AI SUMMIT

CINCINNATI, OH • NOVEMBER 14–16, 2023

Artificial Intelligence Transparency and Explainability

Shawn Forest

Digital Health Specialist, FDA

Mike Salem

Associate Director of Data Science, Gilead Sciences

Transparency for Artificial Intelligence/ Machine Learning (AI/ML)-Enabled Medical Devices

Shawn Forrest
Digital Health Specialist, Center for Devices & Radiological Health (CDRH)
CDRH Digital Health Center of Excellence, US FDA

Outline

- Artificial Intelligence/ Machine Learning (AI/ML)-Enabled Medical Devices
- FDA AI/ML Action Plan
  - Good Machine Learning Practice
  - Transparency

Outline

- **Artificial Intelligence/ Machine Learning (AI/ML)-Enabled Medical Devices**
- FDA AI/ML Action Plan
  - Good Machine Learning Practice
  - Transparency

AI/ML-Enabled Medical Devices:
International Harmonization on Terminology

Artificial Intelligence (AI):
A branch of computer science, statistics, and engineering that uses algorithms or models to perform tasks and exhibit behaviors such as learning, making decisions and making predictions.

Machine Learning (ML):
A subset of AI that allows computer algorithms to learn through data, without being explicitly programmed, to perform a task.

AI/ML-Enabled Medical Device:
A medical device that uses machine learning to achieve its intended medical purpose.

Adapted from IMDRF Artificial Intelligence Medical Devices Key Terms & Definitions proposed document

FDA Resource on AI/ML-Enabled Medical Devices

FDA

Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices

Export Excel   Show 50 entries

Date of Final Decision	Submission Number	Device	Company	Panel (Lead)	Primary Product Code
07/27/2023	K231195	Brainomix 360 Triage ICH	Brainomix Limited	Radiology	QAS
07/26/2023	K231038	Global Hypoperfusion Index (GHI) Algorithm	Edwards Lifesciences, LLC	Cardiovascular	QNL
07/25/2023	K223473	ME-APDS™; MAGENTIQ-COLO™	Magentiq Eye LTD	Gastroenterology/Urology	QNP
07/25/2023	K230365	Sonio Detect	Sonio	Radiology	IYN
07/25/2023	K230913	ANDI	Imeka Solutions, Inc.	Radiology	QIH
07/24/2023	K223347	UltraSight AI Guidance	UltraSight Inc	Radiology	QJU
07/21/2023	K230150	OptimMRI	RebrAIn, SAS	Radiology	QIH

https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices

AI/ML-Enabled Medical Devices: Opportunities & Challenges

FDA

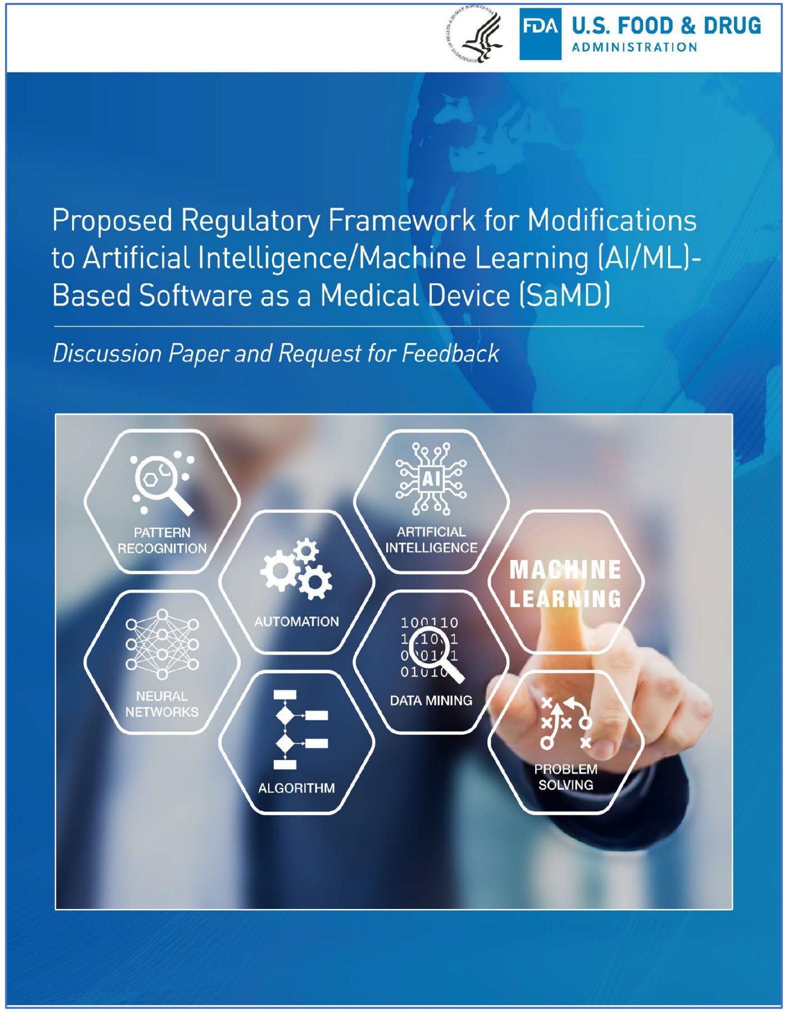
OPPORTUNITIES

- **Significant positive impact on health care**
  - Earlier disease detection
  - More accurate diagnosis
  - New insights into human physiology
  - Personalized diagnostics and therapeutics
- **Applications across all medical fields**
- **Ability to learn, adapt, and improve performance**

