ACR®

Making imaging safe, effective and accessible to those who need it.

November 4, 2019

Medical Device Innovation Consortium (MDIC) Digital Pathology & Artificial Intelligence Meeting

Bibb Allen, Jr. MD FACR

Chief Medical Officer, American College of Radiology Data Science Institute

Grandview Medical Center

Birmingham, Alabama

DATA SCIENCE INSTITUTE™

AMERICAN COLLEGE OF RADIOLOGY

No Commercial Conflicts Of Interest

Neither I nor my immediate family have a financial relationship with a commercial organization that may have a direct or indirect interest in the content of this presentation

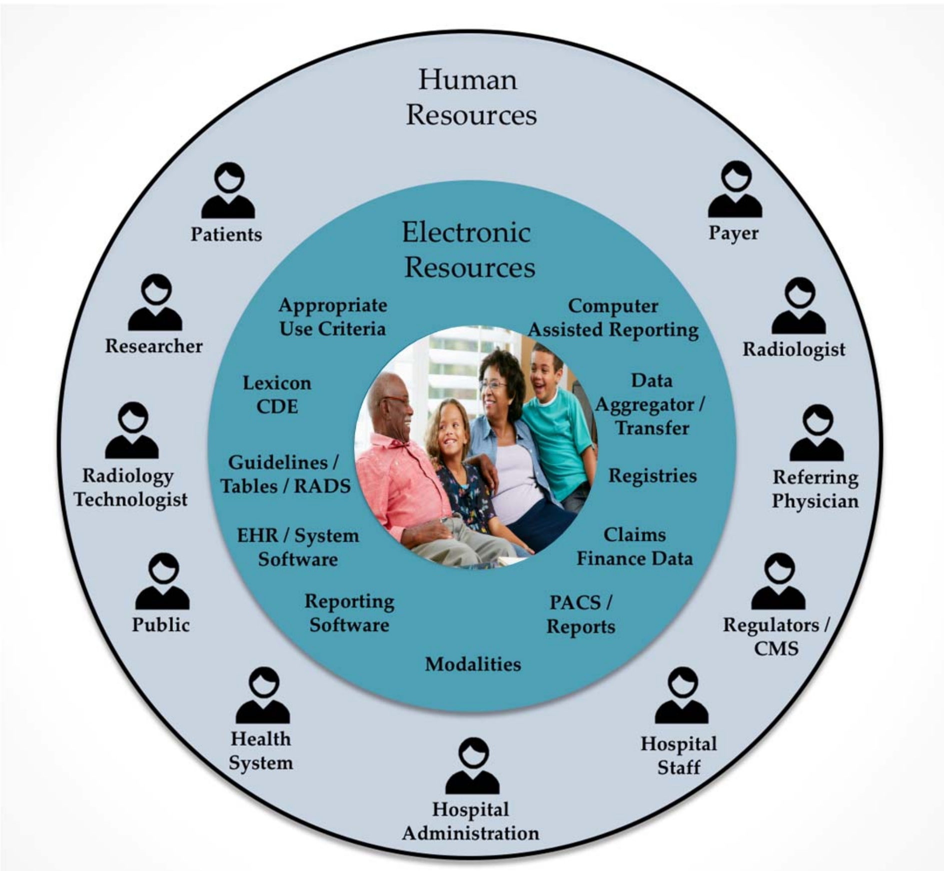
DATA SCIENCE INSTITUTE™
AMERICAN COLLEGE OF RADIOLOGY

- Chief Medical Officer American College of Radiology Data Science Institute
- Former Board Chair and President ACR

Objectives

- Artificial intelligence will be transformational technology for improving how we care for our patients
- The democratization of AI will accelerate the advancement of AI in healthcare and radiologists should play a leading role
- Physicians and medical specialty societies can facilitate the development, deployment and clinical use of AI by fostering an ecosystem between disparate stakeholders including public-private partnerships with regulatory agencies

Healthcare Ecosystem



Radiology AI Ecosystem

- Patients
- Radiology professionals
- Researchers and academic centers
- Industry developers
- Governmental agencies
- Hospitals and health systems
- Insurers and third-party payers

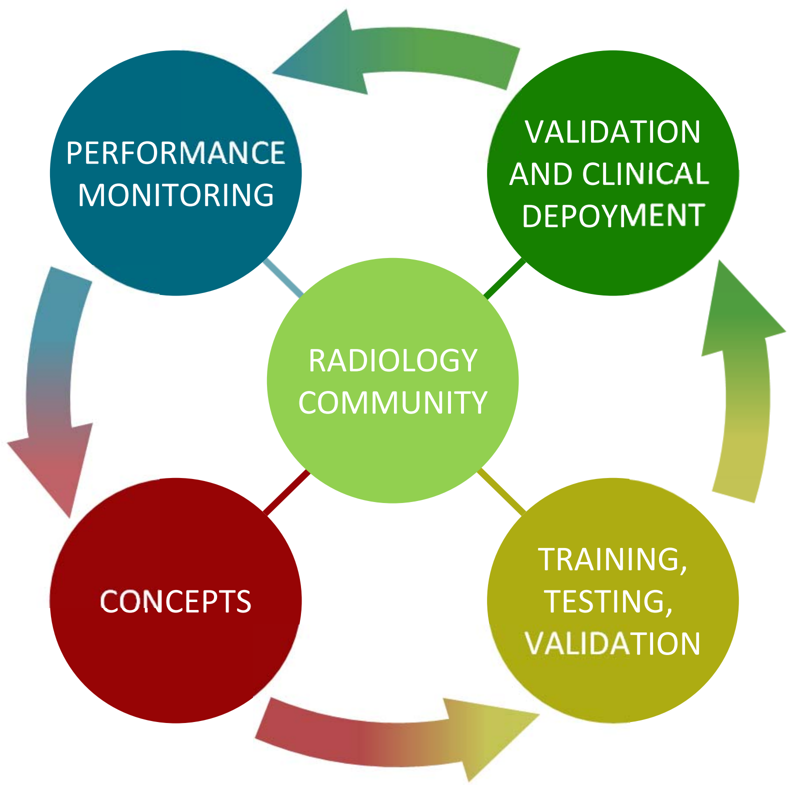
Concept To Clinical Practice – Radiology AI Ecosystem

The Radiology AI Ecosystem

Ideas to Clinical Practice

Define standard methods to integrate and monitor, AI models in clinical practice using real world evidence

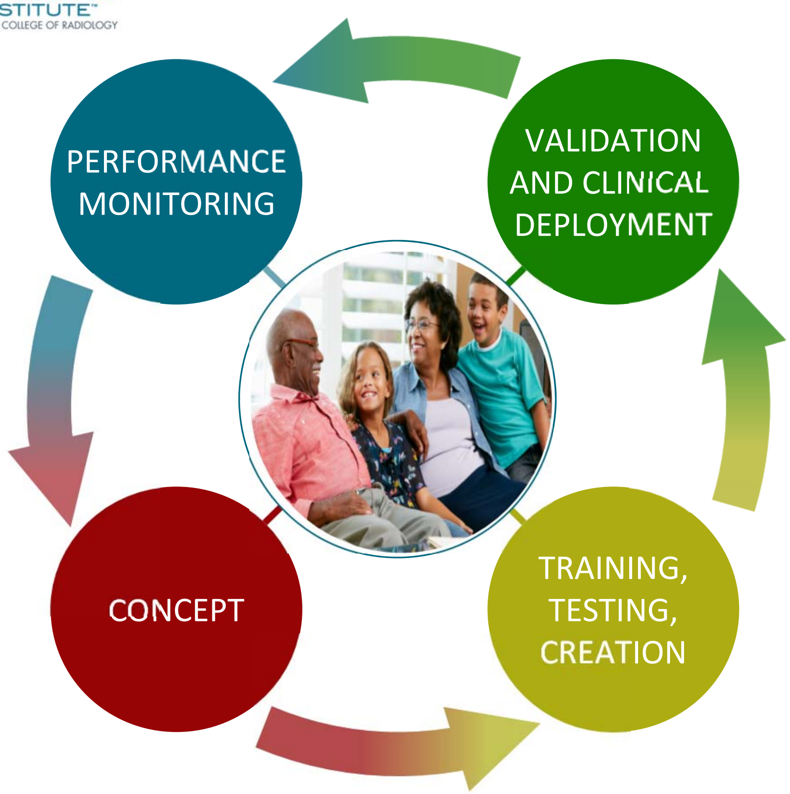
Define standards for use cases considering clinical needs and technical capabilities



Define standardized methods for AI model validation consistent with regulatory processes

Define standard methods to aggregate and annotate data for AI model training and testing

The Radiology AI Ecosystem
Ideas To Clinical Practice



DATA SCIENCE INSTITUTE™
AMERICAN COLLEGE OF RADIOLOGY

ACR Strategic Plan For Data Science

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

- Educate on the appropriate use and ethical issues for AI in radiology
- Define the appropriate uses of AI in radiology
- Help radiologists become global leaders in data science

Radiology’s Value Proposition

Part Of The Solution

Imaging 3.0: Value-Based Radiology

Clinical Decision Support for Ordering Physicians
Providing >24 Million examinations per month

Image Sharing
RSNA / NIH / Vendors

Structured Reporting
Incorporated in all VR reporting platforms

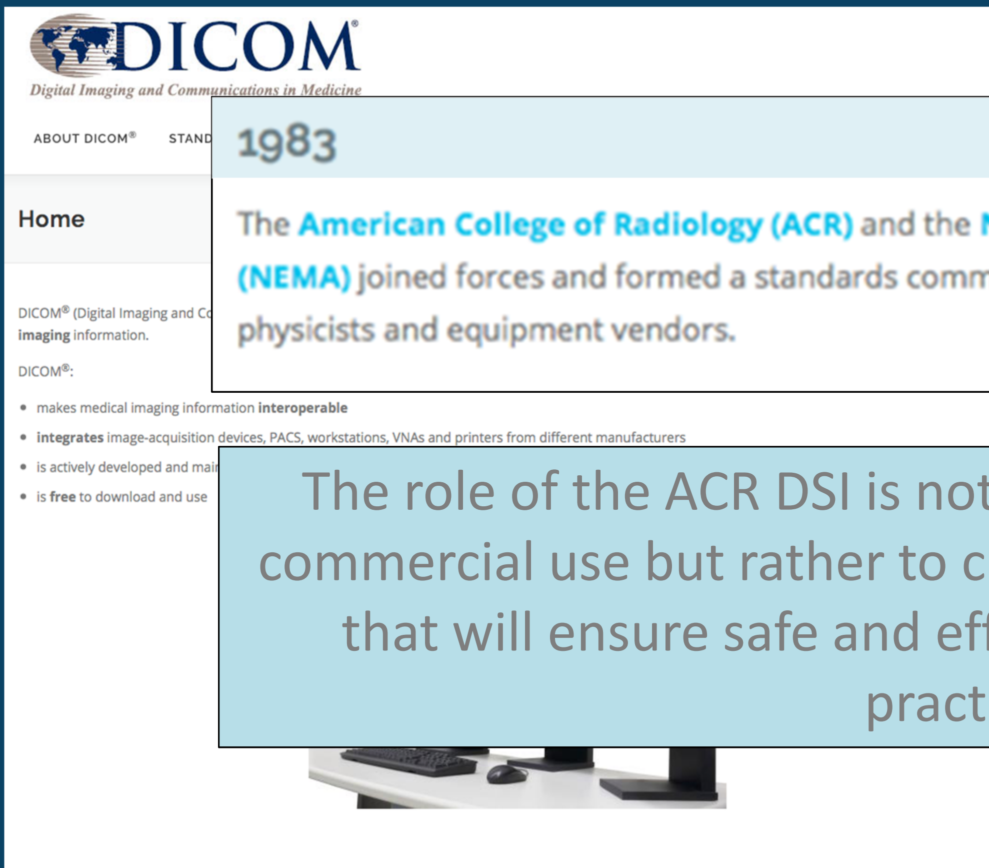
Registries
Radiation Exposure / Patient Outcomes / Quality

