From Theory To Practice:

Implementing Machine Learning Solutions Safely and Effectively in the Clinical Laboratory

Nick Spies, MD

Co-Medical Director - Applied AI & Clinical Chemistry, ARUP Laboratories

Assistant Professor - University of Utah

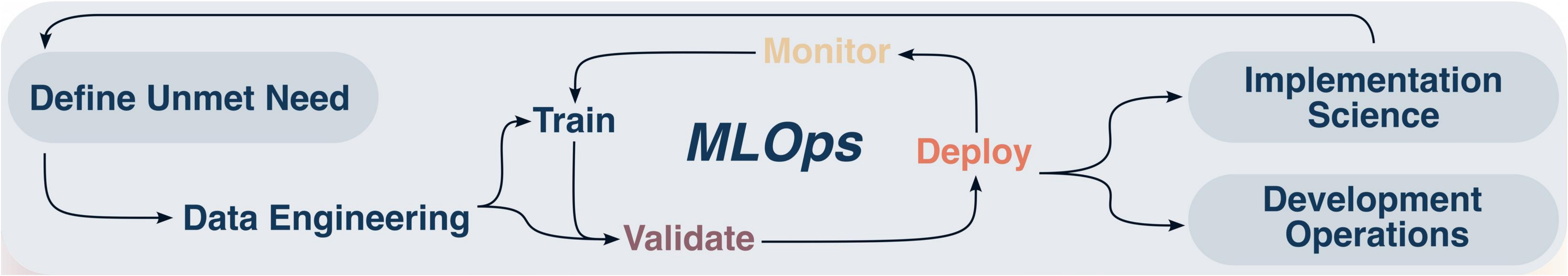
Disclosures

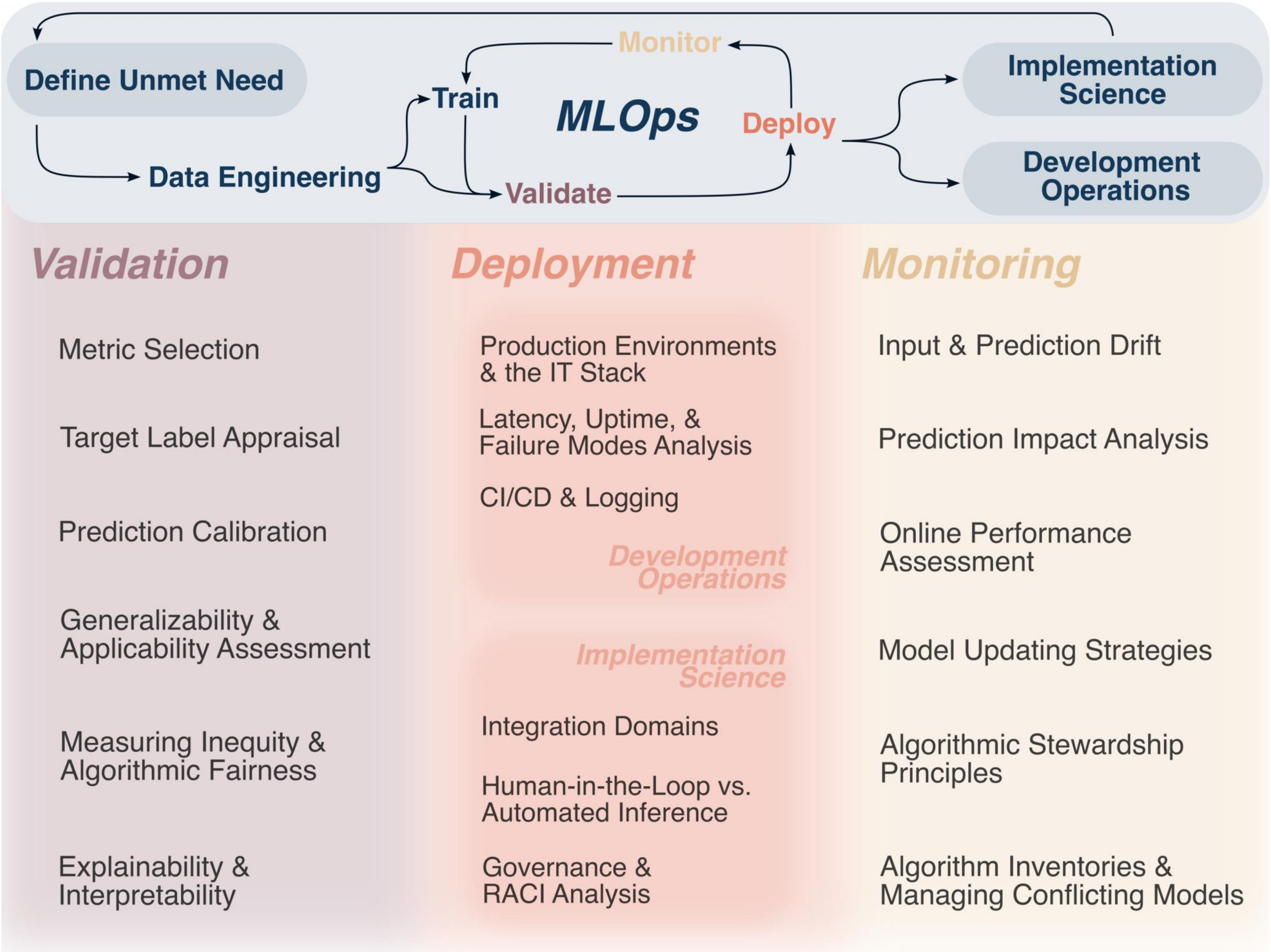
- I have no relevant conflicts of interest to disclose.