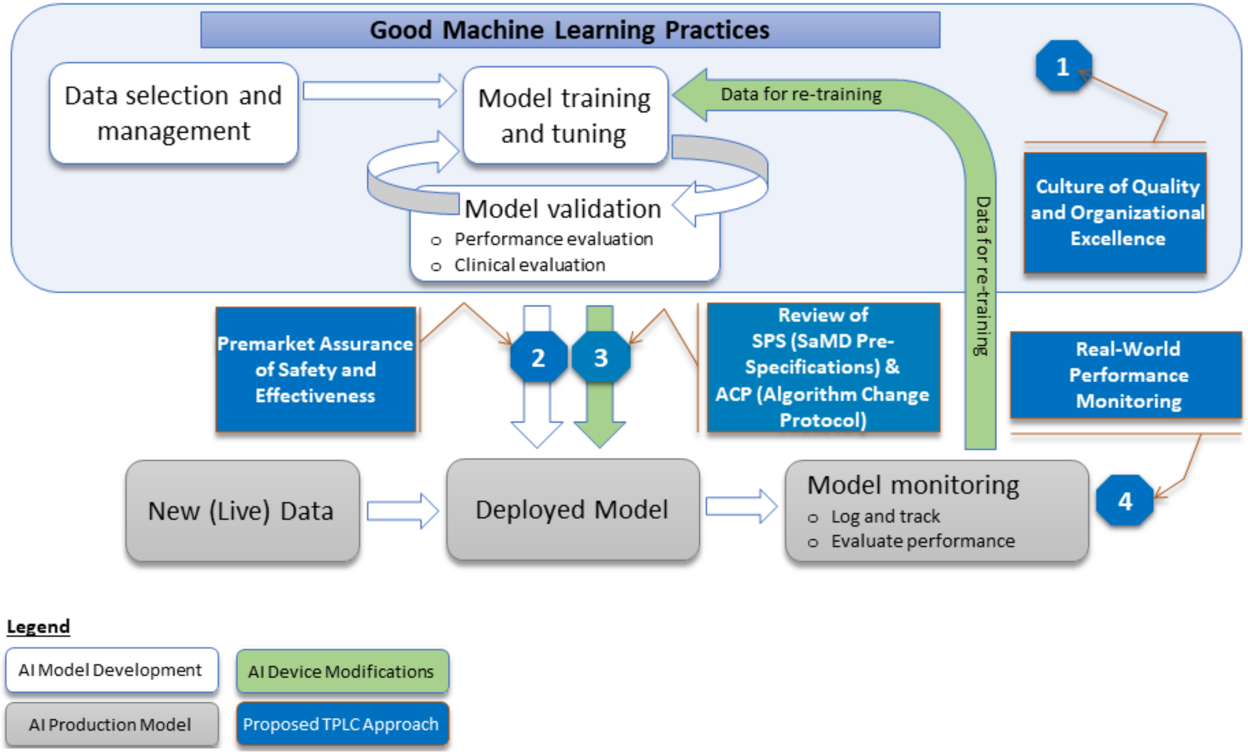
CHALLENGES

- **Fit-for-purpose data sets for development and testing, including diversity**
- **Identification and minimization of bias**
- **Opacity of some algorithms**
- **Providing oversight for an adaptive system**
- **Ensuring transparency to users**

Proposed Regulatory Framework for AI/ML-Enabled Device Software



Overlay of FDA's TPLC Approach on AI/ML Workflow



Continuing our Collaborative Approach

FDA

2019	2020	2021	2022	2023
• Published AI/ML-SaMD Discussion Paper • First joined Collaborative Community related to AI/ML	• Public Workshop on AI/ML in Radiological Imaging • Patient Engagement Advisory Committee Meeting on Patient Trust in AI/ML Devices	• Published AI/ML Medical Device Software Action Plan • Posted List of Currently Marketed AI/ML Devices • Public Workshop on Transparency of AI/ML Devices • Published Good Machine Learning Practice Principles	• Contributed to IMDRF's Key Terms & Definitions: Machine Learning Enabled Medical Devices • Published Clinical Decision Support (CDS) Final Guidance • Updated List of Currently Marketed AI/ML Devices • Recognized new Consensus Standard on AI/ML	• Published Predetermined Change Control Plan for AI/ML Devices Draft Guidance • Updated List of Currently Marketed AI/ML Devices • Published Pre-determined Change Control Plans for Machine Learning-Enabled Medical Devices Guiding Principles

We recognize that by working collaboratively with stakeholders we can lay out a clear path toward building a proactive patient-centered approach to the development and use of AI/ML-enabled devices.

Outline

- Artificial Intelligence/ Machine Learning (AI/ML)-Enabled Medical Devices
- **FDA AI/ML Action Plan**
  - Good Machine Learning Practice
  - Transparency

Stakeholder Feedback on AI/ML Approach

What we heard, and what we'll do

What we heard from stakeholders:

1. **Regulatory Framework**: Requested further development of proposed regulatory framework for AI/ML-based SaMD
2. **Good Machine Learning Practices (GMLP)**: Supported the idea of GMLP and the need for harmonization of its efforts
3. **Transparency**: Asked for further discussion with FDA on how these technologies interact with people, including transparency to users
4. **Regulatory Science**: Described need for improved methods related to algorithmic bias and robustness.
5. **Real-World Performance (RWP)**: Sought clarity on RWP monitoring for AI/ML software.