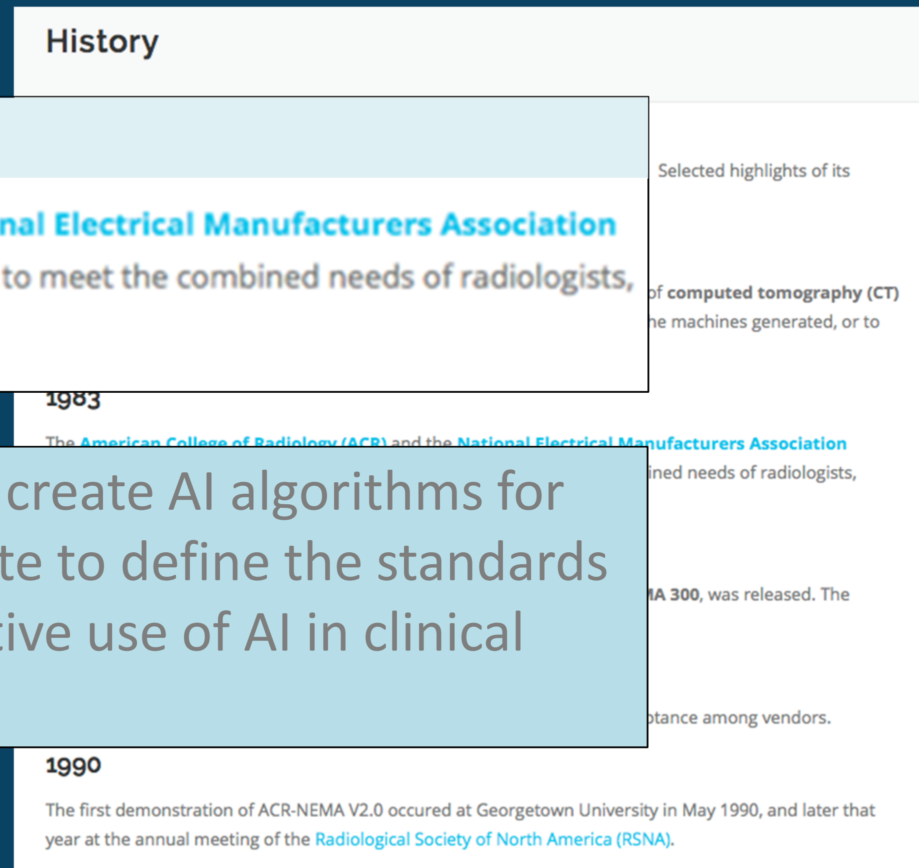
Portfolio of IT Tools

Clinical Decision Support for Image Interpretation
Integrated into >75% of radiologists desktops

Artificial Intelligence

ACR DSI – Developing Standards For Industry And Institutions





The role of the ACR DSI is not to create AI algorithms for commercial use but rather to create to define the standards that will ensure safe and effective use of AI in clinical practice.

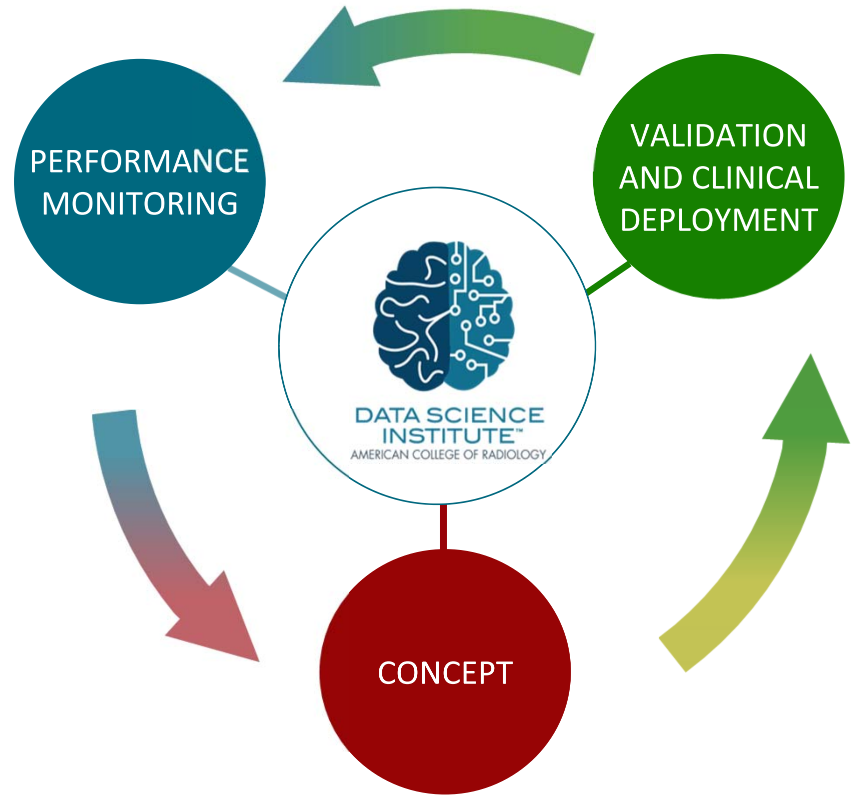
Advancing AI In Clinical Practice While Protecting Patients From Unintended Consequences Of AI

- Algorithms are useful, safe and effective
- Clinically validated
- Transparency in algorithm output
- Monitored in practice
- Free of unintended bias
- Medicare and insurance coverage issues

Artificial Intelligence: Concept To Clinical Practice

How Do We Make Sure AI Is Working In The Real World?

How Do We Validate AI Algorithms For Clinical Practice?



What Are The Most Important Clinical Tasks For AI?

Research Needs to Advance AI in Clinical Practice

DATA SCIENCE INSTITUTE
AMERICAN COLLEGE OF RADIOLOGY

Radiology

ORIGINAL RESEARCH • SPECIAL REPORT

A Roadmap for Foundational Research on Artificial Intelligence in Medical Imaging: From the 2018 NIH/RSNA/ACR/The Academy Workshop

Curtis P. Langlotz, MD, PhD • Bibb Allen, MD • Bradley J. Erickson, MD, PhD • Jayashree Kalpathy-Cramer, PhD • Keith Bigelow, BA • Tessa S. Cook, MD, PhD • Adam E. Flanders, MD • Matthew P. Lungren, MD, MPH • David S. Mendelson, MD • Jeffrey D. Rudie, MD, PhD • Ge Wang, PhD • Krishna Kandarpa, MD, PhD

From the Department of Radiology, Stanford University, Stanford, CA 94305 (C.P.L., M.P.L.); Department of Radiology, Grandview Medical Center, Birmingham, Ala (B.A.); Department of Radiology, Mayo Clinic, Rochester, Minn (B.J.E.); Department of Radiology, Massachusetts General Hospital, Harvard Medical School, Boston, Mass (J.K.C.); GE Healthcare, Chicago, Ill (K.B.); Department of Radiology, Hospital of the University of Pennsylvania, Philadelphia, Pa (T.S.C., J.D.R.); Department of Radiology, Thomas Jefferson University Hospital, Philadelphia, Pa (A.E.F.); Department of Radiology, Icahn School of Medicine at Mount Sinai, New York, NY (D.S.M.); Biomedical Imaging Center, Rensselaer Polytechnic Institute, Troy, NY (G.W.); and National Institute of Biomedical Imaging and Bioengineering, National Institutes of Health, Washington, DC (K.K.). Received March 17, 2019; revision requested March 19; revision received March 24; accepted March 25. **Address correspondence to** C.P.L. (e-mail: *langlotz@stanford.edu*).

Conflicts of interest are listed at the end of this article.

Radiology 2019; 291:781–791 • https://doi.org/10.1148/radiol.2019190613 • Content code: IN

Imaging research laboratories are rapidly creating machine learning systems that achieve expert human performance using open-source methods and tools. These artificial intelligence systems are being developed to improve medical image reconstruction, noise reduction, quality assurance, triage, segmentation, computer-aided detection, computer-aided classification, and radiogenomics. In August 2018, a meeting was held in Bethesda, Maryland, at the National Institutes of Health to discuss the current state of the art and knowledge gaps and to develop a roadmap for future research initiatives. Key research priorities include: 1, new image reconstruction methods that efficiently produce images suitable for human interpretation from source data; 2, automated image labeling and annotation methods, including information extraction from the imaging report, electronic phenotyping, and prospective structured image reporting; 3, new machine learning methods for clinical imaging data, such as tailored, pretrained model architectures, and federated machine learning methods; 4, machine learning methods that can explain the advice they provide to human users (so-called explainable artificial intelligence); and 5, validated methods for image de-identification and data sharing to facilitate wide availability of clinical imaging data sets. This research roadmap is intended to identify and prioritize these needs for academic research laboratories, funding agencies, professional societies, and industry.

JACR Journal of the American College of Radiology

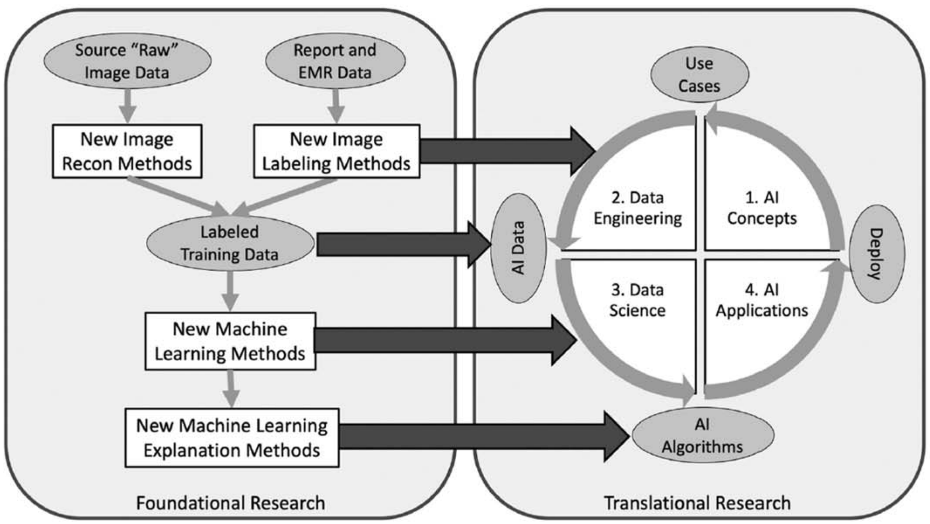
Health Services Research & Policy • Clinical Practice Management • Data Science • Training & Education • Leadership

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



AI DEVELOPMENT IN MEDICAL IMAGING

Fig 1. As in other industries, AI development in medical imaging includes both foundational and translational research activities. The foundational portion of the National Institutes of Health Workshop considered research priorities to accelerate and improve the development of AI algorithms for medical imaging [8]. The translational portion of the workshop considered medical imaging use cases for algorithm development and how these applications will be validated, deployed, and monitored in routine clinical practice. The diagram shows how foundational and translational research activities are connected. Foundational research leads to new image reconstruction and labeling methods, new machine learning algorithms, and new explanation methods, each of which enhance the data sets, data engineering, and data science that lead to the successful deployment of AI applications in medical imaging. AI = artificial intelligence; EMR = electronic medical record; Recon = reconstruction. The figure was developed by the authors for publication in both *Radiology* and *JACR*. This figure also published in reference 8.

