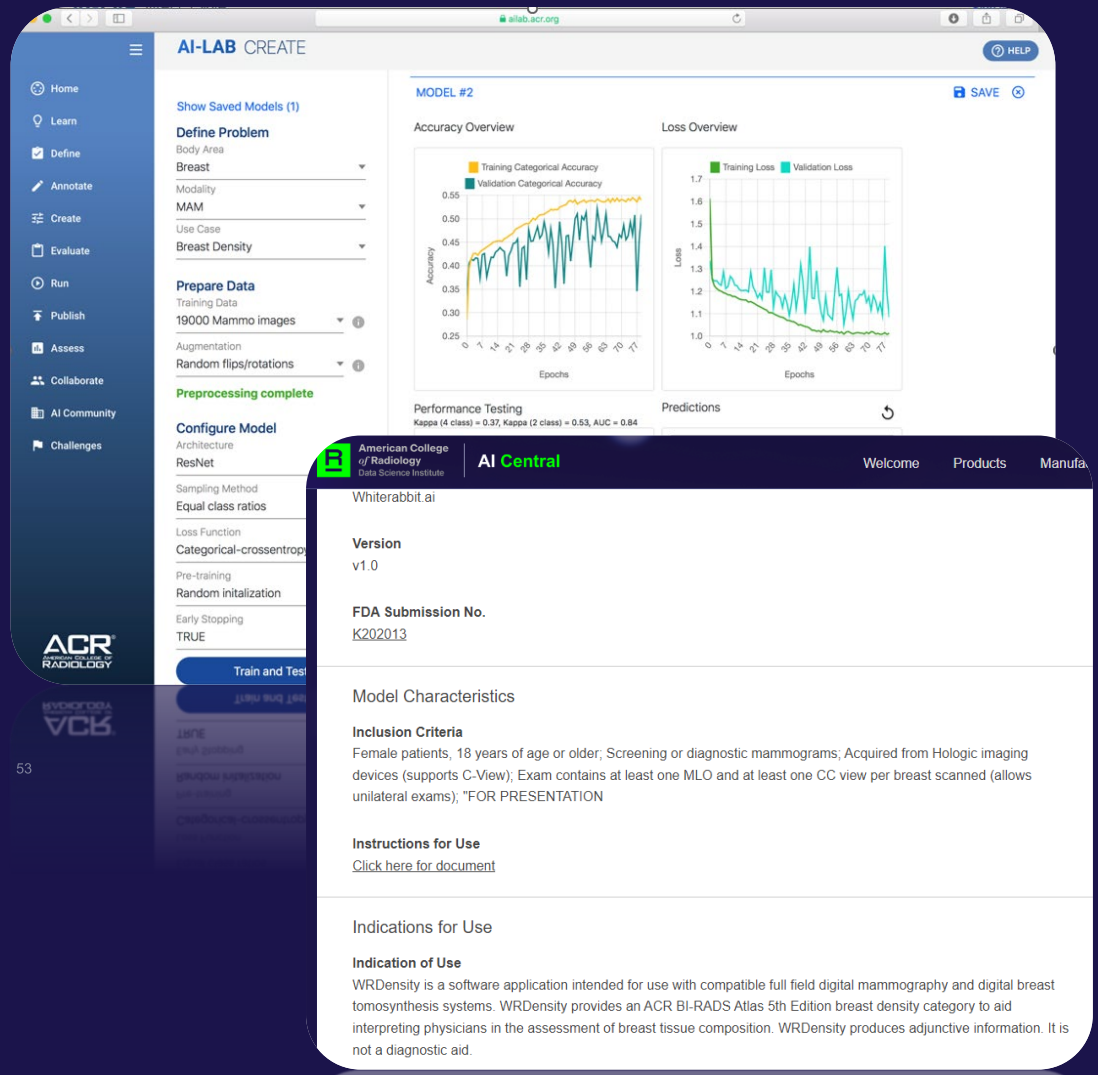
- Consulting with AI Central (or similar) for selecting AI models
- Document the inclusion criteria or DICOM routing rules
- Document the purpose of the AI model

```python
def is_head_ct(dicom_file):
    """
    Check if the DICOM file is a head CT scan.
    """
    ds = pydicom.dcmread(dicom_file)
    if ds.Modality == "CT" and "HEAD" in ds.SeriesDescri
        return True
    return False
```