Learning Objectives

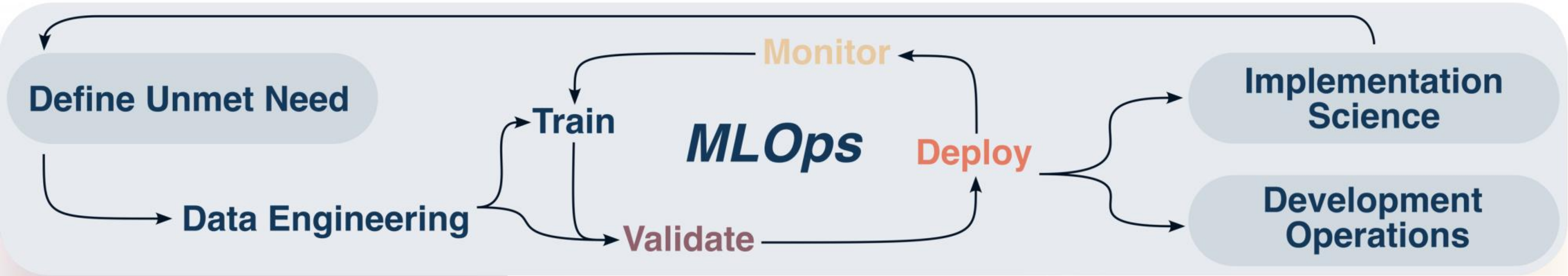
- Define key roles and responsibilities in the machine learning life-cycle.
- Explore techniques for validating, deploying, and monitoring models.
- Reinforce these concepts within a relevant, lab-based example.

The Machine Learning Life Cycle





Define Unmet Need
Data Engineering
Train
Monitor
MLOps
Deploy
Validate
Implementation Science
Development Operations
Validation
Metric Selection
Target Label Appraisal
Prediction Calibration
Generalizability & Applicability Assessment
Measuring Inequity & Algorithmic Fairness
Explainability & Interpretability
Deployment
Production Environments & the IT Stack
Latency, Uptime, & Failure Modes Analysis
CI/CD & Logging
Development Operations
Implementation Science
Integration Domains
Human-in-the-Loop vs. Automated Inference
Governance & RACI Analysis
Monitoring
Input & Prediction Drift
Prediction Impact Analysis
Online Performance Assessment
Model Updating Strategies
Algorithmic Stewardship Principles
Algorithm Inventories & Managing Conflicting Models