Radiology AI Ecosystem

- Structured use cases
- Data access
- Patient safety
- Clinical integration

FDA Discussion Paper on Continuously Learning Algorithms and the FDA Software Precertification Program

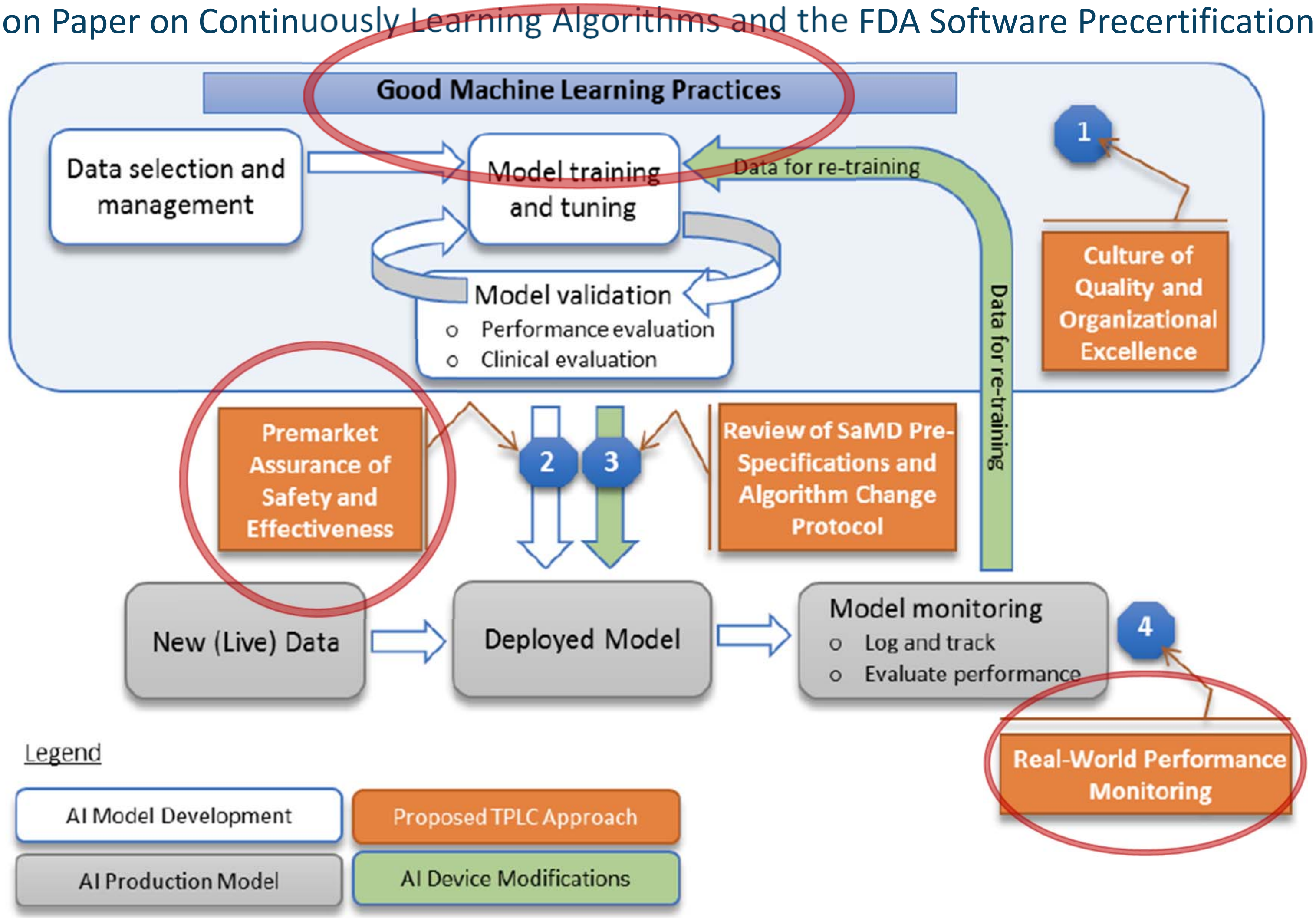


Figure 2: Overlay of FDA's TPLC approach on AI/ML workflow

FDA Discussion Paper on Continuously Learning Algorithms and the FDA Software Precertification Program

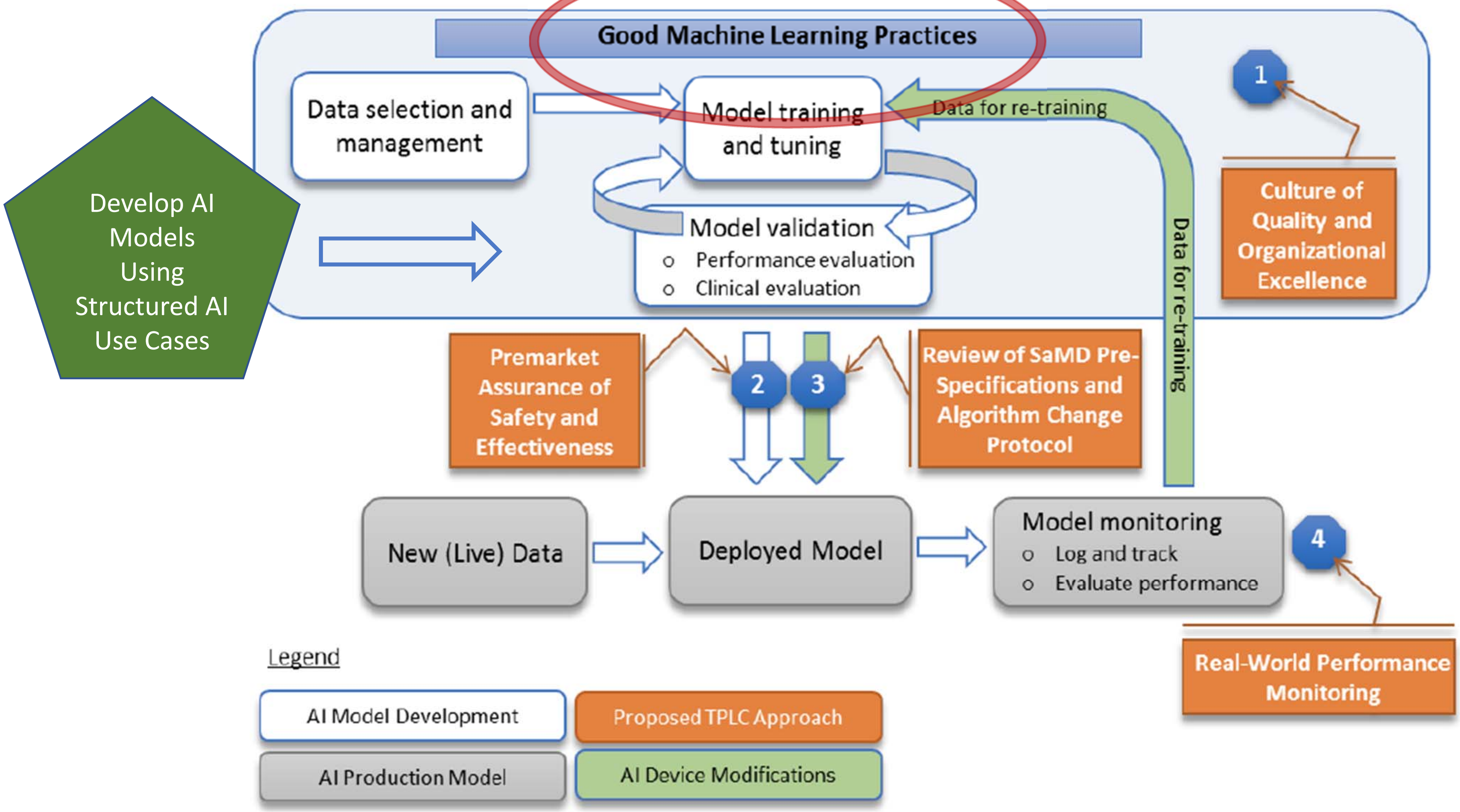
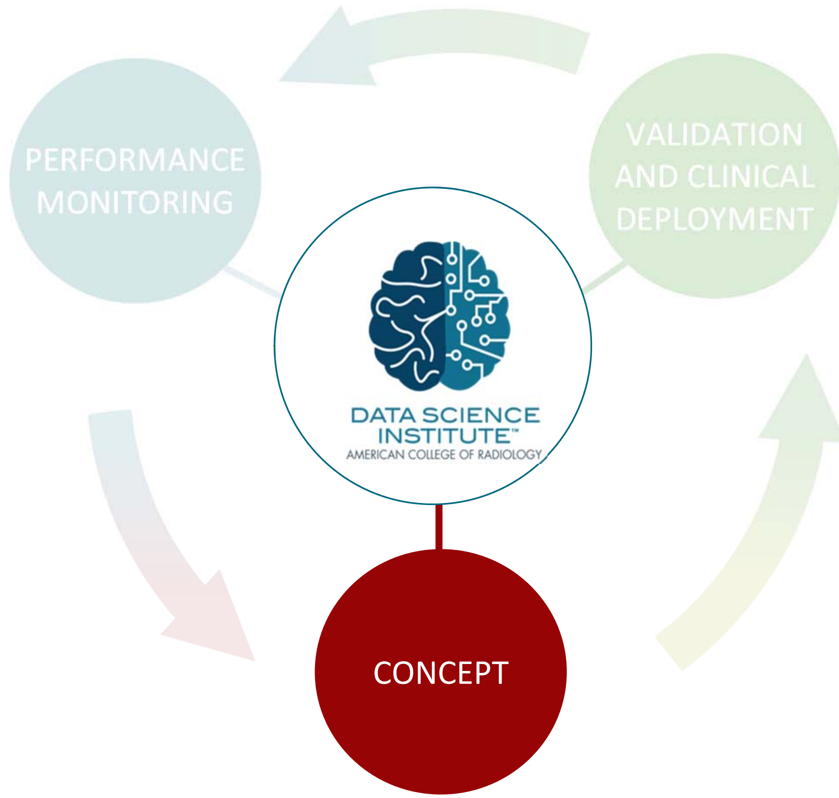


Figure 2: Overlay of FDA's TPLC approach on AI/ML workflow

Artificial Intelligence: Concept To Clinical Practice

How Do We Make Sure AI Is Working In The Real World?



How Do We Validate AI Algorithms For Clinical Practice?

What Are The Most Important Clinical Tasks For AI?

Artificial Intelligence: Concept To Clinical Practice

How Do We Make Sure AI Is Working In The Real World?

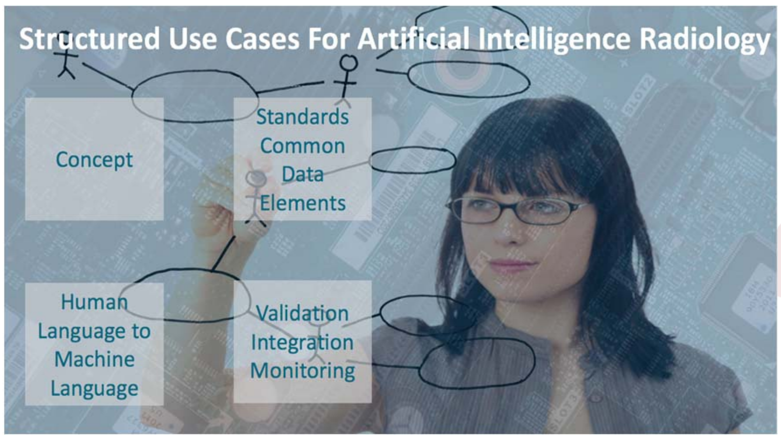
PERFORMANCE MONITORING

VALIDATION AND CLINICAL DEPLOYMENT

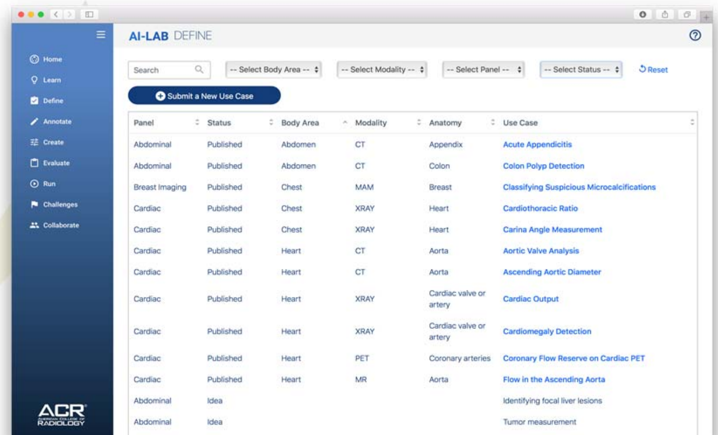
How Do We Validate AI Algorithms For Clinical Practice?

DATA SCIENCE INSTITUTE™
AMERICAN COLLEGE OF RADIOLOGY

Define-AI



ACR DSI STRUCTURED AI USE CASES



Panel	Status	Body Area	Modality	Anatomy	Use Case
Abdominal	Published	Abdomen	CT	Appendix	Acute Appendicitis
Abdominal	Published	Abdomen	CT	Colon	Colon Polyp Detection
Breast Imaging	Published	Chest	MAM	Breast	Classifying Suspicious Microcalcifications
Cardiac	Published	Chest	XRAY	Heart	Cardiothoracic Ratio
Cardiac	Published	Chest	XRAY	Heart	Carina Angle Measurement
Cardiac	Published	Heart	CT	Aorta	Aortic Valve Analysis
Cardiac	Published	Heart	CT	Aorta	Ascending Aortic Diameter
Cardiac	Published	Heart	XRAY	Cardiac valve or artery	Cardiac Output
Cardiac	Published	Heart	XRAY	Cardiac valve or artery	Cardiomegaly Detection
Cardiac	Published	Heart	PET	Coronary arteries	Coronary Flow Reserve on Cardiac PET
Cardiac	Published	Heart	MR	Aorta	Flow in the Ascending Aorta
Abdominal	Idea				Identifying focal liver lesions
Abdominal	Idea				Tumor measurement

RadElement.org

What Are The Most Important Clinical Tasks For AI?