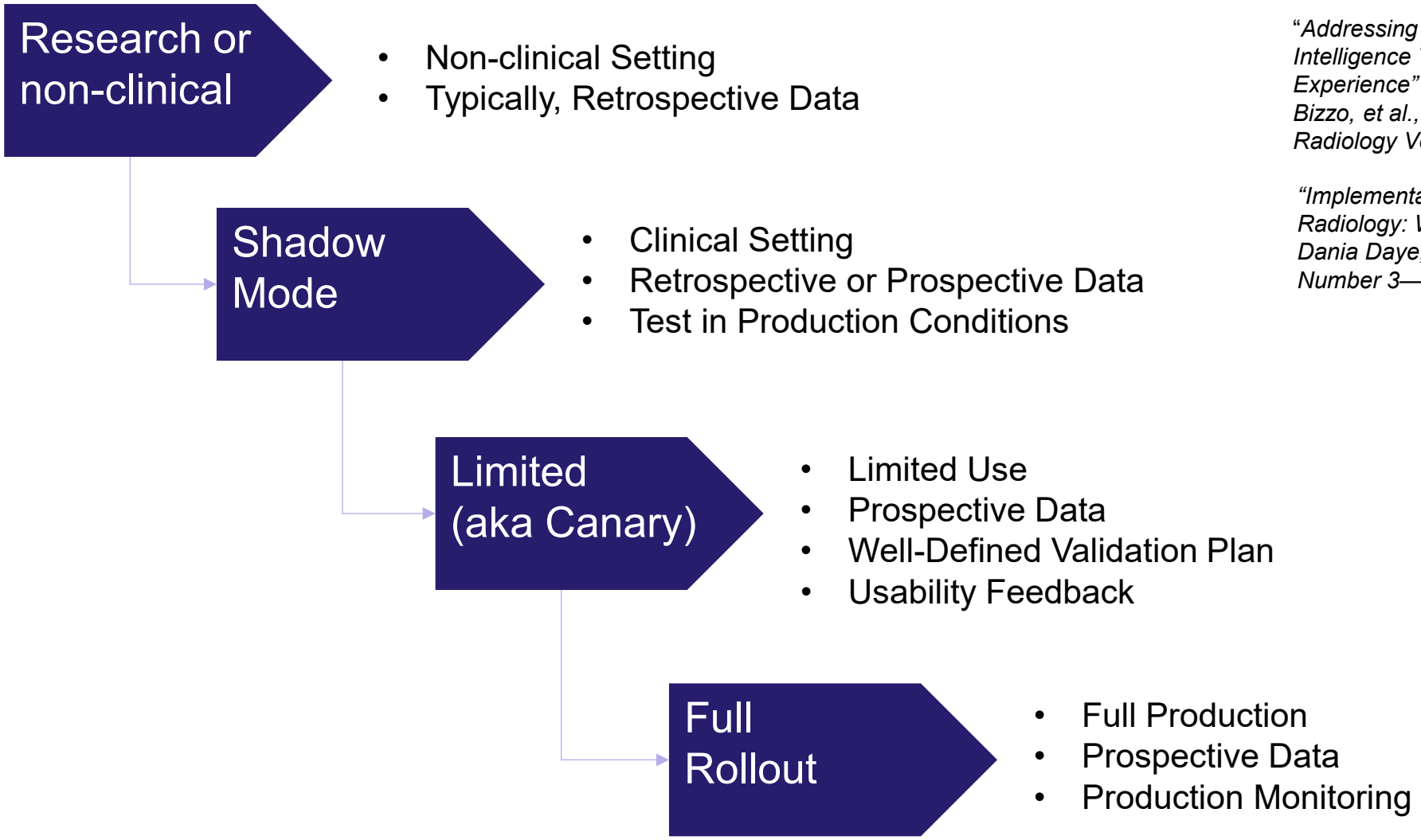


Acceptance Testing

- Document process to assess model performance on local data
- Document training procedures for users of the model
- Disseminate information (e.g., Instructions for Use) to local users