Stakeholder Feedback on AI/ML Approach

What we heard, and what we'll do

What we heard from stakeholders:

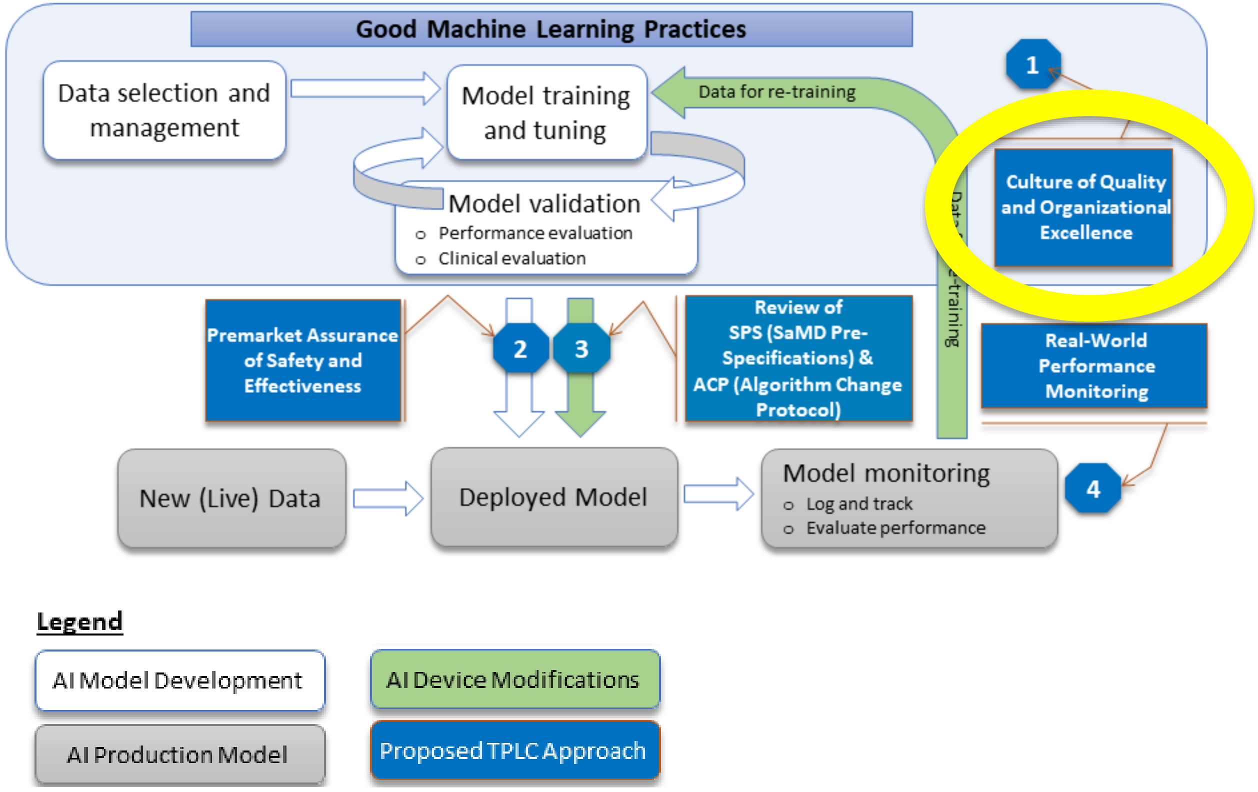
1. **Regulatory Framework**: Requested further development of proposed regulatory framework for AI/ML-based SaMD
2. **Good Machine Learning Practices (GMLP)**: Supported the idea of GMLP and the need for harmonization of its efforts
3. **Transparency**: Asked for further discussion with FDA on how these technologies interact with people, including transparency to users
4. **Regulatory Science**: Described need for improved methods related to algorithmic bias and robustness.
5. **Real-World Performance (RWP)**: Sought clarity on RWP monitoring for AI/ML software.

What we'll do -- The AI/ML Action Plan:

1. **Update the proposed AI/ML framework**, including through Guidance
2. **Strengthen FDA's role in harmonizing GMLP** through standards development & other initiatives
3. **Foster a patient-centered approach**, starting with a workshop on transparency to users
4. **Support development of regulatory science methods** related to algorithm bias and robustness
5. **Advance real-world performance pilots** in coordination with stakeholders and other programs

Good Machine Learning Practice (GMLP)

- Accepted practices in ML/AI algorithm design, development, training, and testing that facilitate the quality development and assessment of ML/AI-enabled devices
- Based on concepts from quality systems, software reliability, machine learning, and data analysis

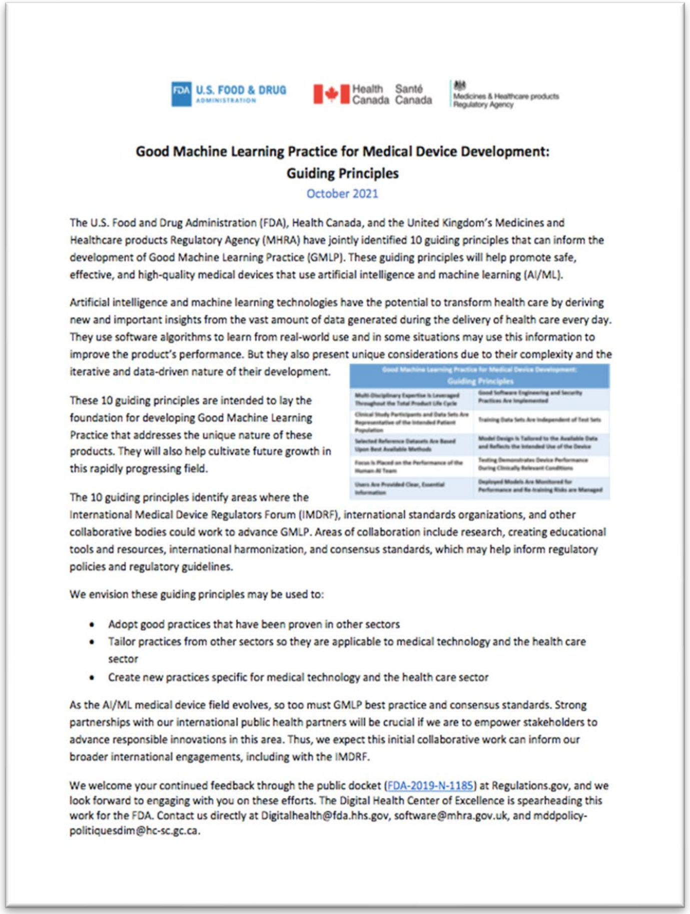


Overlay of FDA's TPLC approach on AI/ML workflow

Adapted from Proposed Regulatory Framework for Artificial Intelligence/Machine Learning (AI/ML)-Based SaMD

