FDA Regulatory Framework and Evaluation Methods for AI/ML-based Decision Tools

Nicholas Petrick. Ph.D.

Division of Imaging, Diagnostics and Software Reliability (DIDSR)
Office of Science and Engineering Labs
Center for Devices and Radiological Health (CDRH)
U.S. Food and Drug Administration (FDA)
nicholas.petrick@fda.hhs.gov

DISCLOSURES

- None

OUTLINE

- Overview of medical device regulatory framework
  - Regulatory overview
  - ML-based medical devices
  - Software as a medical device (SaMD)

- Assessment
  - Imaging-based machine learning (ML) SaMD assessment
  - Framework for assessing ML SaMD modifications

Center for Devices and Radiological Health

- Protect and promote the health of the public by ensuring the safety and effectiveness of medical devices and the safety of radiation-emitting electronic products

Device Class & Pre-Market Requirements

Device Class	Controls	Premarket Review Process
Class I (lowest risk)	General Controls	Most are exempt
Class II	General Controls Special Controls	Premarket Notification [510(k)]
Class III (highest risk)	General Controls Premarket Approval	Premarket Approval [PMA]

Device Class & Pre-Market Requirements

Device Class		Premarket Review Process
Class I (lowest risk)		**Most are exempt**
Class II	**General Controls** **Special Controls**	**Premarket Notification [510(k)]**
Class III (highest risk)	**General Controls** **Premarket Approval**	**Premarket Approval [PMA]**

Demonstrate substantial equivalence to predicate device

Device Class & Pre-Market Requirements

Device Class		Premarket Review Process
Class I (lowest risk)		**Most are exempt**
Class II	**General Controls** **Special Controls**	**De Novo**
Class III (highest risk)	**General Controls** **Premarket Approval**	**Premarket Approval [PMA]**

Means for new device, without a valid predicate, to be classified into Class I or II

Device Class & Pre-Market Requirements

Device Class	Controls	Premarket Review Process
Class I (lowest risk)	General Controls	**Most are exempt**
Class II	Special Controls	**Premarket Notification [510(k)]**
Class III (highest risk)	**General Controls Premarket Approval**	**Premarket Approval [PMA]**

Demonstrate reasonable assurance of safety and effectiveness

Examples of ML-based Medical Software

FDA News Release

FDA permits marketing of clinical decision support software for alerting providers of a potential stroke in patients

February 13, 2018

Viz.Ai

FDA News Release

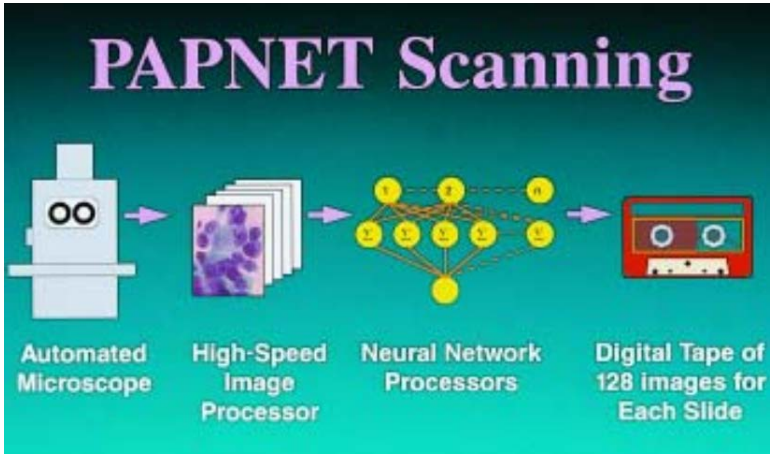
FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems

April 11 2018

IDx-DR

ML-BASED MEDICAL DEVICES ARE NOT NEW

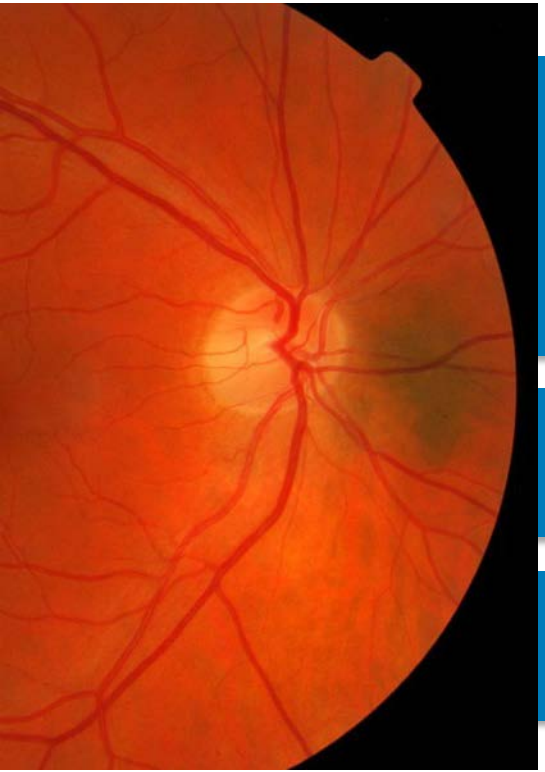
- Mostly imaging or physiological signal analysis applications
  - ECG signal analysis
  - Analysis of radiology images
  - Analysis of cytology/pathology images



- Semi-automated cervical cytology slide reader
- Reduce false-negatives due to human error
- FDA approval in 1994

ML-Based Medical Devices

IDx-DR



Potential to fundamentally transform the delivery of health care:

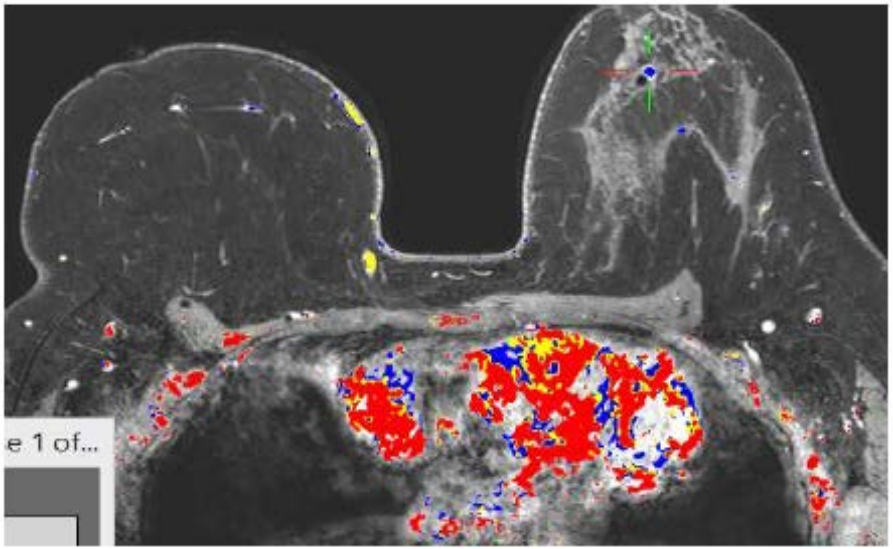
E.g., Earlier disease detection, more accurate diagnosis, new insights into human physiology, personalized diagnostics and therapeutics

Ability for ML to learn from the wealth of real-world data and improve performance