Common Data Elements

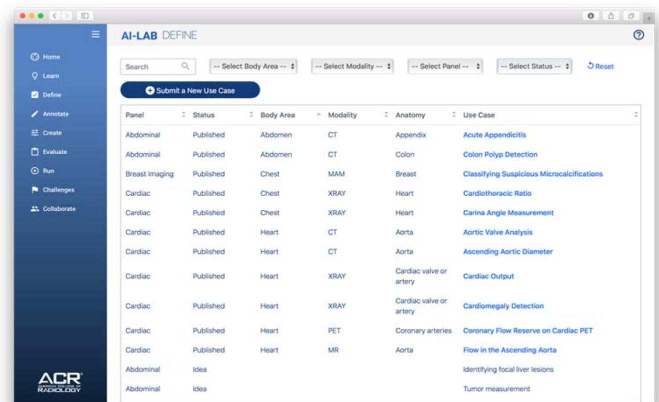
“Good Machine Learning Practices”

Structured AI Use Cases

- Standardized inputs and outputs
- Common data elements
- Defined pathways for clinical integration

DATA SCIENCE INSTITUTE™
AMERICAN COLLEGE OF RADIOLOGY

Define-AI



FDA Discussion Paper on Continuously Learning Algorithms and the FDA Software Precertification Program

DATA SCIENCE INSTITUTE™
AMERICAN COLLEGE OF RADIOLOGY

Define-AI

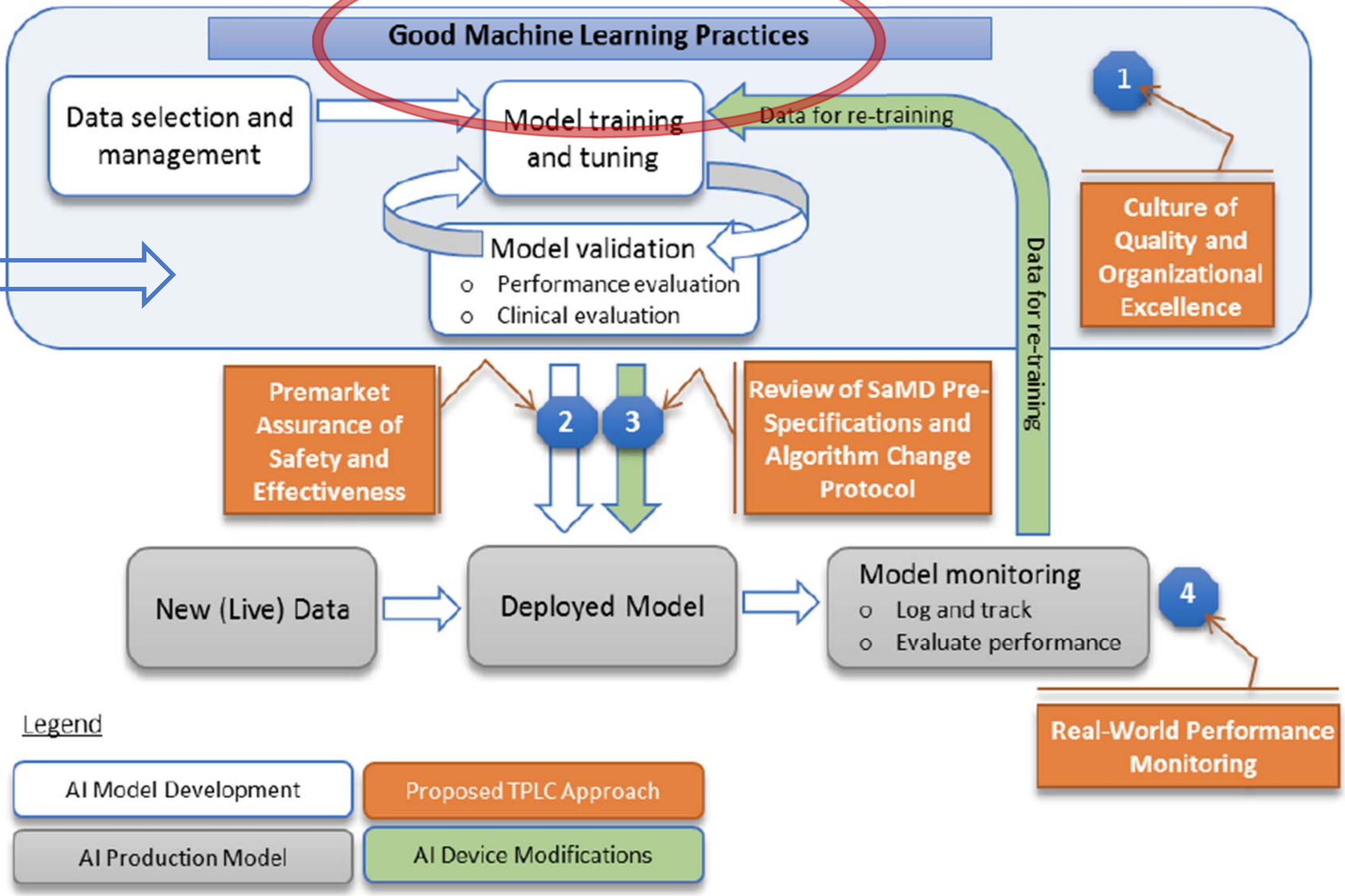


Figure 2: Overlay of FDA's TPLC approach on AI/ML workflow

FDA Discussion Paper on Continuously Learning Algorithms and the FDA Software Precertification Program

DATA SCIENCE INSTITUTE™
AMERICAN COLLEGE OF RADIOLOGY

Define-AI

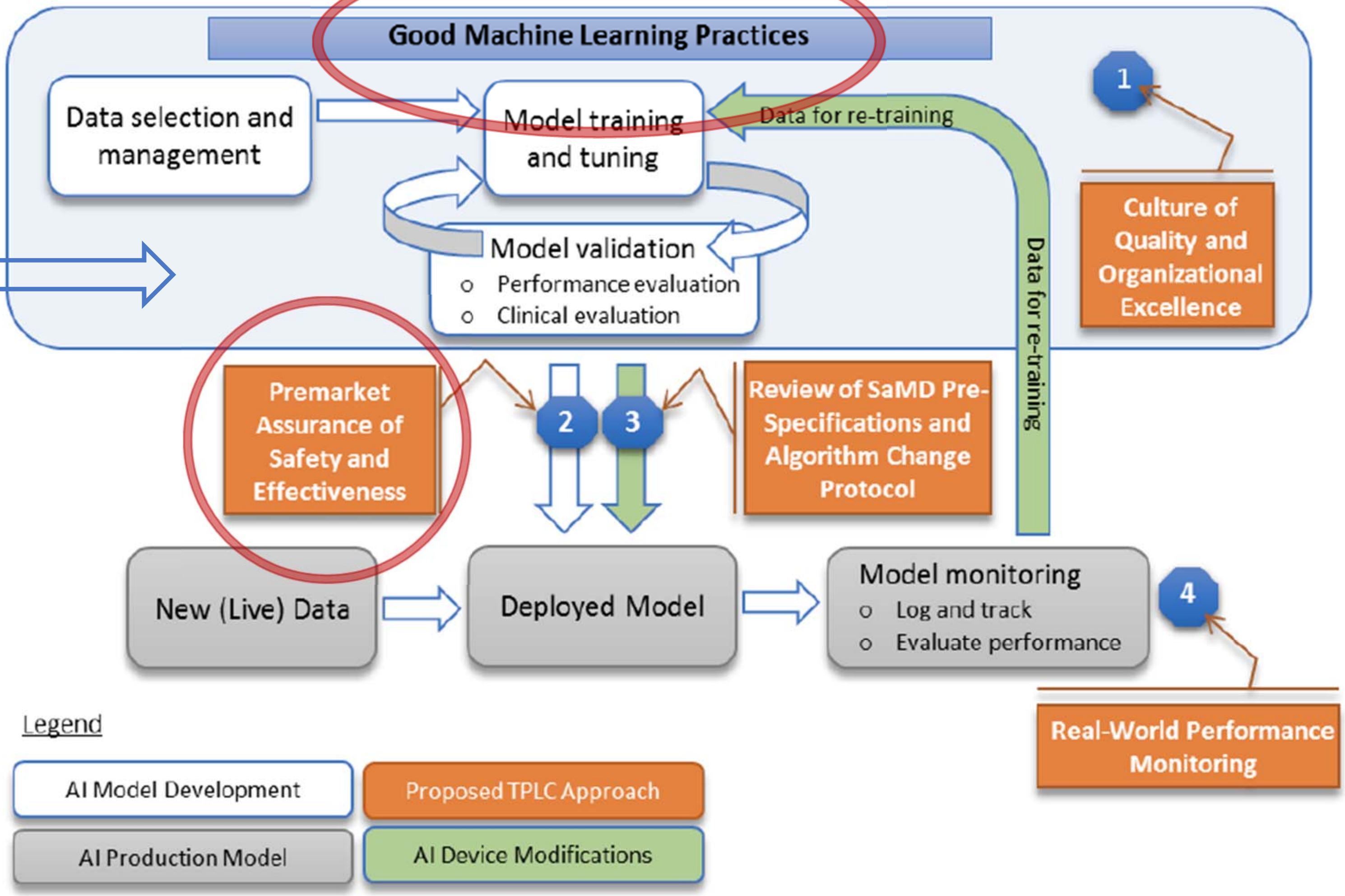
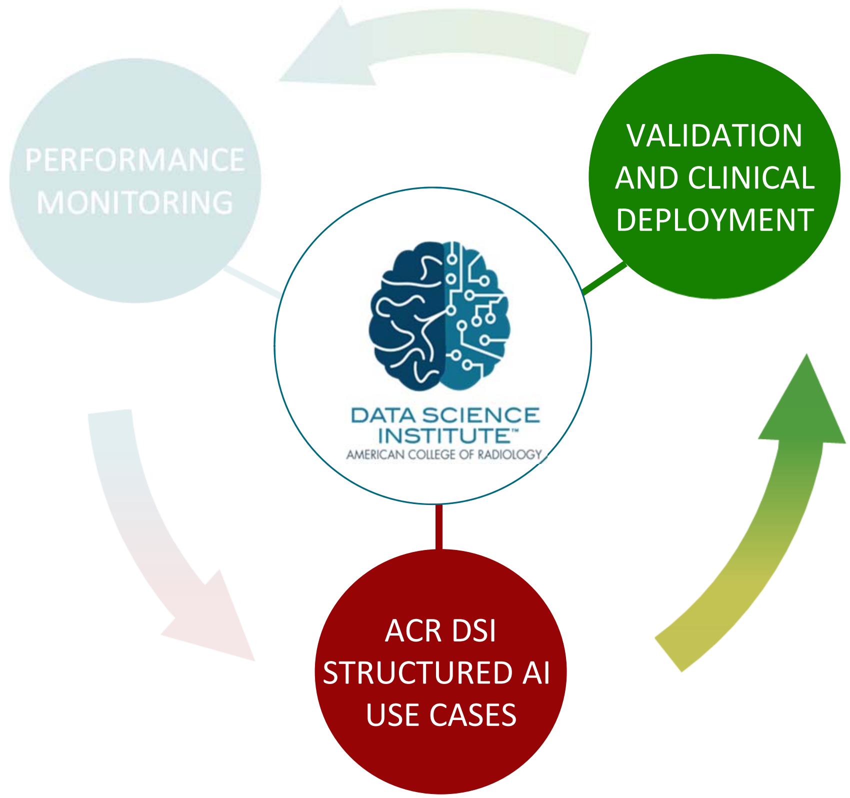


Figure 2: Overlay of FDA's TPLC approach on AI/ML workflow

Artificial Intelligence: Protecting Patients And The Public

How Do We Make Sure AI Is Working In The Real World?



How Do We Validate AI Algorithms For Clinical Practice?