Typical Deployment Pathways



Resources

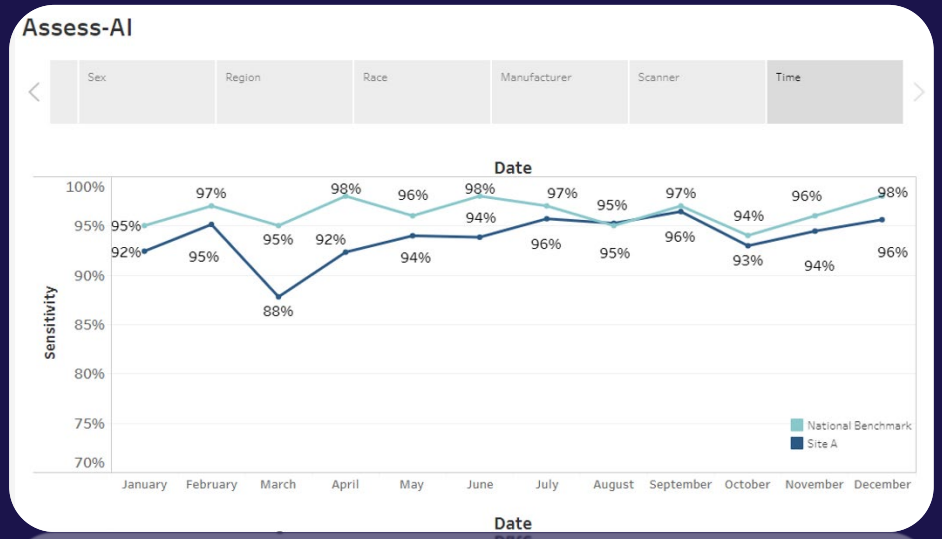
"*Addressing the Challenges of Implementing Artificial Intelligence Tools in Clinical Practice: Principles From Experience*"
Bizzo, et al., Journal of the American College of Radiology Volume 20, March 2023

"Implementation of Clinical Artificial Intelligence in Radiology: Who Decides and How?"
Dania Daye, MD, et al, PhD, Radiology: Volume 305: Number 3—December 2022

Monitoring

Local monitoring

Participation in registry



Locally Developed AI

Trusted personnel

Follows standard procedures

Beth Israel Lahey Health
Lahey Hospital & Medical Center

Texas Children's Hospital®

Yale NewHaven Health
Yale New Haven Hospital

MAYO CLINIC

EMORY UNIVERSITY SCHOOL OF MEDICINE | Department of Radiology and Imaging Sciences