Already seeing ML lead to the development of novel medical devices

ML-Based Medical Devices: Challenges

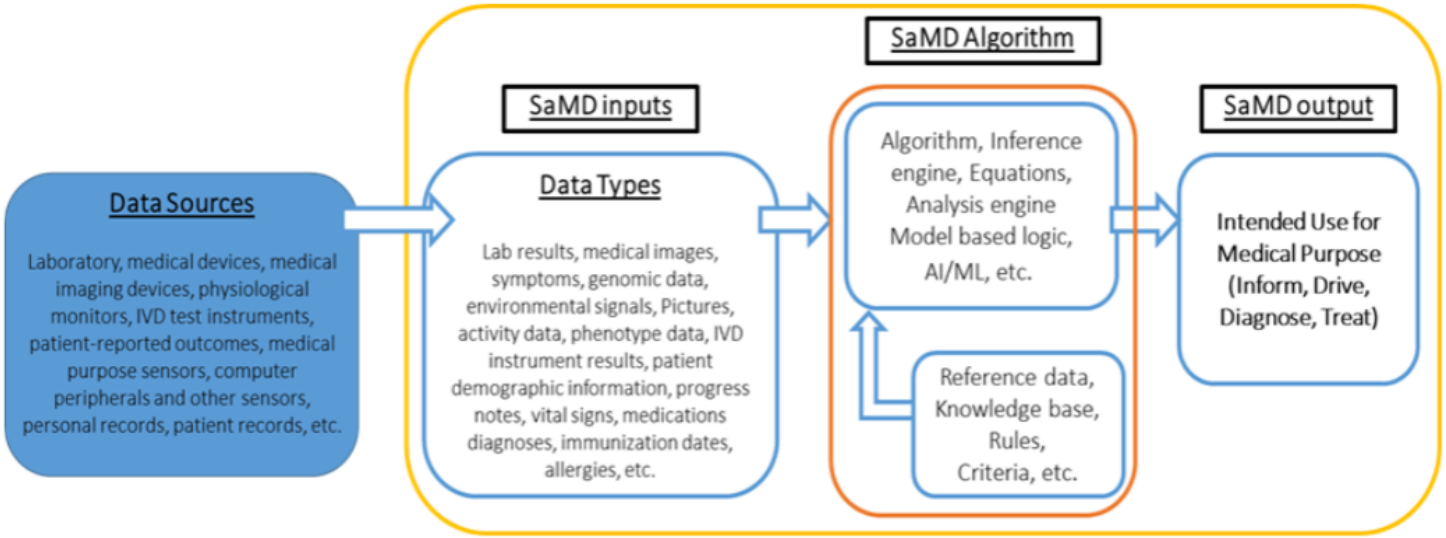
- Need for large, high quality, well-curated data sets
- Explainability of “black box” approaches
- Identifying and removing bias
- Oversight to ML-based algorithms that learn/change over time



QuantX

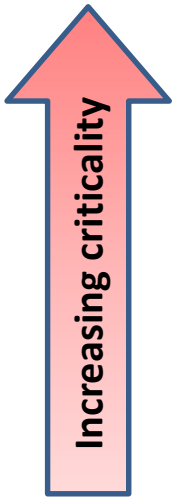
SOFTWARE AS A MEDICAL DEVICE (SAMD)

"Software as a Medical Device" (SaMD) is defined as software intended to be used for one or more medical purposes that perform these purposes without being part of a hardware medical device.[5]



[5] http://www.imdrf.org/workitems/wi-samd.asp

SaMD Risk Categorization



State of Healthcare Situation or Condition	Significance of Information Provided by SaMD to Healthcare Decision		
	Treat or Diagnose	Drive Clinical Management	Inform Clinical Management
Critical	IV	III	II
Serious	III	II	I
Non-Serious	II	I	I

[5]http://www.imdrf.org/workitems/wi-samd.asp

IMDRF International Medical Device Regulators Forum

Fundamentals of image-based ML Assessment

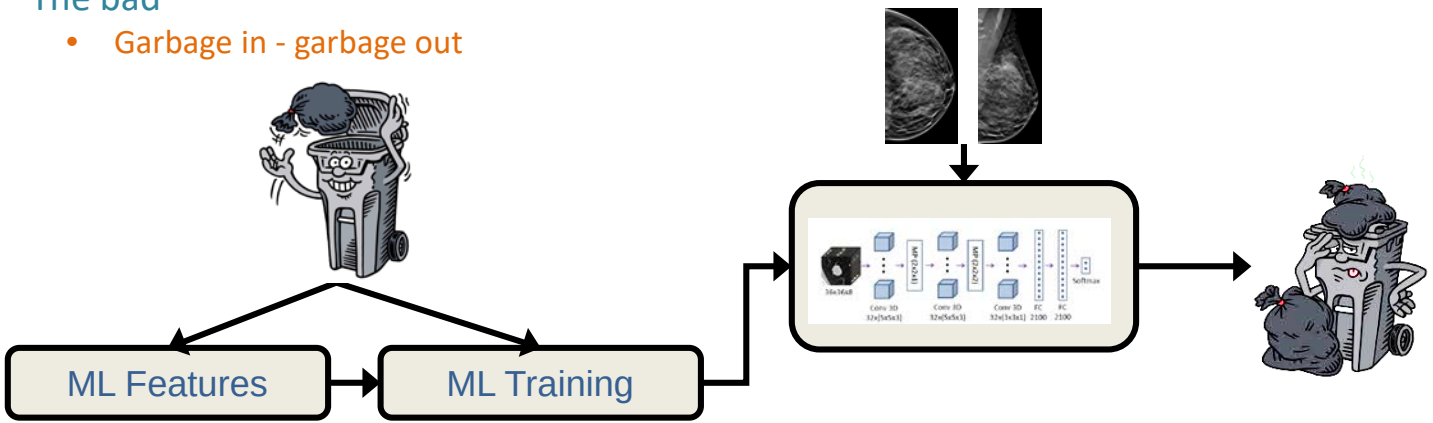
- Device description
- Data
- Performance assessment
  - Standalone performance
  - Reader performance (when appropriate)
  - …
- Human factors or other information/testing as appropriate
- …