What Are The Most Important Clinical Tasks For AI?

Data Availability For AI Development

- Single institution data
- Diversity and bias

Facial Recognition Underperforms in Women and Women and Men of Color

Original Article | Artificial Intelligence

eISSN 2005-8330
https://doi.org/10.3348/kjr.2019.0025
Korean J Radiol 2019;20(3):405-410

Korean Journal of Radiology
KJR

Check for updates

Design Characteristics of Studies Reporting the Performance of Artificial Intelligence Algorithms for Diagnostic Analysis of Medical Images: Results from Recently Published Papers

Dong Wook Kim, MD[1]*, Hye Young Jang, MD[2]*, Kyung Won Kim, MD, PhD[2], Youngbin Shin, MS[2], Seong Ho Park, MD, PhD[2]

[1]Department of Radiology, Taean-gun Health Center and County Hospital, Taean-gun, Korea; [2]Department of Radiology and Research Institute of Radiology, University of Ulsan College of Medicine, Asan Medical Center, Seoul, Korea

Objective: To evaluate the design characteristics of studies that evaluated the performance of artificial intelligence (AI) algorithms for the diagnostic analysis of medical images.
Materials and Methods: PubMed MEDLINE and Embase databases were searched to identify original research articles published between January 1, 2018 and August 17, 2018 that investigated the performance of AI algorithms that analyze medical images to provide diagnostic decisions. Eligible articles were evaluated to determine 1) whether the study used external validation rather than internal validation, and in case of external validation, whether the data for validation were collected, 2) with diagnostic cohort design instead of diagnostic case-control design, 3) from multiple institutions, and 4) in a prospective manner. These are fundamental methodologic features recommended for clinical validation of AI performance in real-world practice. The studies that fulfilled the above criteria were identified. We classified the publishing journals into medical vs. non-medical journal groups. Then, the results were compared between medical and non-medical journals
Results: Of 516 eligible published studies, only 6% (31 studies) performed external validation. None of the 31 studies adopted all three design features: diagnostic cohort design, the inclusion of multiple institutions, and prospective data collection for external validation. No significant difference was found between medical and non-medical journals.
Conclusion: Nearly all of the studies published in the study period that evaluated the performance of AI algorithms for diagnostic analysis of medical images were designed as proof-of-concept technical feasibility studies and did not have the design features that are recommended for robust validation of the real-world clinical performance of AI algorithms.
Keywords: *Artificial intelligence; Machine learning; Deep learning; Clinical validation; Clinical trial; Accuracy; Study design; Quality; Appropriateness; Systematic review; Meta-analysis*

Validating AI For Clinical Use

- 516 eligible studies from the literature
- 6% performed external validation

THE LANCET
Digital Health

A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis

Xiaoxuan Liu*, Livia Faes*, Aditya U Kale, Siegfried K Wagner, Dun Jack Fu, Alice Bruynseels, Thushika Mahendiran, Gabriella Moraes, Mohith Shamdas, Christoph Kern, Joseph R Ledsam, Martin K Schmid, Konstantinos Balaskas, Eric J Topol, Lucas M Bachmann, Pearse A Keane, Alastair K Denniston

Implications of all the available evidence

Deep learning models achieve equivalent levels of diagnostic accuracy compared with health-care professionals. The methodology and reporting of studies evaluating deep learning models is variable and often incomplete. New international standards for study protocols and reporting that recognise specific challenges of deep learning are needed to ensure quality and interpretability of future studies.

Our review found the diagnostic performance of deep learning models to be equivalent to that of health-care professionals. **However, a major finding of the review is that few studies presented externally validated results or compared the performance of deep learning models and health-care professionals using the same sample. Additionally, poor reporting is prevalent in deep learning studies, which limits reliable interpretation of the reported diagnostic accuracy.**

HOW DO WE MAKE SURE AI IS WORKING IN THE REAL WORLD?

PERFORMANCE MONITORING

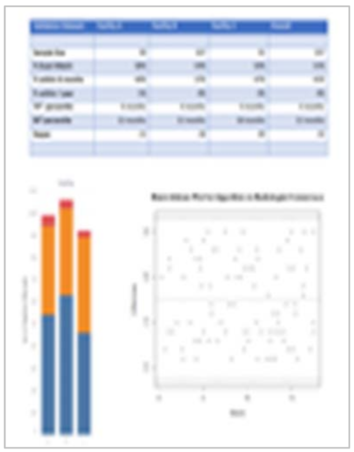
ACR DSI VALIDATION SERVICE

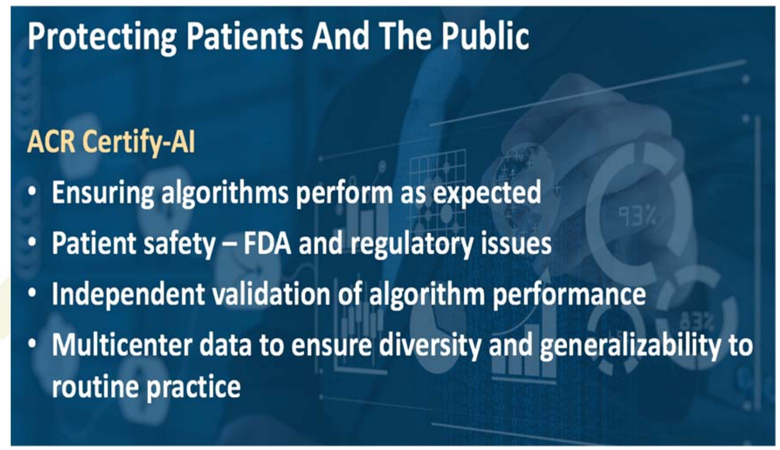
HOW DO WE VALIDATE AI ALGORITHMS FOR CLINICAL PRACTICE?

DATA SCIENCE INSTITUTE™
AMERICAN COLLEGE OF RADIOLOGY

Certify-AI

ACR DSI STRUCTURED AI USE CASES





WHAT ARE THE MOST IMPORTANT CLINICAL TASKS FOR AI?

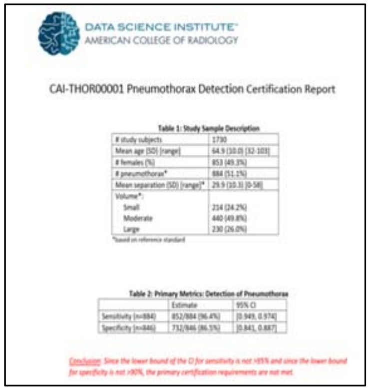
“Premarket Assurance of Safety and Effectiveness”

Algorithm Validation

- Diverse validation data sets
  - Multiple institutions
  - Diverse patient demographics
  - Diverse imaging equipment
- Built according to the use case
- Reasonable costs for developers as compared to reader studies
- Access to diverse data for validation

DATA SCIENCE INSTITUTE™
AMERICAN COLLEGE OF RADIOLOGY

Certify-AI


DATA SCIENCE INSTITUTE™
AMERICAN COLLEGE OF RADIOLOGY

CAI-THOR00001 Pneumothorax Detection Certification Report

Table 1: Study Sample Description

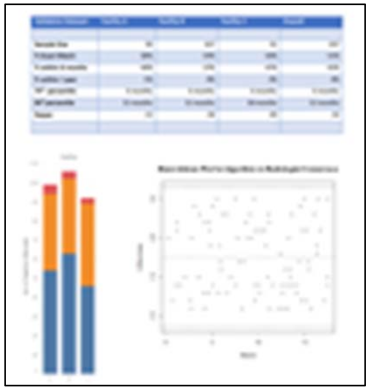
# study subjects	1730
Mean age (SD) [range]	64.9 (10.0) [32-103]
# females (%)	853 (49.3%)
# pneumothorax*	884 (51.1%)
Mean separation (SD) [range]*	29.9 (10.3) [0-58]
Volume*:	
Small	214 (24.2%)
Moderate	440 (49.8%)
Large	230 (26.0%)

*based on reference standard

Table 2: Primary Metrics: Detection of Pneumothorax

	Estimate	95% CI
Sensitivity (n=884)	852/884 (96.4%)	[0.949, 0.974]
Specificity (n=846)	732/846 (86.5%)	[0.841, 0.887]

Conclusion: Since the lower bound of the CI for sensitivity is not >95% and since the lower bound for specificity is not >90%, the primary certification requirements are not met.



"Premarket Assurance of Safety and Effectiveness"

FDA MDDT Program

- "The FDA's Medical Device Development Tools (MDDT) program is a way for the FDA to qualify tools that medical device sponsors can use in the development and evaluation of medical devices"
- "Qualification means that the FDA has evaluated the tool and concurs with available supporting evidence that the tool produces scientifically-plausible measurements and works as intended within the specified context of use"

PNEUMOTHORAX DETECTION

Purpose	Detection of pneumothorax on chest radiograph
Tag(s)	
Panel	Thoracic
Certify-AI ID	CAI-THOR00001

REFERENCE DATASET

Sample Size Requirements	The images from a sample of 1730 subjects is required in order to construct 95% CIs with a precision of ±0.02. Each co-morbidity listed above should be represented by at least 10% of subjects.
Sample Size	1730
# of facilities contributing	12
Reference Standard	Expert review by a panel of 3 radiologists independently interpreting the images in the test set, along with any available follow-up imaging. The majority decision of the 3 radiologists regarding presence/absence of pneumothorax and presence/absence of a chest tube, and the mean of the 3 radiologists' measurements of pleural separation and volume will serve as ground truth.
Prevalence	46% prevalence of suspected pneumothorax

FDA – ACR MDDT Demonstration

FDA Discussion Paper on Continuously Learning Algorithms and the FDA Software Precertification Program