Good Machine Learning Practice (GMLP) Guiding Principles

- Ten guiding principles issued by US FDA, MHRA (UK) and Health Canada
- Intended to help inform the development of GMLP and encourage broad stakeholder engagement
- Promotes global harmonization in efforts for the identification of best practices and the creation of standards
- Being considered by broader IMDRF international community



U.S. FOOD & DRUG ADMINISTRATION | Health Canada Santé Canada | Medicines & Healthcare products Regulatory Agency

Good Machine Learning Practice for Medical Device Development: Guiding Principles

October 2021

The U.S. Food and Drug Administration (FDA), Health Canada, and the United Kingdom's Medicines and Healthcare products Regulatory Agency (MHRA) have jointly identified 10 guiding principles that can inform the development of Good Machine Learning Practice (GMLP). These guiding principles will help promote safe, effective, and high-quality medical devices that use artificial intelligence and machine learning (AI/ML).

Artificial intelligence and machine learning technologies have the potential to transform health care by deriving new and important insights from the vast amount of data generated during the delivery of health care every day. They use software algorithms to learn from real-world use and in some situations may use this information to improve the product's performance. But they also present unique considerations due to their complexity and the iterative and data-driven nature of their development.

These 10 guiding principles are intended to lay the foundation for developing Good Machine Learning Practice that addresses the unique nature of these products. They will also help cultivate future growth in this rapidly progressing field.

The 10 guiding principles identify areas where the International Medical Device Regulators Forum (IMDRF), international standards organizations, and other collaborative bodies could work to advance GMLP. Areas of collaboration include research, creating educational tools and resources, international harmonization, and consensus standards, which may help inform regulatory policies and regulatory guidelines.

We envision these guiding principles may be used to:

- Adopt good practices that have been proven in other sectors
- Tailor practices from other sectors so they are applicable to medical technology and the health care sector
- Create new practices specific for medical technology and the health care sector

As the AI/ML medical device field evolves, so too must GMLP best practice and consensus standards. Strong partnerships with our international public health partners will be crucial if we are to empower stakeholders to advance responsible innovations in this area. Thus, we expect this initial collaborative work can inform our broader international engagements, including with the IMDRF.

We welcome your continued feedback through the public docket (FDA-2019-N-1185) at Regulations.gov, and we look forward to engaging with you on these efforts. The Digital Health Center of Excellence is spearheading this work for the FDA. Contact us directly at Digitalhealth@fda.hhs.gov, software@mhra.gov.uk, and mddpolicy-politiquesdim@hc-sc.gc.ca.

U.S. FOOD & DRUG ADMINISTRATION | Health Canada / Santé Canada | Medicines & Healthcare products Regulatory Agency

Good Machine Learning Practice for Medical Device Development: Guiding Principles	
Multi-Disciplinary Expertise Is Leveraged Throughout the Total Product Life Cycle	Good Software Engineering and Security Practices Are Implemented
Clinical Study Participants and Data Sets Are Representative of the Intended Patient Population	Training Data Sets Are Independent of Test Sets
Selected Reference Datasets Are Based Upon Best Available Methods	Model Design Is Tailored to the Available Data and Reflects the Intended Use of the Device
Focus Is Placed on the Performance of the Human-AI Team	Testing Demonstrates Device Performance During Clinically Relevant Conditions
Users Are Provided Clear, Essential Information	Deployed Models Are Monitored for Performance and Re-training Risks are Managed

https://www.fda.gov/medical-devices/software-medical-device-samd/good-machine-learning-practice-medical-device-development-guiding-principles

Good Machine Learning Practice (GMLP) Guiding Principles

Principle 7

Focus Is Placed on the Performance of the Human-AI Team

- Where the model has a "human in the loop," human factors considerations and the human interpretability of the model outputs are addressed with emphasis on the performance of the Human-AI team, rather than just the performance of the model in isolation.

Good Machine Learning Practice (GMLP) Guiding Principles

FDA

Principle 9

Users Are Provided Clear, Essential Information:

- Users are provided ready access to clear, contextually relevant information that is appropriate for the intended audience (such as health care providers or patients) including:
  - the product's intended use and indications for use,
  - performance of the model for appropriate subgroups,
  - characteristics of the data used to train and test the model,
  - acceptable inputs,
  - known limitations,
  - user interface interpretation,
  - and clinical workflow integration of the model.
- Users are also made aware of device modifications and updates from real-world performance monitoring, the basis for decision-making when available, and a means to communicate product concerns to the developer.

Patient-Centered Approach Incorporating Transparency to Users

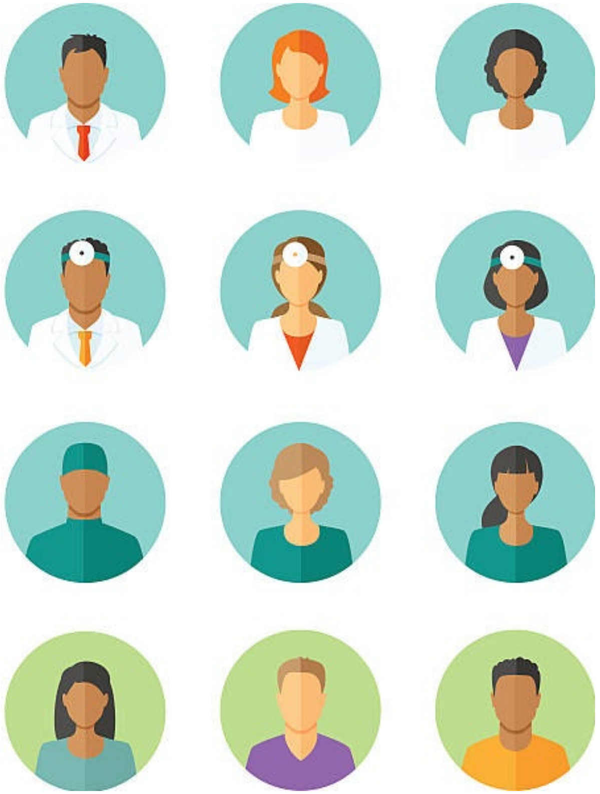
FDA

AI/ML-enabled devices have unique considerations that necessitate a proactive patient-centered approach:

- that takes into account issues including usability, equity, trust, and accountability
- that promotes transparency to all users and to patients more broadly

Patient Engagement Advisory Committee (PEAC) Meeting held Oct 2020

Workshop on Transparency of AIML-enabled devices held Oct 2021



Working Definition

TRANSPARENCY:

Degree to which appropriate information about a device

– including its intended use, development,

performance, and, when available, logic –

is clearly communicated to stakeholders

**Working definition of Transparency adapted from ISO/IEC 25059:2023* Software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Quality model for AI systems

Transparency is fundamental to a patient-centered approach

FDA

Transparency supports the safe and effective use of AI/ML-enabled devices

1. Allows patients, providers, and caregivers to make informed decisions
2. Supports proper use of a device
3. Promotes health equity
4. Facilitates evaluation and monitoring of device performance
5. Fosters trust and promotes adoption

Continuing to Improve Transparency

FDA

- What are the needs of specific stakeholders?
- What is the appropriate information to communicate?
- What is the best way to communicate that information?
  - How can device labeling be improved?
  - How can other public-facing information be improved?
  - What else can be done to promote transparency?

We carefully consider the discussions held in our public workshop on the Transparency of AI/ML-enabled Medical Devices, as well as comments from the public docket to inform our next steps toward improving transparency.

Transparency Workshop

- **3800 workshop participants**
  - Patients, healthcare professionals, academia, advocacy groups, and industry
- **Discussion Themes (Workshop and Docket)**

- Health equity and bias
- Labeling
- Public education efforts
- Decision summaries
- Databases
- Post market pathways
- Real world performance monitoring
- Industry guidance
- Data set requirements
- Validation of transparency measures
- Promoting GMLP

Topics of AI/ML Transparency Workshop Discussion

What does AI/ML Transparency mean?

- Safe and effective
  - Clear intended use
  - Works as described
- Health equity
  - Fair to all people
  - Bias management
- Real world performance
  - Assurance of improved health outcomes

How to promote AI/ML Transparency?

- User facing information/ labeling
  - Accessible language/terminology
  - Clear functionality and limitations
- Public education on AI/ML
- Dataset requirements
- Pre-market guidance
- Post-market monitoring

Planned Guidance

FDA

A-List: Prioritized Guidance Documents that CDRH Intends to Publish in FY2024

Final Guidance Topics

- Marketing Submission Recommendations for A Predetermined Change Control Plan for Artificial Intelligence/Machine Learning (AI/ML)-Enabled Device Software Functions

Draft Guidance Topics

- Artificial Intelligence/Machine Learning (AI/ML)-enabled Device Software Functions: Lifecycle Management Considerations and Premarket Submission Recommendations

https://www.fda.gov/medical-devices/guidance-documents-medical-devices-and-radiation-emitting-products/cdrh-proposed-guidances-fiscal-year-2024fy2024#a

Continuing our Collaborative Approach

2019	2020	2021	2022	2023
• Published AI/ML-SaMD Discussion Paper • First joined Collaborative Community related to AI/ML	• Public Workshop on AI/ML in Radiological Imaging • Patient Engagement Advisory Committee Meeting on Patient Trust in AI/ML Devices	• Published AI/ML Medical Device Software Action Plan • Posted List of Currently Marketed AI/ML Devices • Public Workshop on Transparency of AI/ML Devices • Published Good Machine Learning Practice Principles	• Contributed to IMDRF's Key Terms & Definitions: Machine Learning Enabled Medical Devices • Published Clinical Decision Support (CDS) Final Guidance • Updated List of Currently Marketed AI/ML Devices • Recognized new Consensus Standard on AI/ML	• Published Predetermined Change Control Plan for AI/ML Devices Draft Guidance • Updated List of Currently Marketed AI/ML Devices • Published Pre-determined Change Control Plans for Machine Learning-Enabled Medical Devices Guiding Principles

We recognize that by working collaboratively with stakeholders we can lay out a clear path toward building a proactive patient-centered approach to the development and use of AI/ML-enabled devices.

Further Questions or Feedback:

www.fda.gov/digitalhealth

DigitalHealth@fda.hhs.gov

Shawn Forrest

CDRH Digital Health Center of Excellence

Office of Strategic Partnerships & Technology Innovation (OST)

Center for Devices and Radiological Health (CDRH), U.S. Food and Drug Administration