Device Description

- Device & algorithm descriptions
  - Device usage (mode of operation, patient population, …)
  - Algorithm design and function
    - Including structure of traditional and deep learning networks
    - Inputs
      - Type and range of signals/data
    - Outputs
  - Training process
  - Training/test database
  - Reference standard
  - …

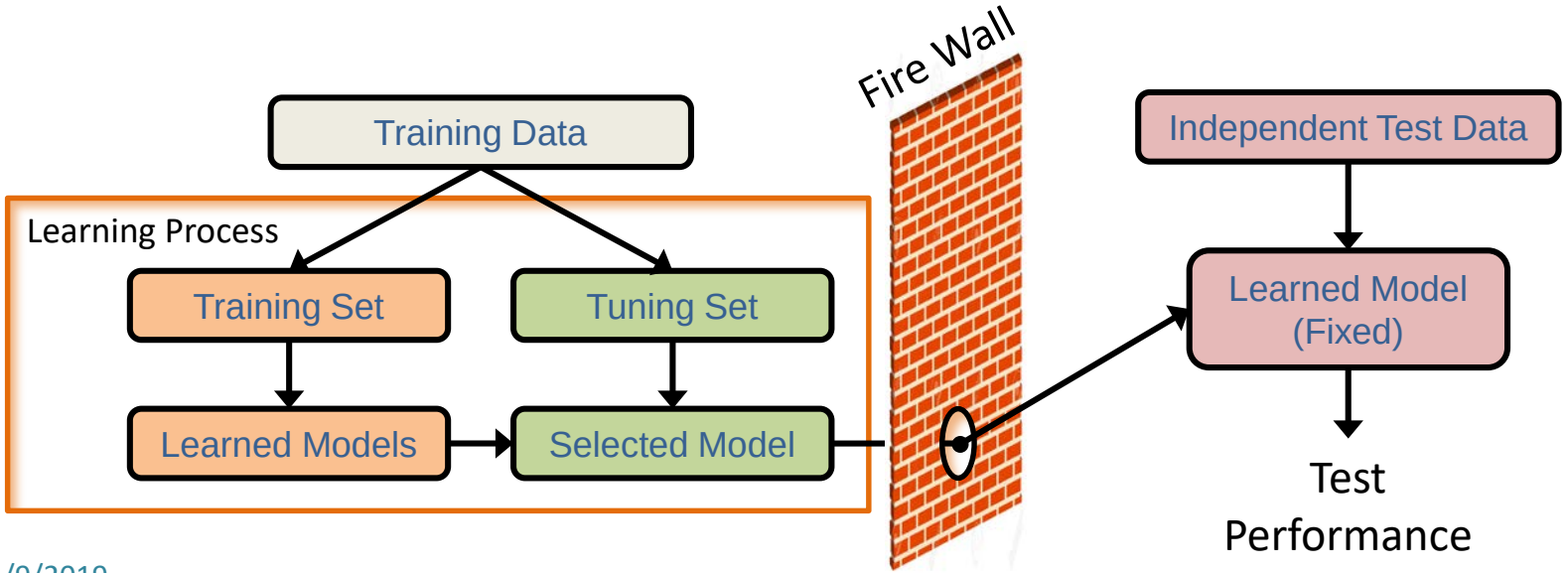
DATA

- ML algorithms are data-driven
  - Versus, for example, physics or biology based
- ML algorithm development now facilitated by standardized ML platforms
  - Brings ML to a wider array of users
  - The good
    - Access to high-quality data streamlines design of novel ML applications
  - The bad
    - Garbage in - garbage out