Jefferson
HOME OF SIDNEY KIMMEL MEDICAL COLLEGE

Cleveland Clinic

University Hospitals
Cleveland Medical Center

UNIVERSITY of MARYLAND
SCHOOL OF MEDICINE

Penn Medicine

Mount Sinai

UHealth
UNIVERSITY OF MIAMI HEALTH SYSTEM

Atlantic Health System

UF UNIVERSITY of FLORIDA

UNIVERSITY RADIOLOGY
Leaders in Subspecialty Imaging

SOUTHCOAST
HEALTH SYSTEM

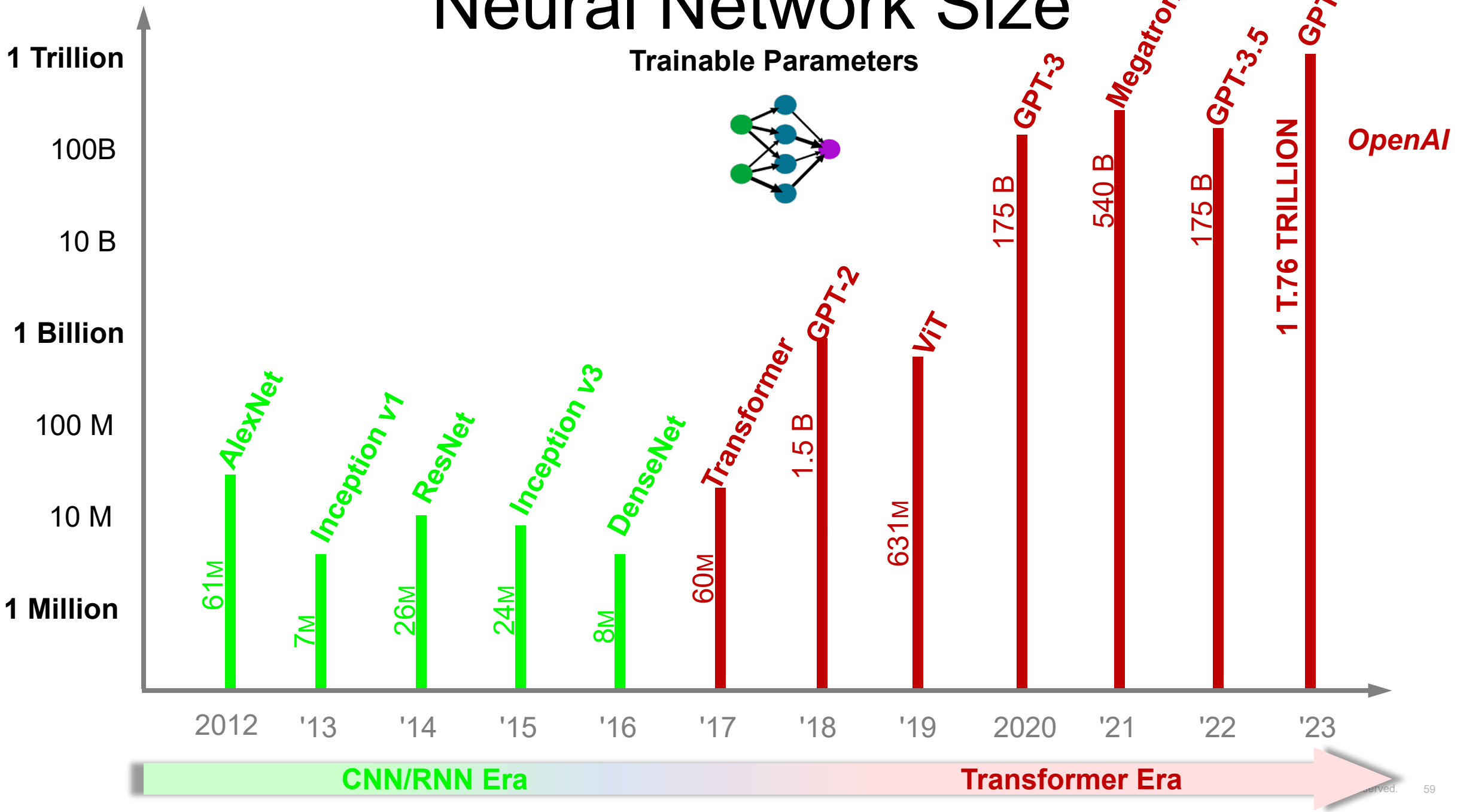
UT Southwestern
Medical Center

UW Medicine
UNIVERSITY OF WASHINGTON
MEDICAL CENTER

UWHealth | School of Medicine and Public Health
UNIVERSITY OF WISCONSIN-MADISON

Notes on LLMs and GenAI

Neural Network Size



Simple vs. Complex Tasks

General	VIEW DOCS
PREDICTED CONCEPT	PROBABILITY
dog	0.999
mammal	0.999
canine	0.997
pet	0.994
animal	0.987
one	0.985
portrait	0.983
no person	0.967
cute	0.957
domestic	0.950

General

VIEW DOCS

PREDICTED CONCEPT	PROBABILITY
dog	0.987
no person	0.984
mammal	0.981
one	0.962
pet	0.952
portrait	0.945
canine	0.937
people	0.926
family	0.896
sit	0.886
animal	0.870
two	

Focused.

You

what is happening in this picture

ChatGPT

In the picture, there is a dog standing in front of a green chair. The chair's cushion appears to have been torn apart, with the stuffing scattered all around on the floor, suggesting that the dog may have destroyed it. The dog has a guilty or sheepish expression, which is commonly interpreted as a sign of knowing it did something wrong. The pillow from the chair is on the floor, undamaged, which might imply that it was removed before or after the cushion was torn apart.

ChatGPT

In the picture, there is a dog standing in front of a green chair. The chair's cushion appears to have been torn apart, with the stuffing scattered all around on the floor, suggesting that the dog may have destroyed it. The dog has a guilty or sheepish expression, which is commonly interpreted as a sign of knowing it did something wrong. The pillow from the chair is on the floor, undamaged, which might imply that it was removed before or after the cushion was torn apart.

Home
Learn
Define
Annotate
Create
Evaluate
Run
Publish
Assess
Collaborate
AI Community
Generative AI
Challenges

Generative AI: *Radiology Meets ChatGPT*

Welcome to ACR's GPT reference, research and translation tool. Our GPT assisted tool will help patients, physicians and radiologists discover previously hard to find documentation, definitions and reports concerning a multitude of imaging procedures.

For Patients

Explore how AI can be used to translate a radiology report to a different language or to layman's terms for a patient.

For Radiologists

As an added feature AI can now be implemented to support a radiologist reading by summarizing the patient history.