shawn.forrest@fda.hhs.gov

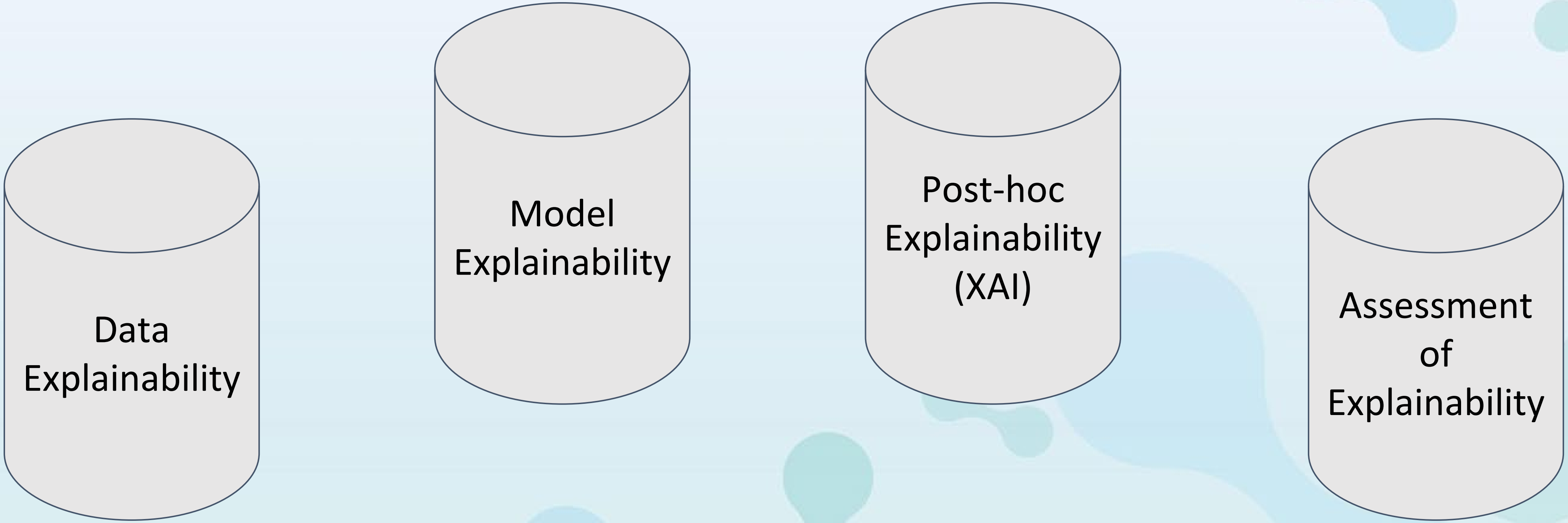
Utilizing AI for health and patient safety

Data science teams should strive to build and develop AI models that positively impact patient lives by providing information and decisions that are trustworthy and transparent.

If there is a lack of transparency, data teams will be less likely to interpret the results, and stakeholders may begin to question the usefulness of models without justification.

Being well versed in model transparency and explainable AI (XAI) will help data teams interpret the results while helping stakeholders build confidence in model outputs.

Model transparency is built on the four pillars of explainable AI (XAI)



How well do you know your data? (Data Explainability)

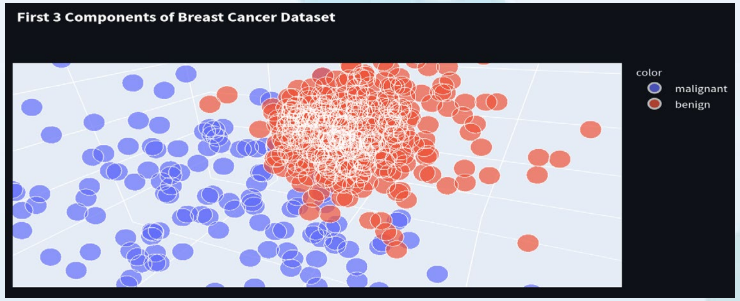
Statistical Analysis:

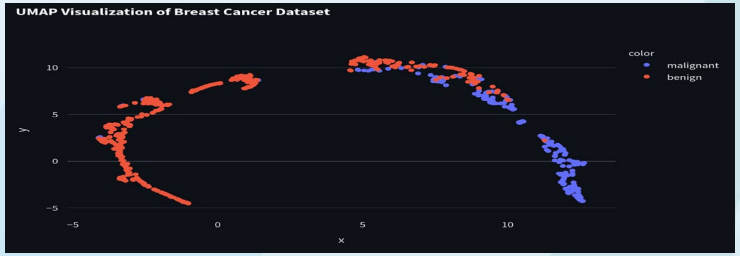
- Exploratory Data Analysis (EDA)
- Feature Engineering
- Physics Informed Neural Networks

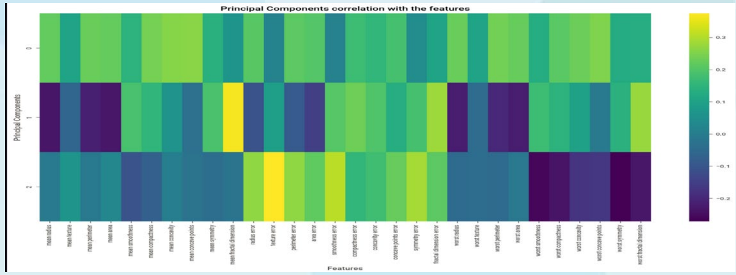
- Data Dimensionality Reduction / Visualization:
  - Principal Component Analysis (PCA)
  - t-distributed Stochastic Neighbor Embedding (TSNE)
  - Uniform Manifold Approximation and Projection (UMAP)

Causal / Relationship Mapping:

- Knowledge Graphs







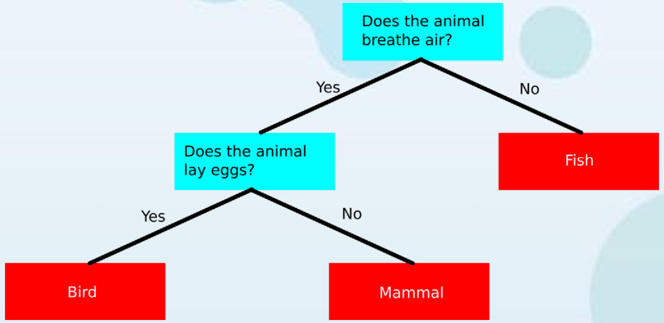
How well do you know your model(s)? (Model Explainability)