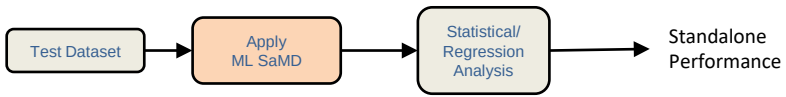
Performance Testing

- Performance of ML algorithm on an independent data
  - Ideally, identifies problems with training process

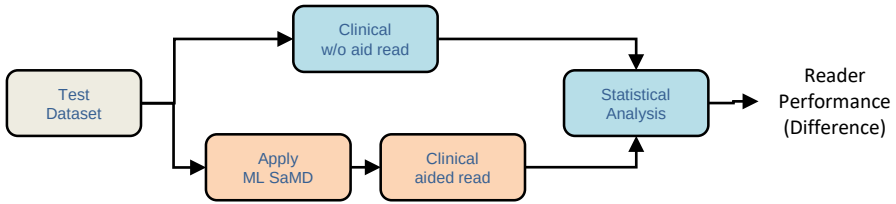


PERFORMANCE TESTING

- Standalone performance
  - Performance of algorithm alone
  - Assesses robustness and generalizability of algorithm



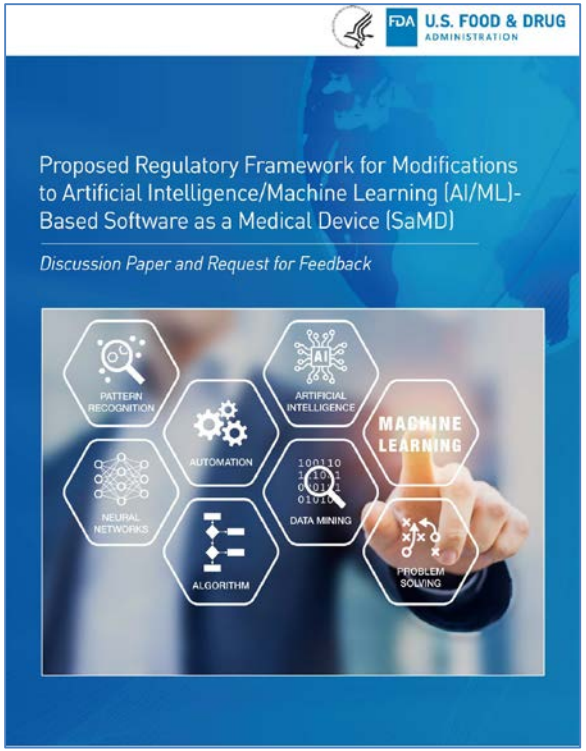
- Clinical reader performance
  - Assessment of clinical aids
  - Clinicians' performance utilizing device
    - Multi-reader multi-case designs
    - Compare clinician's performance with the ML SaMD aid to without the aid



Proposed Regulatory Framework for AI/ML Algorithms Modifications

- Agency proposing framework to give manufacturers option to submit a plan for AI/ML modifications during initial premarket review

- Initial premarket phase would include
  - Review initial SaMD performance
  - Review plan for modifications
  - Review ability to manage/control resultant risks of modifications
- FDA asking for community feedback on this document