Define-AI

Certify-AI

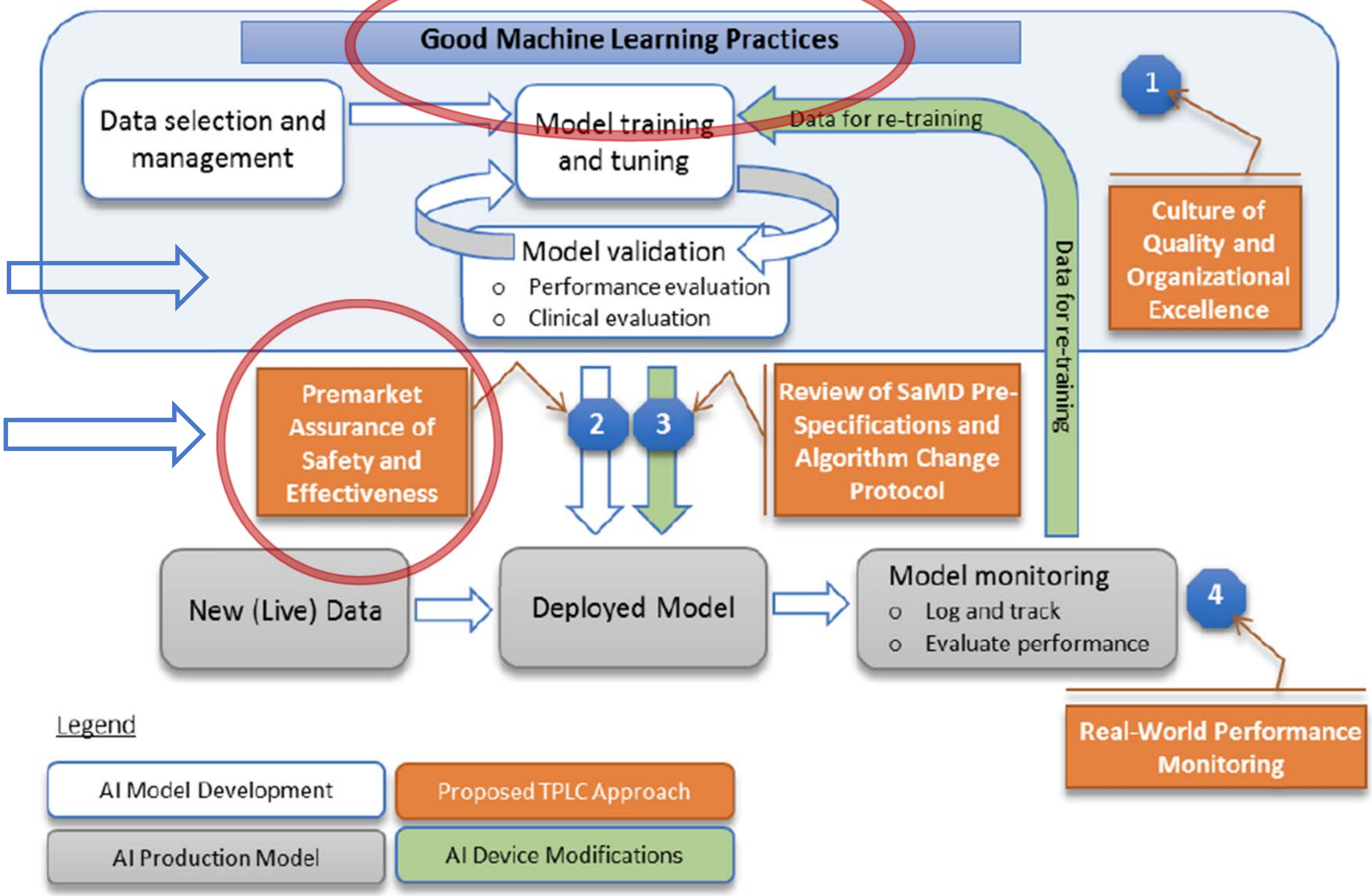
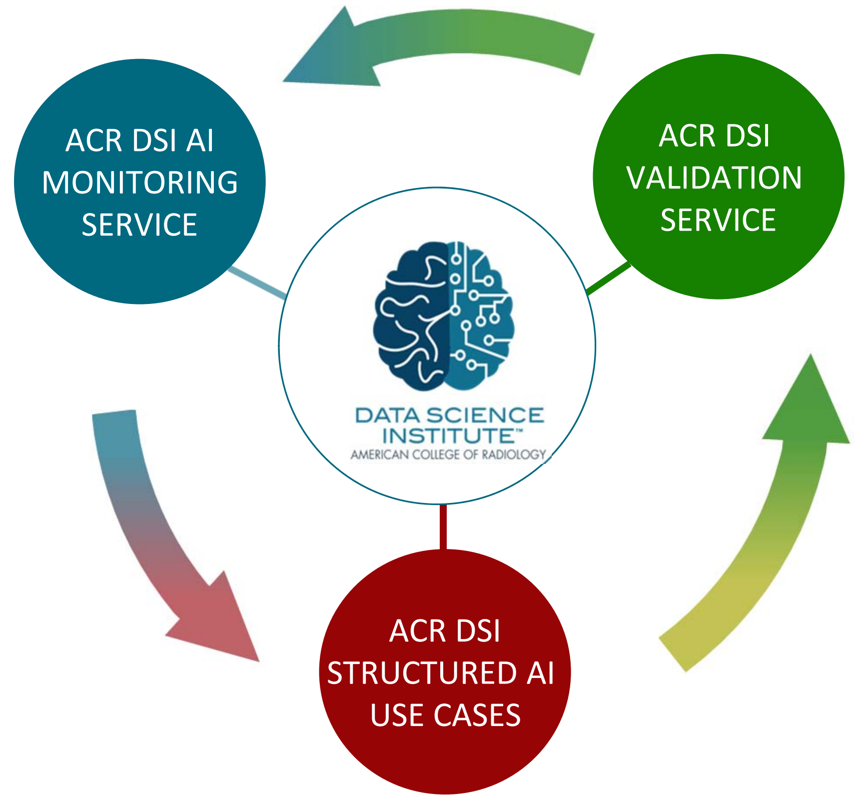


Figure 2: Overlay of FDA's TPLC approach on AI/ML workflow

ARTIFICIAL INTELLIGENCE: CONCEPT TO CLINICAL PRACTICE

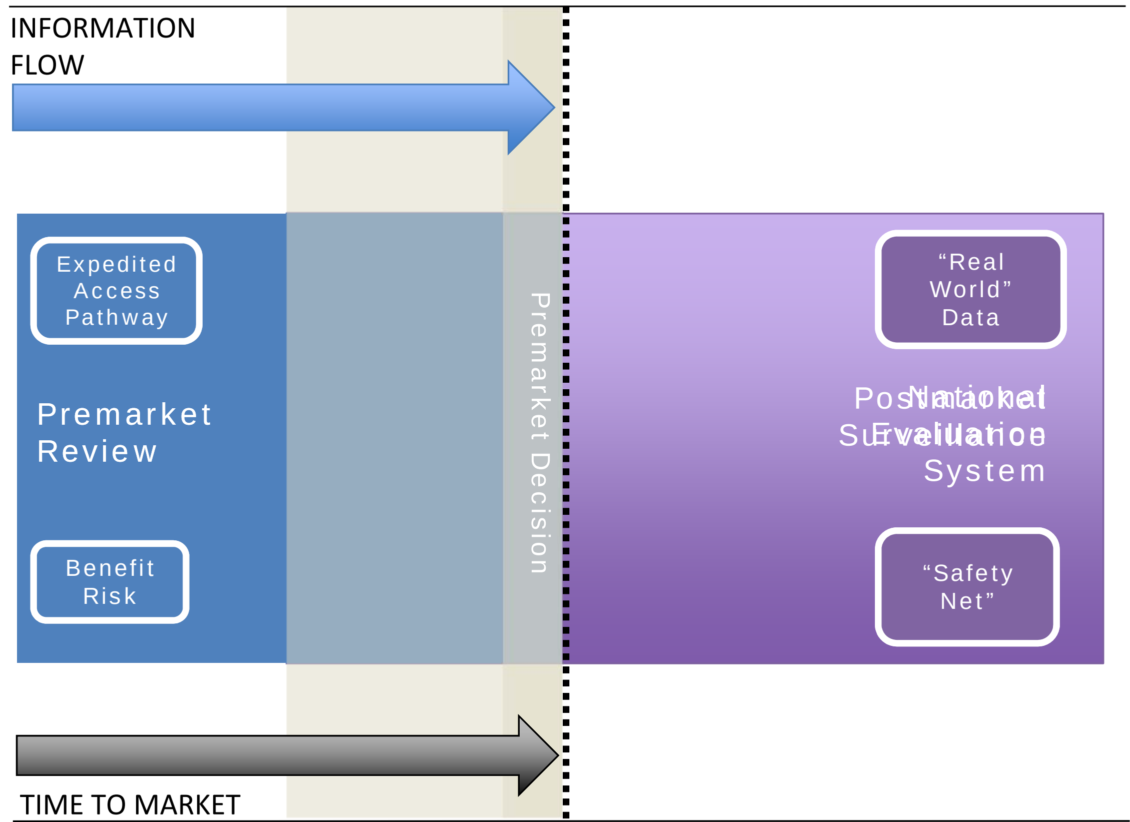
HOW DO WE MAKE SURE AI IS WORKING IN THE REAL WORLD?

HOW DO WE VALIDATE AI ALGORITHMS FOR CLINICAL PRACTICE?



WHAT ARE THE MOST IMPORTANT CLINICAL TASKS FOR AI?

FDA Review And Surveillance – Real World Data



Courtesy Greg Pappas, FDA

HOW DO WE MAKE SURE AI IS WORKING IN THE REAL WORLD?

ACR DSI AI MONITORING SERVICE

ACR DSI VALIDATION SERVICE

HOW DO WE VALIDATE AI ALGORITHMS FOR CLINICAL PRACTICE?

DATA SCIENCE INSTITUTE™
AMERICAN COLLEGE OF RADIOLOGY

Assess-AI

ACR DSI STRUCTURED AI USE CASES

Protecting Patients And The Public

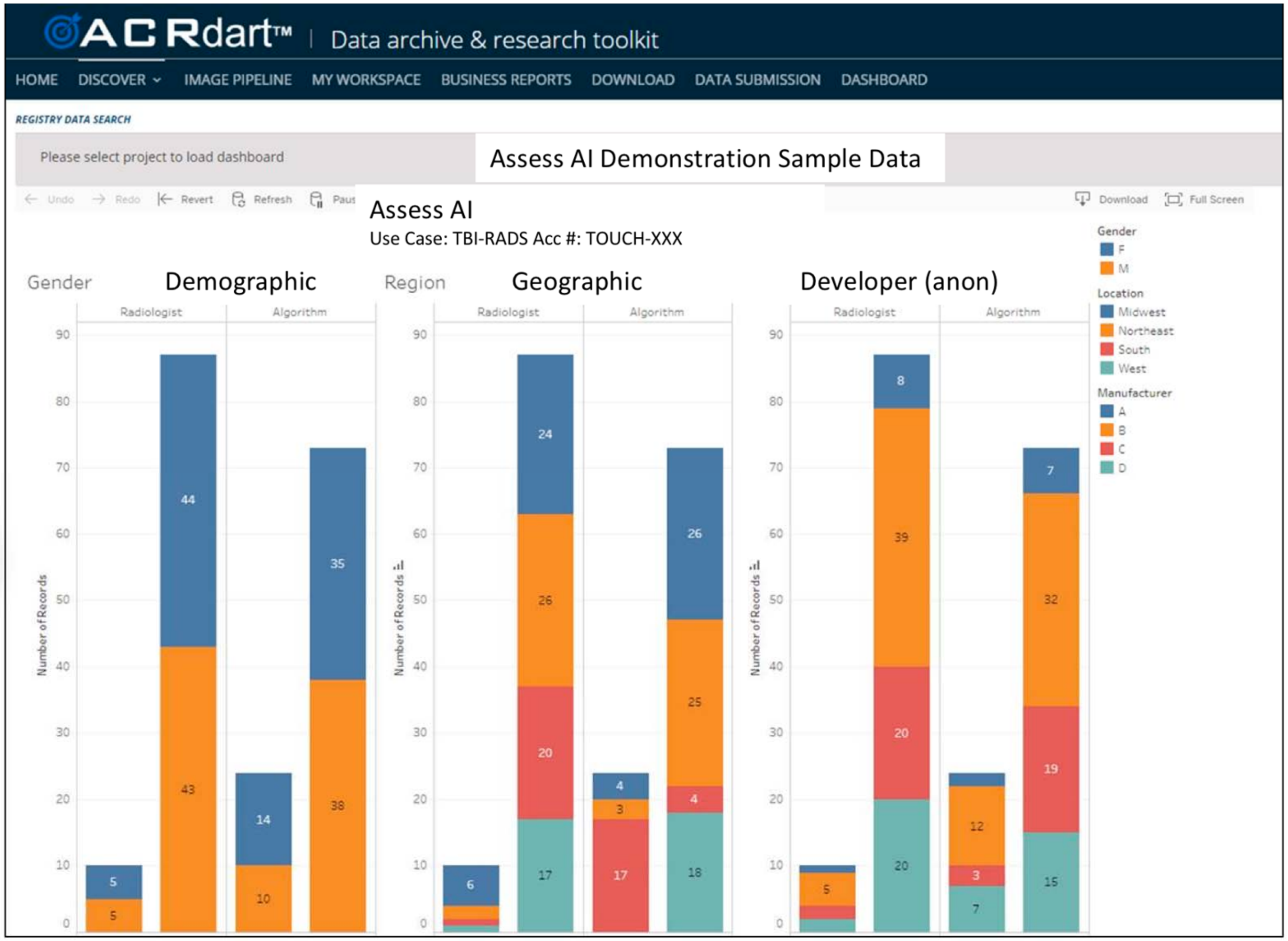
ACR Assess-AI

- Ensuring algorithms perform as expected in clinical practice
- Patient safety – FDA and regulatory issues
- Real world / real time monitoring of algorithm performance
- Radiology professionals providing feedback to developers and regulatory agencies to ensure safe use of AI