White Box Models - Usually basic and don't capture complexity

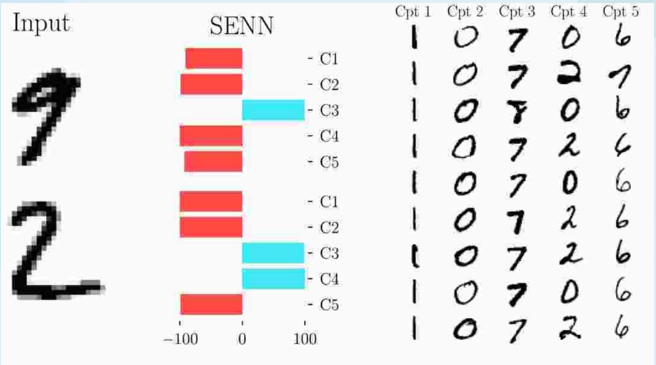
- Decision sets
- Rule Sets
- Cased-based reasoning
- Interpretable Fuzzy Systems
- Generalized Additive Models

Hybrid Explainable Models – Models coupled with black box models

- Deep K-Nearest Neighbors (DkNN)
- Deep Weighted Averaging Classifier (DWAC)
- Self-Explaining Neural Networks (SENN)
- BagNets
- Neural-Symbolic models (NeSy)



https://towardsdatascience.com/a-beginners-guide-to-decision-tree-classification-6d3209353ea



https://omarelb.github.io/self-explaining-neural-networks

How did your model come to a conclusion?

Attribution Methods

Attribution Methods are great for understanding image explanations by looking at relevant pixels

Four Methods:

- Deep Taylor Decomposition (DTD)
- Perturbation Methods
- Backpropagation Methods
- DeepLIFT

Method:
Grad-CAM

Predicted:
Ox

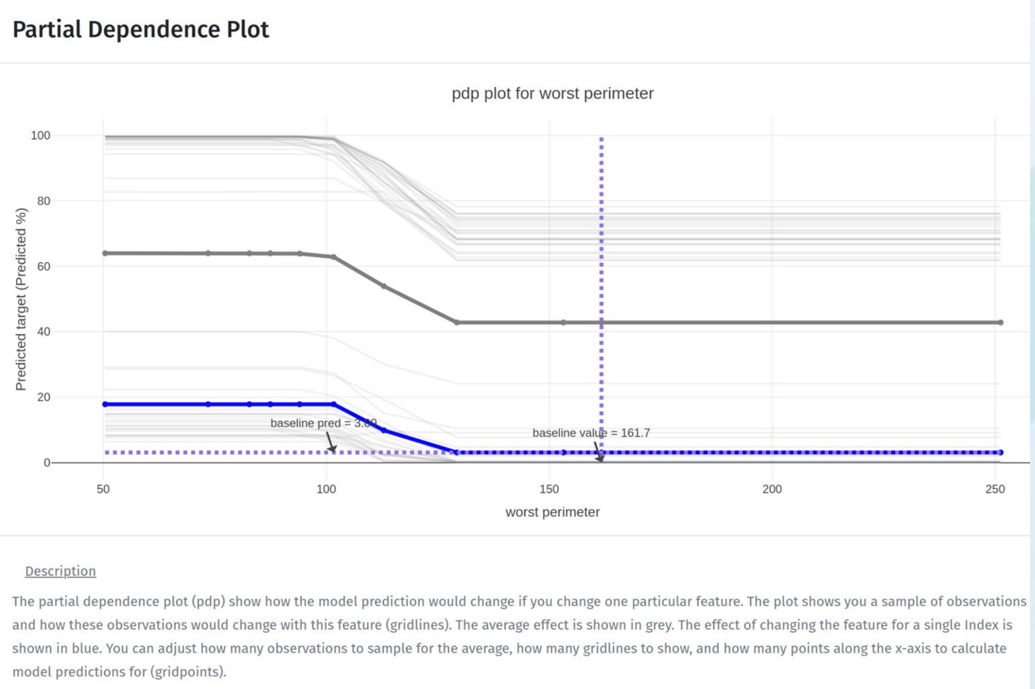
How did your model come to a conclusion?

Visualization Methods

Visualizations work well with supervised learning methods by understanding feature contributions

Three Methods:

- Partial Dependence Plot (PDP)
- Individual Conditional Expectations (ICE)
- Accumulated Local Effects (ALE)



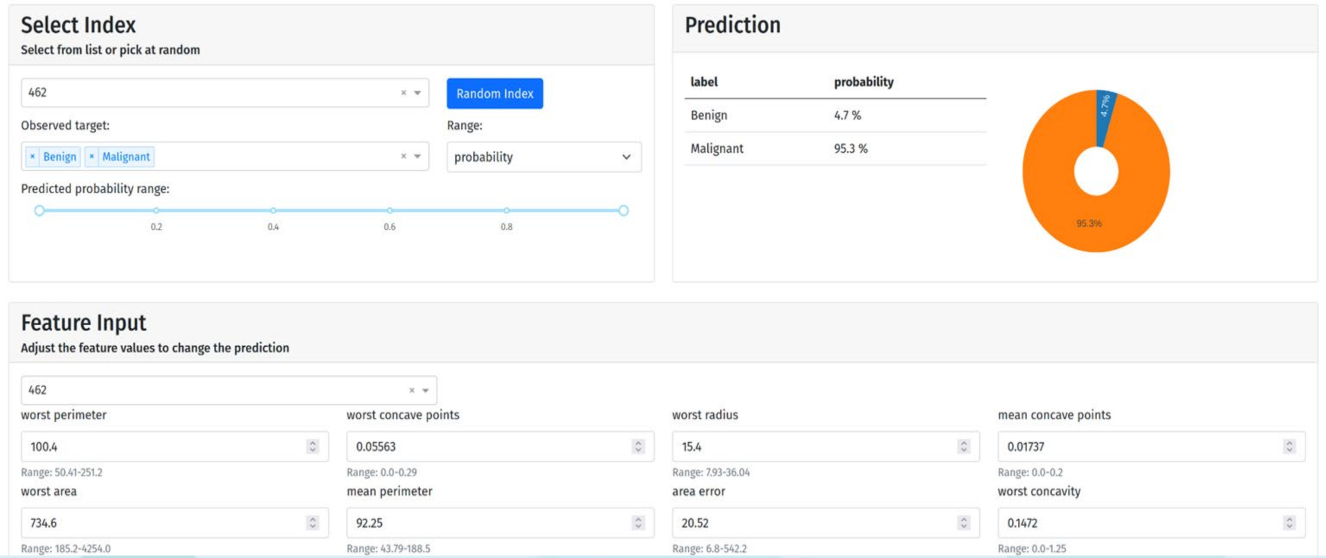
How did your model come to a conclusion?