https://www.regulations.gov/document?D=FDA-2019-N-1185-0001

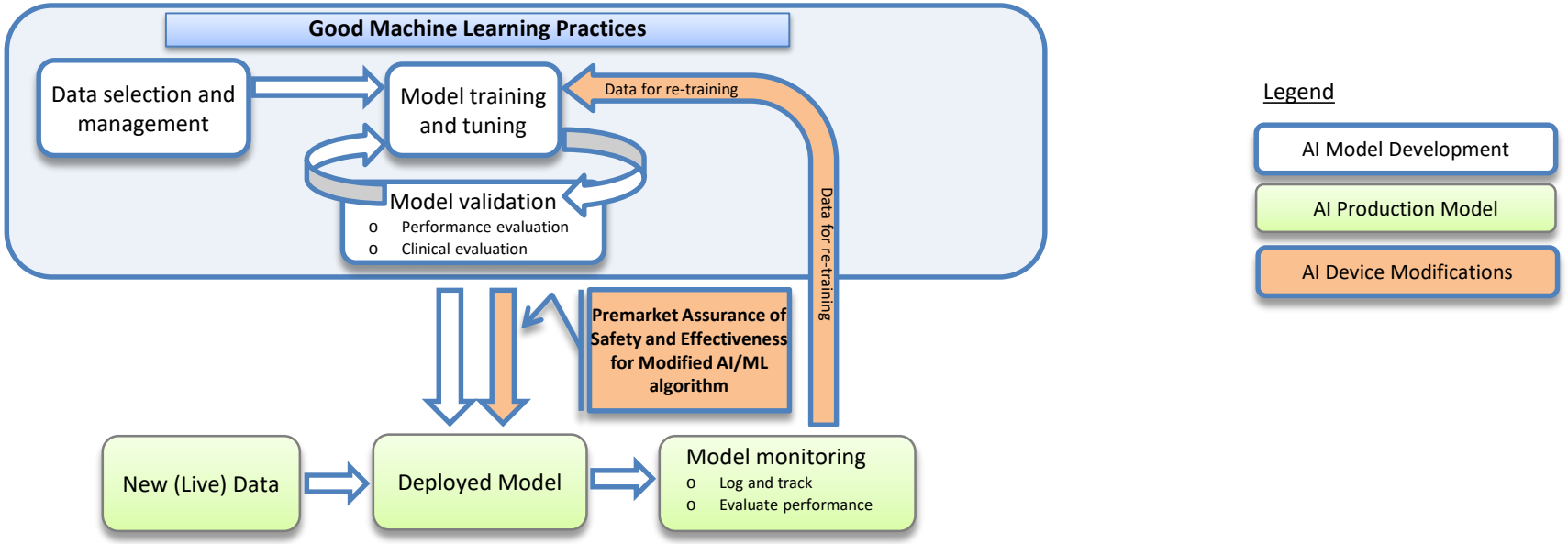
Components of Change Control Plan

- Good ML Practices (GMLP):
  - Accepted practices in AI/ML algorithm design, development, training, and testing that facilitate the quality development and assessment of AI/ML-based algorithms
    - Based on concepts from quality systems, software reliability, machine learning, and data analysis, etc.

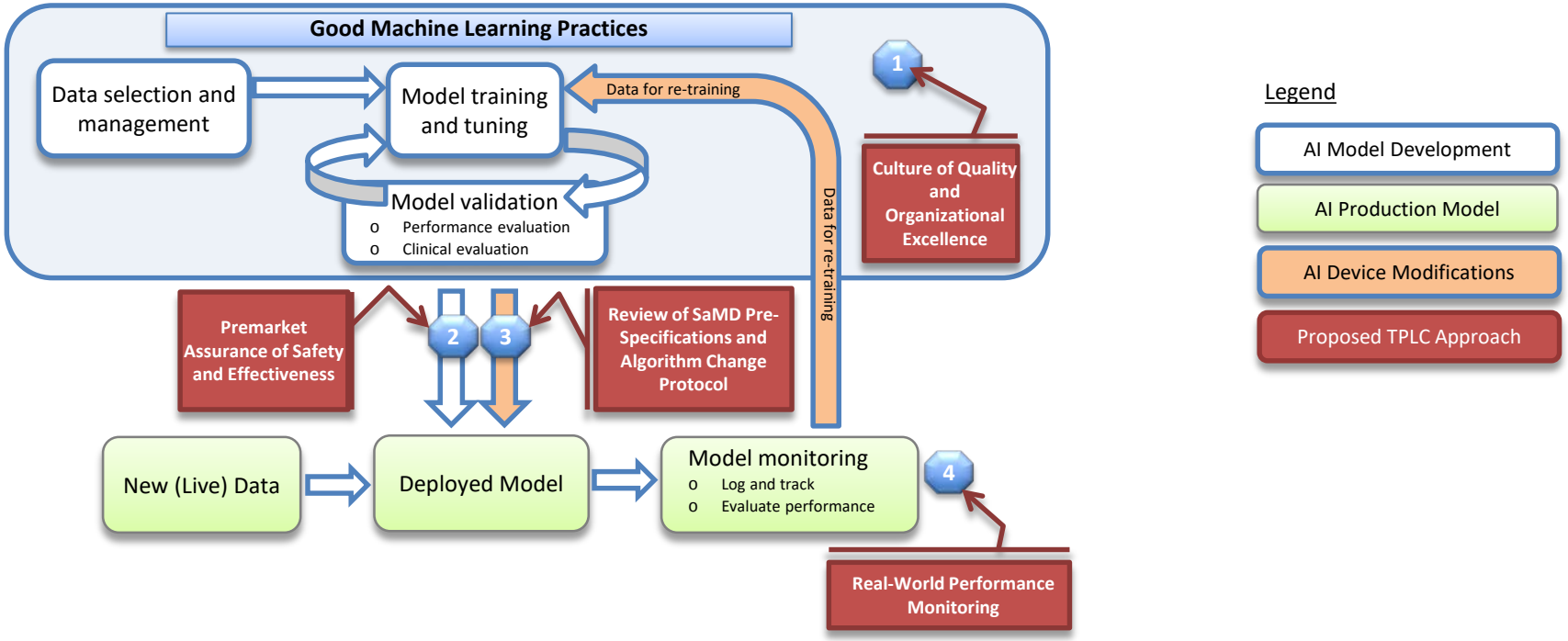
- SaMD Pre-Specifications (SPS):
  - Delineates the proposed types of modifications to the SaMD (i.e., what types of changes the sponsor plans to achieve)
    - Determine “range of potential changes” around the initial specifications and labeling of original device