For Referring Provider

Discover how AI can be used to support ordering appropriate procedures for a patient.

Home
Learn
Define
Annotate
Create
Evaluate
Run
Publish
Assess
Collaborate
AI Community
Generative AI
Challenges

PATIENT 1: Ultrasound Post-Read

Summary:

The radiologist has just read an ultrasound procedure for left pelvic pain on an premenopausal patient.

Report:

Current

Patient: Patrick Smith
Date of Exam: 02-21-2023
Exam Type: Pelvic Ultrasound

Indication:
Left pelvic pain in a premenopausal patient

Findings:
The right kidney measures 10.2 cm in length, and the left kidney measures 11.2 cm in length. No hydronephrosis is identified.
There is a 3.2 cm parapelvic renal cyst seen in the left kidney.
The gallbladder contains multiple mobile gallstones, with no evidence of acute cholecystitis on Murphy's sign.
The liver demonstrates increased echogenicity, suggestive of fatty infiltration. The spleen measures 11.2 cm in length. The aorta appears normal in caliber, measuring 2.1 cm in diameter.
In the left adnexa, a 4.5 cm cyst is identified with internal echoes, without associated blood flow.

Suggestions:

Format the quantitative findings in a table f... | Reset | Export

Format the quantitative findings in a table format

Format the quantitative findings in a table format

Right Kidney Length:	10.2 cm
Left Kidney Length:	11.2 cm
Parapelvic Renal Cyst (left):	3.2 cm
Gallstones (mobile):	Multiple
Liver Echogenicity:	Increased
Spleen Length:	11.2 cm
Aorta Diameter:	2.1 cm
Left Adnexal Cyst:	4.5 cm

Healthcare AI Challenge

American College of Radiology™
Data Science Institute

Mass General Brigham

EMORY HEALTHCARE

UWHealth

UW Medicine

Healthcare AI Challenge

A Collaborative Community for Healthcare AI Validation and Monitoring

1. Geographically & Demographically Diverse, Embargoed Data
2. Unlimited Clinical Use Cases (e.g. Draft Report Generation)
3. Geographically & Institutionally Diverse Experts
4. Expert evaluation of AI Agents with Input Data

→

AI Challenge Results ★

- AI Discordance
- AI Performance
- AI Generalizability
- AI Bias

↑ **AI Performance Ratings (Public Forum)**

↓ **AI Decision Making (Institutional)**

Healthcare AI Challenge

Powered by MGB AI Arena

Description | Analytics

The X-Ray Interpretation Event includes radiography cases from major body parts, organized by Clinical Specialties.

Healthcare professionals can review and rate responses from foundation models that process radiography cases for various Challenges.

Each response should be rated on a five-point clinical skill level scale: Radiologist, Fellow, Resident, Medical Student, or Unacceptable.

Events

- X-Ray Interpretation (Basic)
- X-Ray Interpretation (Advanced)
- CT Head Interpretation
- Clinical Decision Support

Challenges — Select one or more Challenges - default is all

- [ ] Draft Report Generation
- [ ] Findings
- [ ] Differential Diagnosis
- [ ] Recommendations
- [ ] Treatment Considerations

Specialty — Select one or more Specialties - default is all

- Cardiology
- Emergency
- Endocrinology
- Gastroenterology
- General Medicine
- Neurology
- Oncology
- Orthopedics
- Pulmonology
- Rheumatology
- Surgery
- Urology
- Women's Health

Healthcare AI Challenge

Powered by MGB AI Arena

Home About My Account demo@aihealthcarearena.ai

Draft Report Generation Challenge 1 of 1

Findings: The lumbar spine X-rays in the anteroposterior and lateral views show normal alignment of the vertebral bodies. There is no evidence of fracture or dislocation. The intervertebral disc spaces are preserved. There is no significant osteophyte formation or other signs of degenerative changes. The sacroiliac joints appear normal. No abnormal soft tissue calcifications are noted.

Impression: Normal lumbar spine X-ray. No acute osseous abnormalities or significant degenerative changes are observed.

Recommendations: No immediate follow-up is necessary based on the current imaging. Clinical correlation is recommended to determine if further evaluation is needed based on symptoms. If symptoms persist or worsen, consider advanced imaging such as MRI for further assessment.