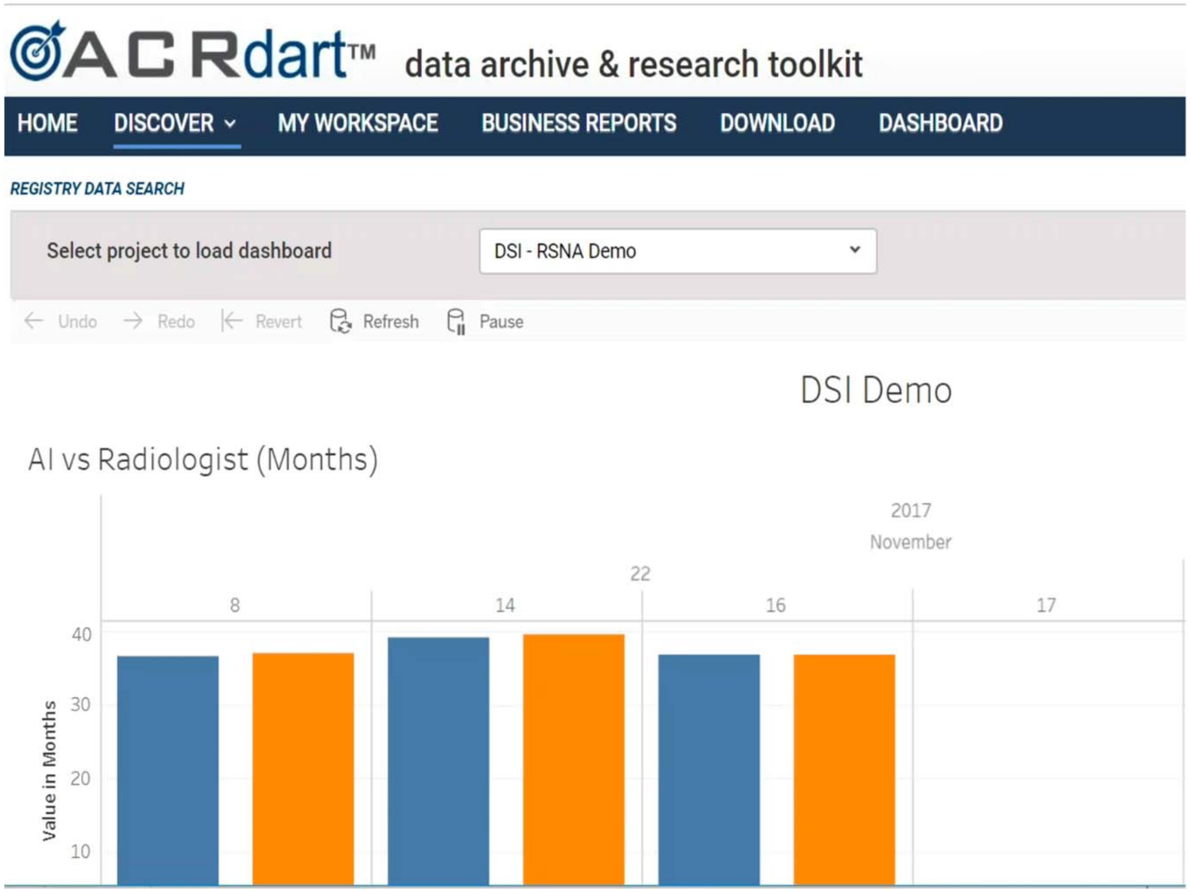
NRDR
AI
ARTIFICIAL INTELLIGENCE REGISTRY™
AMERICAN COLLEGE OF RADIOLOGY
DATA SCIENCE INSTITUTE

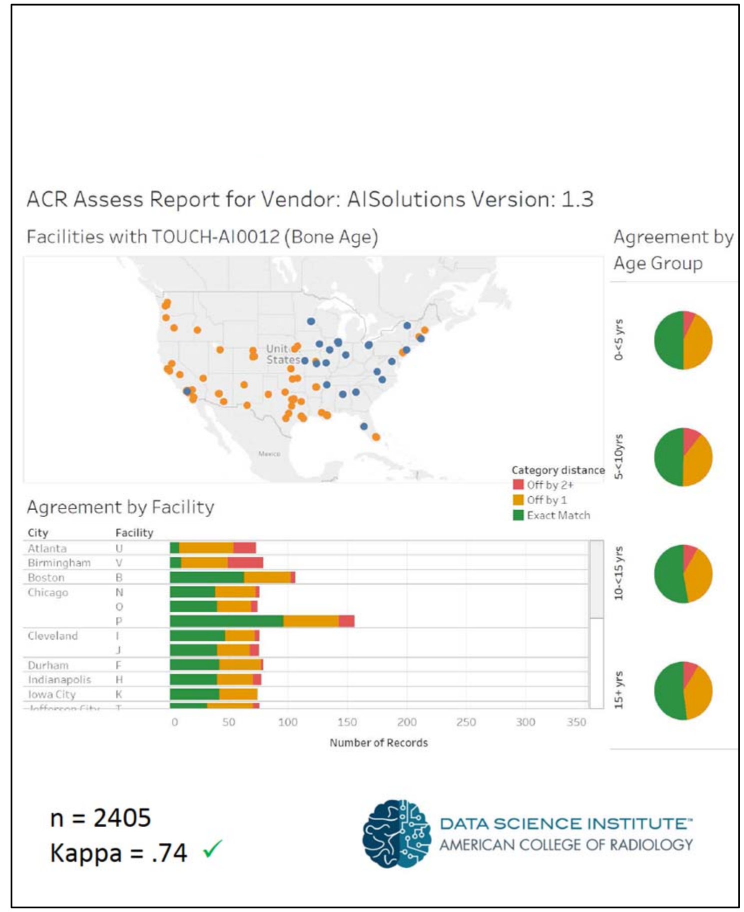
WHAT ARE THE MOST IMPORTANT CLINICAL TASKS FOR AI?

SAMPLE ASSES-AI REPORT



MONITORING ALGORITHM PERFORMANCE IN CLINICAL PRACTICE – REAL WORLD DATA





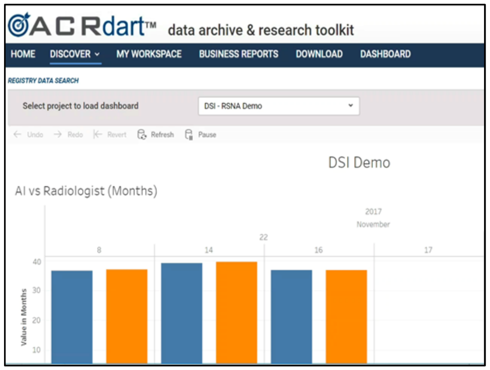
“Real World Performance Monitoring”

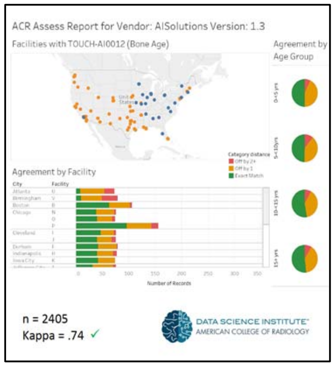
Algorithm Monitoring In Clinical Practice

- AI registries
- Capture algorithm performance from practicing radiologists
- Capture meta-data about the examination
- Feedback to developers / FDA
- Working with FDA to capture data

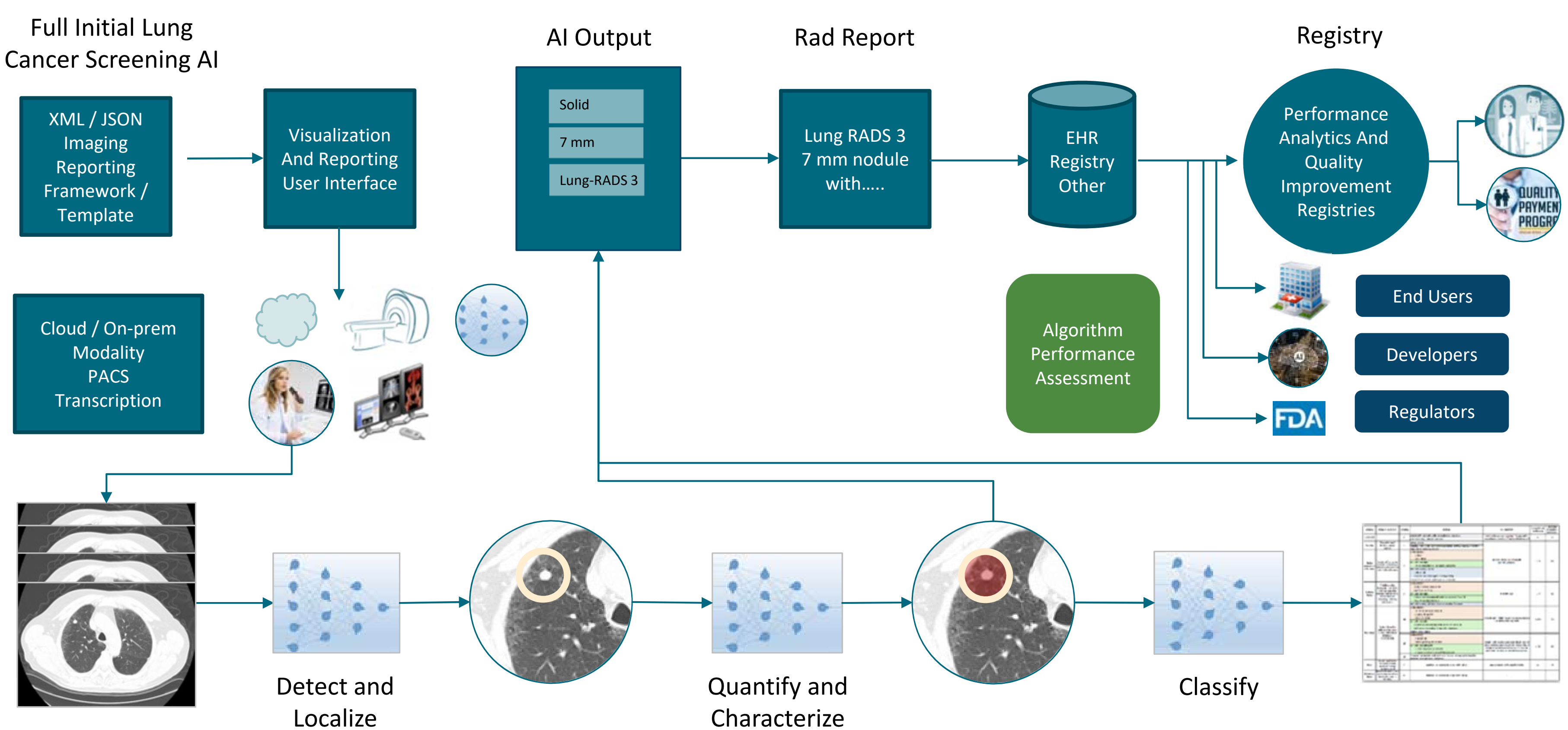
DATA SCIENCE INSTITUTE™
AMERICAN COLLEGE OF RADIOLOGY

Assess-AI



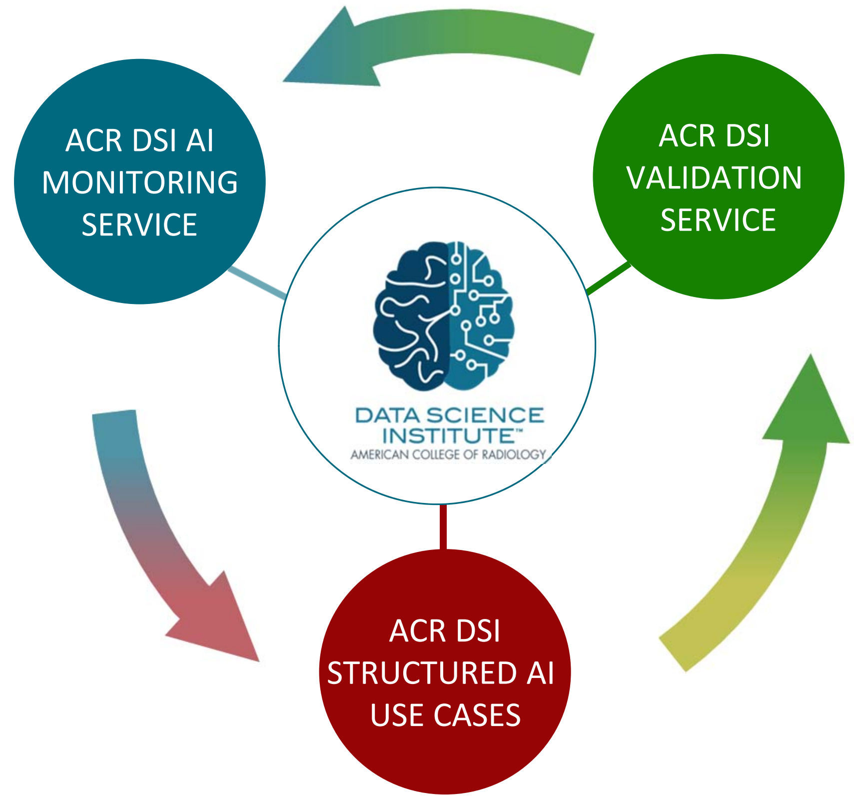


Integrating AI Into Routine Clinical Practice



AI In Clinical Practice With Registry Reporting For Monitoring With Real-World Data

HOW DO WE MAKE SURE AI IS WORKING IN THE REAL WORLD?