- Algorithm Change Protocol (ACP):
  - Describes the methods for performing and validating the changes pre-specified in SPS (i.e. how the sponsor intends to achieve the changes)
    - Typically specific to the device and type of change
    - Expected to contain a step-by-step delineation of the procedures to be followed

CURRENT AI/ML WORKFLOW



Proposed TPLC Approach Overlayed on AI/ML Workflow



AI-AI Questions

1. How can current AI technology, particularly convolutional neural networks (CNN), be enhanced? Are there any gaps or roadblocks?
2. What additional tools beyond AI do we need …?
   - Develop data models & robust study designs for assessing standalone & clinical performance of AI/ML algorithms
3. How should the curated data sets be developed to optimize specific tasks of today and tomorrow?
   - Concurrently maximize impact and durability of datasets
4. What are the models for scaling up and sustainment? …
   - Develop data, study design and statistical methods for assessing AI/ML algorithm modifications
   - Determine elements for software pre-specifications (SPS) & algorithm change protocol (ACP)
   - Determine role of real-world role evidence in:
     - Assessing benefit of AI/ML algorithms
     - Supporting transparency to end user

DIDSR AI/ML Research

PI Name	Project Title
Chen	Technical and statistical assessment of AI/ML in digital pathology for clinical deployment
Gallas	High-throughput truthing of microscope slides to validate AI algorithms analyzing digital scans of same slides
Gavrielides	Improving pathologist performance for diagnosing ovarian cancer histological subtypes using ML
Glick	Development of deep learning model observer to assess x-ray breast imaging systems
Li and Petrick	Vascular calcium and material characterization in women using dual-energy CT (quantitative imaging biomarkers)
Pezeshk	Recurrent conv. networks for nodule detection in thoracic CT scans
Pezeshk, Graff	Comparison of quality assessment methods for deep-learning-based MR image reconstruction
Pezeshk	Assessment of AI systems that use un-annotated or weakly labeled datasets in training
Sahiner, Cha	Use of synthetic data in deep learning algorithm training and test
Sahiner, Chen	Leveraging imperfect post-market reference indices for the evaluation of artificial intelligence and machine learning devices
Sahiner, Gossmann	Assessment of adaptive machine learning systems: Methods for re-use of holdout sets and application to deep learning systems for medical image and physiological signal analysis
Zeng, Graff	Deep learning-based image reconstruction and denoising in radiological imaging

ACKNOWLEDGMENTS

- I’ d like to acknowledge Berkman Sahiner and Matthew Diamond for their help in developing this presentation


FDA
U.S. FOOD & DRUG
ADMINISTRATION