Select the clinical skill level you would assign to the model's response

Attending (Outstanding) | Fellow (Good) | Resident (Acceptable) | Student (Poor) | Unacceptable

Next Case

Healthcare AI Challenge
Powered by MGB AI Arena

Home About My Account demo@aihealthcarearena.ai

Images All Models



Draft Report Generation Challenge 1 of 1

Findings: The lumbar spine X-rays in the anteroposterior and lateral views show normal alignment of the vertebral bodies. There is no evidence of fracture or dislocation. The intervertebral disc spaces are preserved. There is no significant osteophyte formation or other signs of degenerative changes. The sacroiliac joints appear normal. No abnormal soft tissue calcifications are noted.

Impression: Normal lumbar spine X-ray. No acute osseous abnormalities or significant degenerative changes are observed.

Recommendations: No immediate follow-up is necessary based on the current imaging. Clinical correlation is recommended to determine if further evaluation is needed based on symptoms. If symptoms persist or worsen, consider advanced imaging such as MRI for further assessment.

The above answer was generated by OpenAI (GPT 4o mini)

Enter Comments (Optional) Save

Select the clinical skill level you would assign to the model's response

Attending (Outstanding) | Fellow (Good) | Resident (Acceptable) | Student (Poor) | Unacceptable

Next Challenge New Case

Healthcare AI Challenge

Powered by MGB AI Arena

Home About My Account demo@aihealthcarearena.ai

Images | All Models

Anthropic (Haiku) | Anthropic (Opus) | Anthropic (Sonnet) | Gemini (Flash) | Gemini (Pro)

OpenAI (GPT 4o) | OpenAI (GPT 4o mini) | Harrison (RAD1) | Microsoft (Draft CXR)

Instruction Prompt **Unformatted Result** Formatting Prompt Result

Findings:The images show multiple vertebral bodies with irregularity and decreased attenuation, suggesting spinal degenerative changes. There is also evidence of mild scoliosis with curvature of the spine. The spinal canal appears patent without significant compromise. The soft tissues appear unremarkable.Impression:The findings are consistent with age-related degenerative changes of the spine, including spondylosis and mild scoliosis. There is no evidence of acute injury or significant spinal canal narrowing.Recommendations:No immediate intervention is required. The patient should continue with conservative management, including physical therapy and pain management as needed. If symptoms persist or worsen, further imaging evaluation, such as MRI, may be considered to assess for any underlying pathology.

Draft Report Generation Challenge 1 of 1

Findings: The lumbar spine X-rays in the anteroposterior and lateral views show normal alignment of the vertebral bodies. There is no evidence of fracture or dislocation. The intervertebral disc spaces are preserved. There is no significant osteophyte formation or other signs of degenerative changes. The sacroiliac joints appear normal. No abnormal soft tissue calcifications are noted.

Impression: Normal lumbar spine X-ray. No acute osseous abnormalities or significant degenerative changes are observed.

Recommendations: No immediate follow-up is necessary based on the current imaging. Clinical correlation is recommended to determine if further evaluation is needed based on symptoms. If symptoms persist or worsen, consider advanced imaging such as MRI for further assessment.

The above answer was generated by OpenAI (GPT 4o mini)

don't count my vote Save

Select the clinical skill level you would assign to the model's response

Attending (Outstanding) | Fellow (Good) | Resident (Acceptable) | Student (Poor) | Unacceptable

Next Challenge | New Case

Healthcare AI Challenge

Powered by MGB AI Arena

Home About My Account demo@aihealthcarearena.ai

Images | All Models

Anthropic (Haiku) | Anthropic (Opus) | Anthropic (Sonnet) | Gemini (Flash) | Gemini (Pro)

OpenAI (GPT 4o) | OpenAI (GPT 4o mini) | Harrison (RAD1) | Microsoft (Draft CXR)

Instruction Prompt Unformatted Result Formatting Prompt Result

* You are a Radiologist, create a draft report for the images that contains the findings, a draft impression, and any follow-up recommendations. Put them all in separate paragraphs and exclude any patient information. Do not include any disclaimers as this is for educational purposes only.

Draft Report Generation — Challenge 1 of 1

Findings: The lumbar spine X-rays in the anteroposterior and lateral views show normal alignment of the vertebral bodies. There is no evidence of fracture or dislocation. The intervertebral disc spaces are preserved. There is no significant osteophyte formation or other signs of degenerative changes. The sacroiliac joints appear normal. No abnormal soft tissue calcifications are noted.

Impression: Normal lumbar spine X-ray. No acute osseous abnormalities or significant degenerative changes are observed.