HOW DO WE VALIDATE AI ALGORITHMS FOR CLINICAL PRACTICE?

WHAT ARE THE MOST IMPORTANT CLINICAL TASKS FOR AI?

FDA Discussion Paper on Continuously Learning Algorithms and the FDA Software Precertification Program

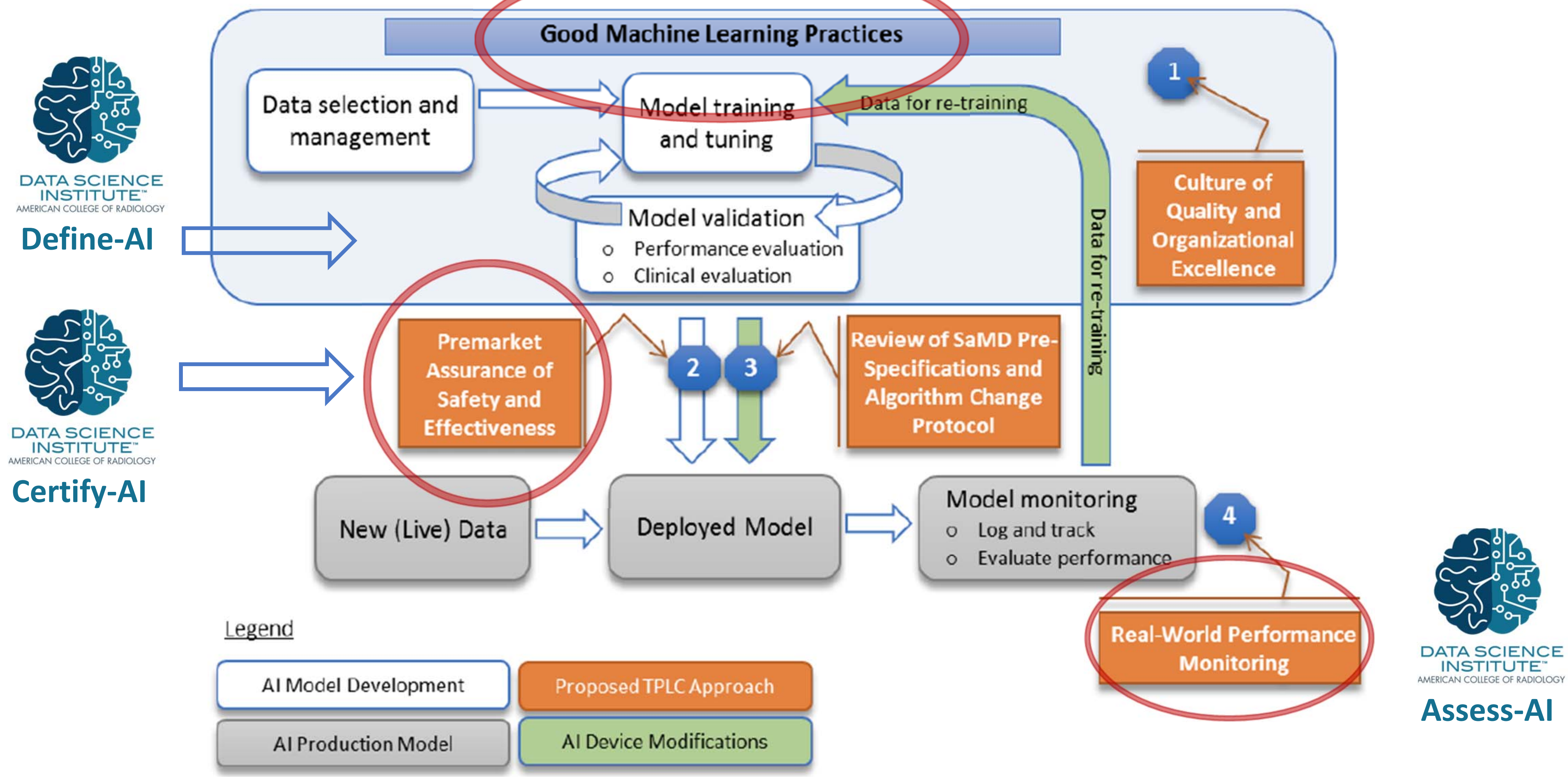


Figure 2: Overlay of FDA's TPLC approach on AI/ML workflow

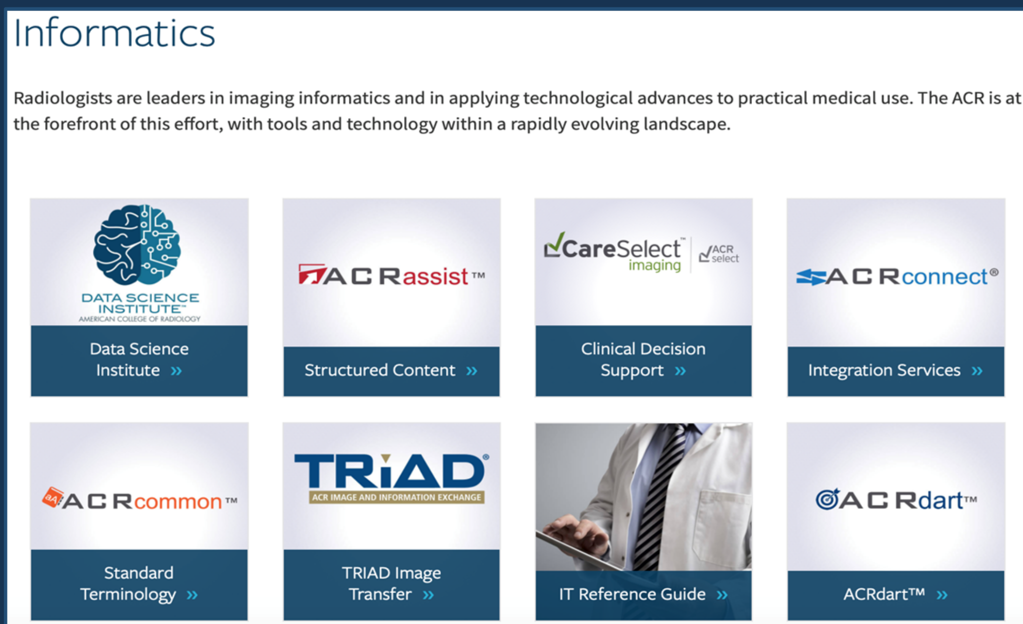
ACR Assets For Implementing The DSI Toolkit

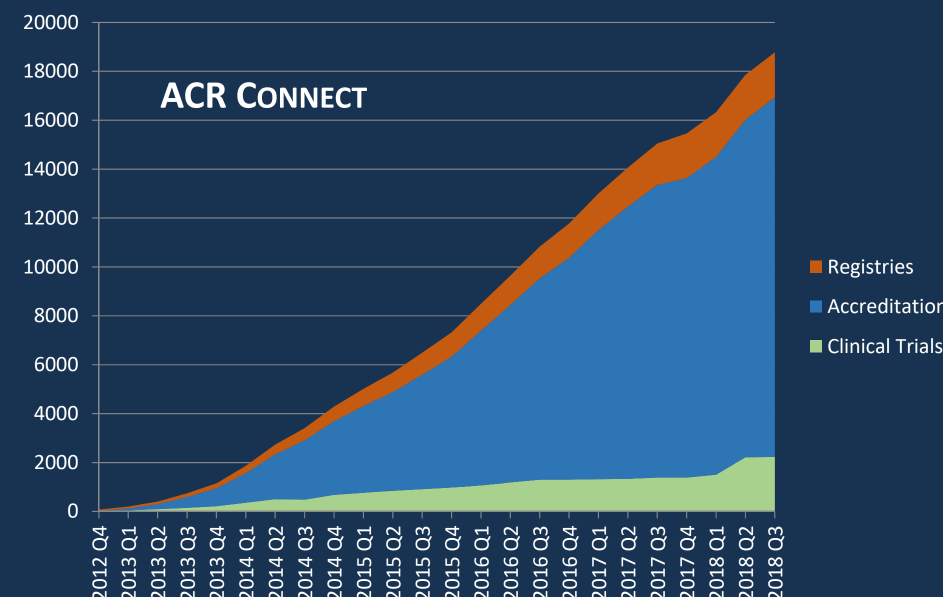
Mission

Resources

Infrastructure

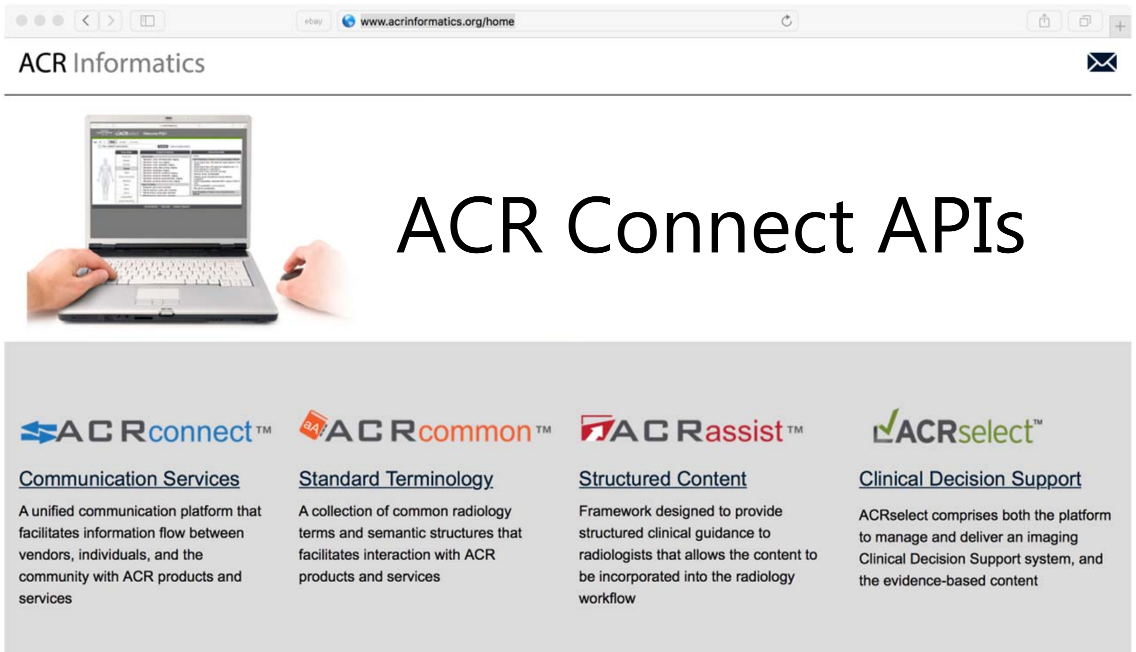
ASSETS FOR ACR TOOLKIT





CONNECTING THE AI ECOSYSTEM

Vendor Neutral Integrations



SIEMENS
SECTRA
INFINITT North America
DoseMonitor
Invivo
GE
PRIMORDIAL
iMALOGIX
MERGE An IBM Company
RADIMETRICS
MEDMYNE
PHILIPS
radiance
MATRIX ANALYTICS
NUANCE
NEXTGEN HEALTHCARE
Jambeyang Research
NVIDIA
Allscripts
MRS Medical Reporting Software
MModal
PenRad Read. Report. Track. Manage.
medinformatix
MCKESSON
SWEARINGEN
AWARE
DOLBEY SINCE 1914
Insight
Cerner
MEDITECH
MagView Healthcare Information Management Solutions


ASSETS FOR ACR TOOLKIT
PATIENT DATA STAYS ON PREM
USE CASE DEVELOPMENT
AI-LAB
DISTRIBUTED VALIDATION
AI EDUCATION
ACR® AMERICAN COLLEGE OF RADIOLOGY
QUALITY IS OUR IMAGE
FEDERATED LEARNING
MODEL EVALUATION
CONNECTING THE AI ECOSYSTEM

Public Private Partnerships