Validation

Metric Selection

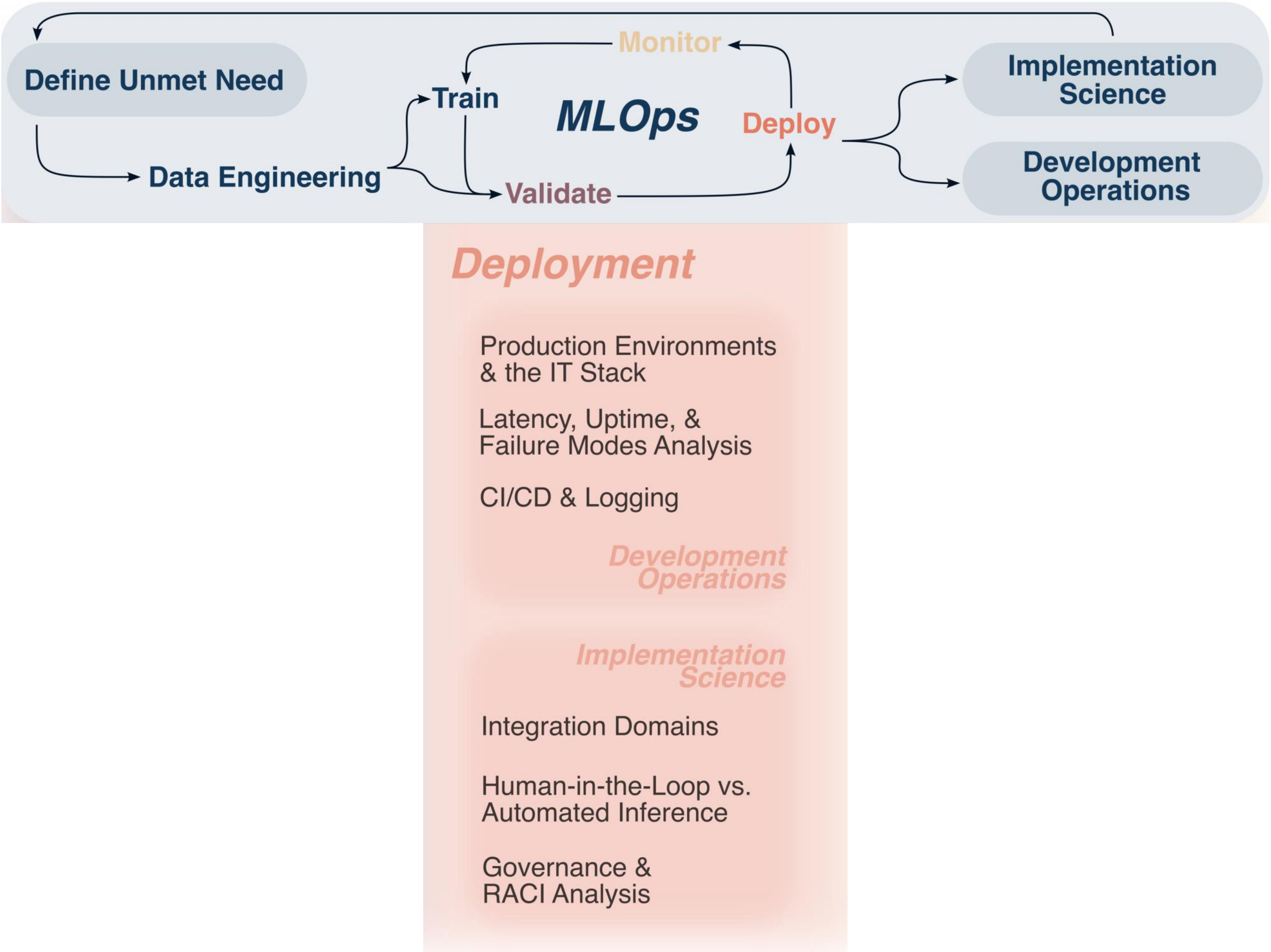
Target Label Appraisal

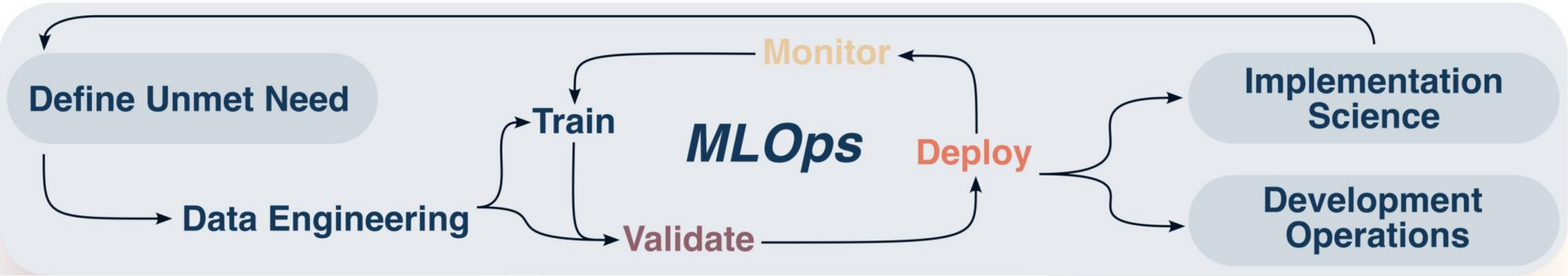
Prediction Calibration

Generalizability & Applicability Assessment

Measuring Inequity & Algorithmic Fairness

Explainability & Interpretability


Define Unmet Need
Data Engineering
Train
Validate
Monitor
Deploy
MLOps
Implementation Science
Development Operations
Deployment
Production Environments & the IT Stack
Latency, Uptime, & Failure Modes Analysis
CI/CD & Logging
Development Operations
Implementation Science
Integration Domains
Human-in-the-Loop vs. Automated Inference
Governance & RACI Analysis