Example-based Explanation Methods

Example-based Explanation Methods focus on single instances in the dataset to explain the underlying data distributions.

Three Methods:

- Prototype and Criticisms
- Counterfactuals
- Adversarial Examples



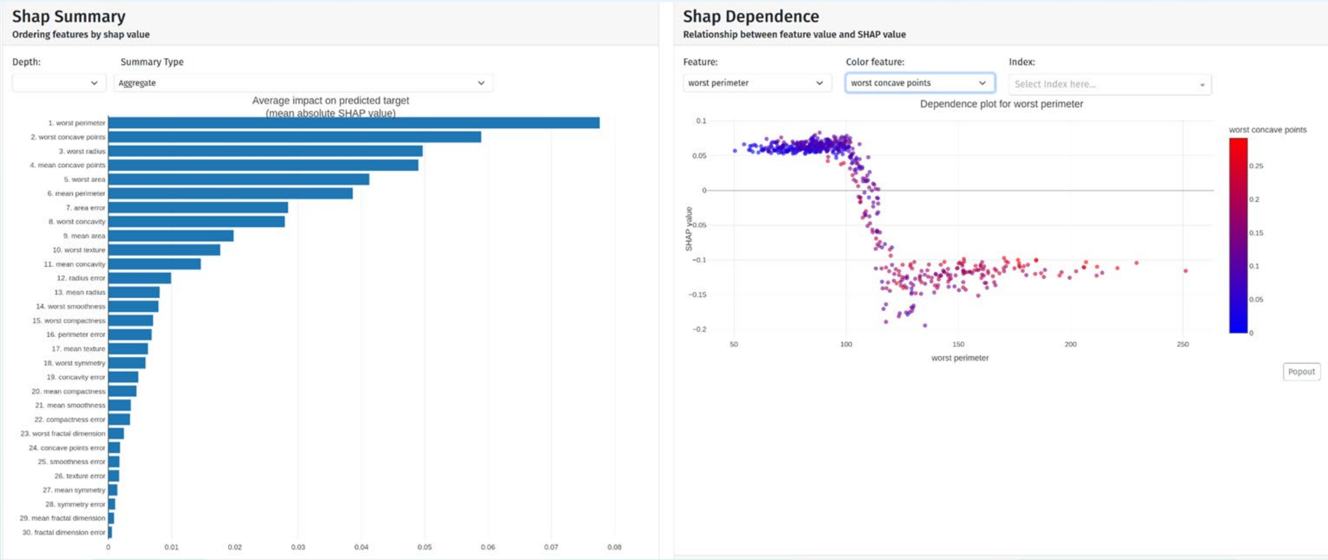
How did your model come to a conclusion?

Game Theory Methods

Game Theory Methods treat machine learning models like games where the features show how much they contribute to the prediction outcome

Two Methods:

- Shap Values
- Shapley Additive Explanation (SHAP)



How did your model come to a conclusion?

Knowledge Extraction Methods and Neural Methods

Knowledge Extraction Methods and Neural Methods look to see what features are being used by a neural network features:

Four Methods:

- Rule Extraction
- Model Distillation
- Influence Methods
- Concept Methods

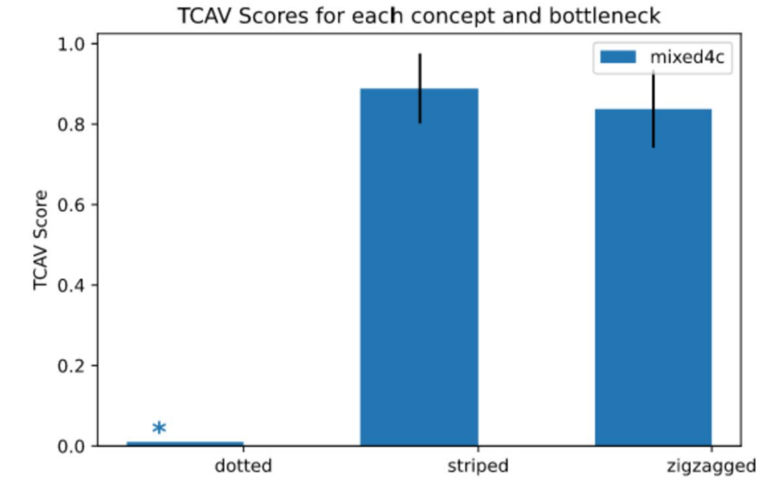


FIGURE 10.11: The example of measuring TCAV scores of three concepts for the model predicting "zebra". The targeted bottleneck is a layer called "mixed4c". A star sign above "dotted" indicates that "dotted" has not passed the statistical significance test, i.e. having the p-value larger than 0.05. Both "striped" and "zigzagged" have passed the test, and both concepts are useful for the model to identify "zebra" images according to TCAV. Figure originally from the TCAV GitHub.

https://christophm.github.io/interpretable-ml-book/detecting-concepts.html

How well do you understand the results? (Assessment of Explanations)

- Understandability and Satisfaction
  - Qualitative approaches
    - Explain the answer provided
  - Quantitative approaches
    - Rate the explanation of the answer / result (Likert Scales)
- Trust and Transparency
  - Swift Trust – The user immediately accepts the model's output
  - Default Trust – The user depends on the model's output
  - Suspicious Trust – The user has apprehension to the model's output

XAI in action!

Auto-EDA

XAI Dashboard

Visualizations

GRAD-CAM