Recommendations: No immediate follow-up is necessary based on the current imaging. Clinical correlation is recommended to determine if further evaluation is needed based on symptoms. If symptoms persist or worsen, consider advanced imaging such as MRI for further assessment.

The above answer was generated by OpenAI (GPT 4o mini)

don't count my vote

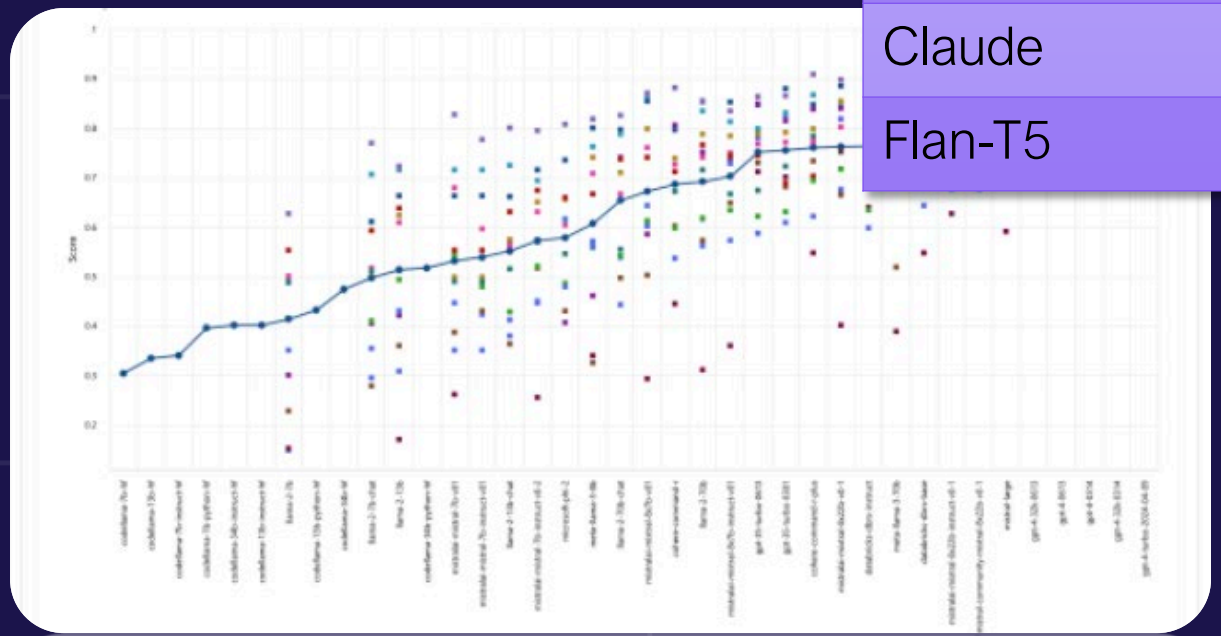
Save

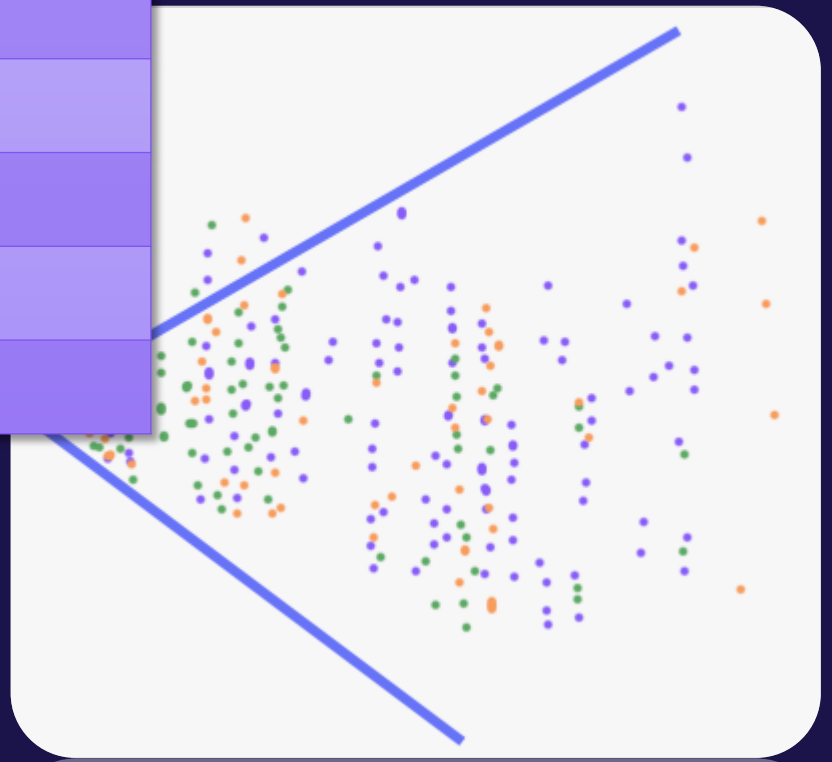
Select the clinical skill level you would assign to the model's response

Attending (Outstanding) | Fellow (Good) | Resident (Acceptable) | Student (Poor) | Unacceptable

Next Challenge | New Case

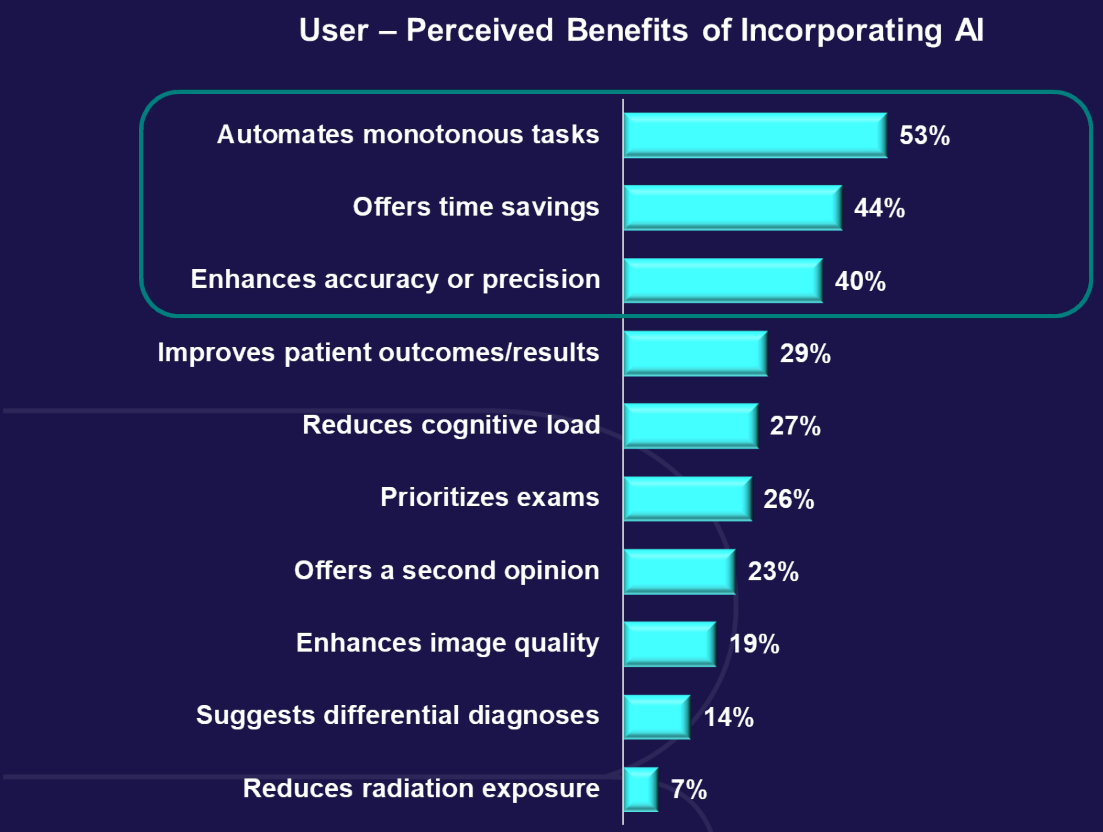
Model	ELO
GPT-4	1059.63
GPT-3.5	940.38
Gemini	960.42
LLaMa	770.02
PaLM 2	862.82
Claude	607.58
Flan-T5	400.28





Conclusion

Risk vs. Reward: Factors Influencing the Calculation



Factors	Examples of Influences
Subjective Risk Attitudes	Individual, Practice, Local Conditions/Market
Type of Intended Use	Triage, Detection, Diagnosis
Clinical Use Case	Different clinical use cases influence risk attitudes
Operational Safeguards	Transparency, Monitoring, Best Practice Implementation

Thank You

Focused. Forward. Together.™