Monitoring

Input & Prediction Drift

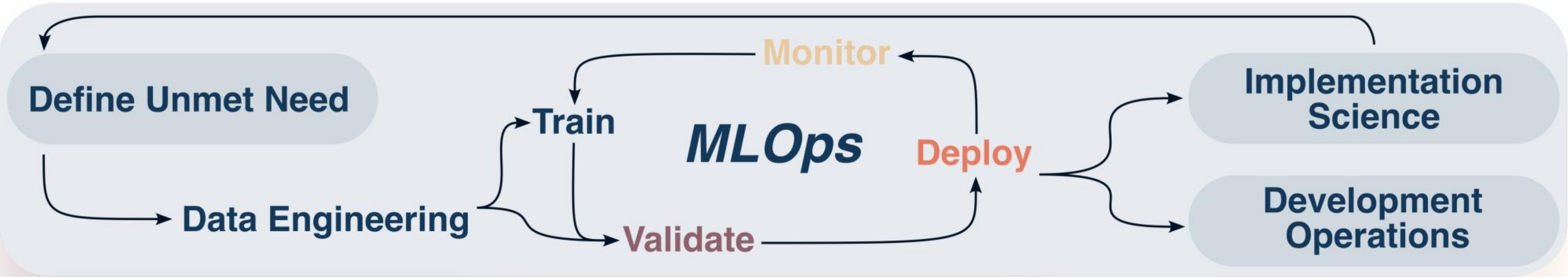
Prediction Impact Analysis

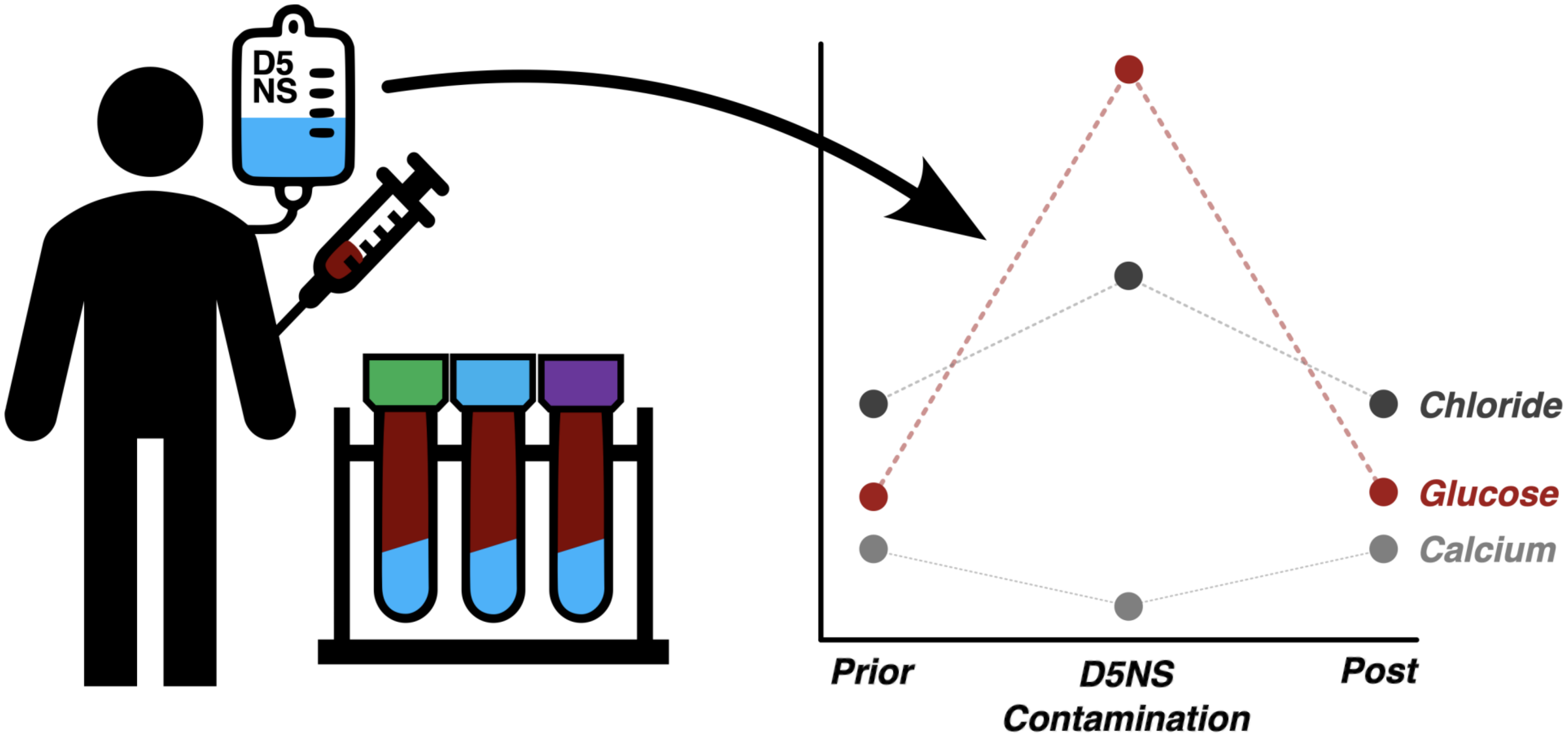
Online Performance Assessment

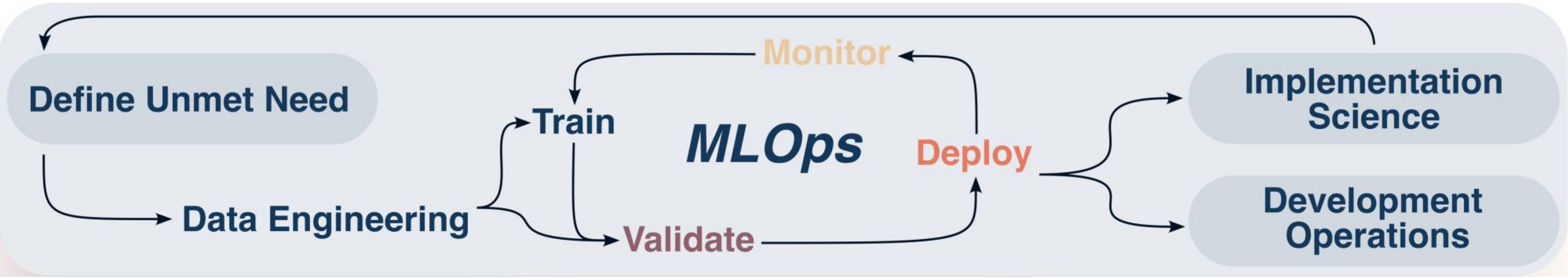
Model Updating Strategies

Algorithmic Stewardship Principles

Algorithm Inventories & Managing Conflicting Models


Define Unmet Need
Data Engineering
Train
Monitor
MLOps
Deploy
Validate
Implementation Science
Development Operations


D5 NS
Chloride
Glucose
Calcium
Prior
D5NS Contamination
Post



JOURNAL ARTICLE

Validating, Implementing, and Monitoring Machine Learning Solutions in the Clinical Laboratory Safely and Effectively

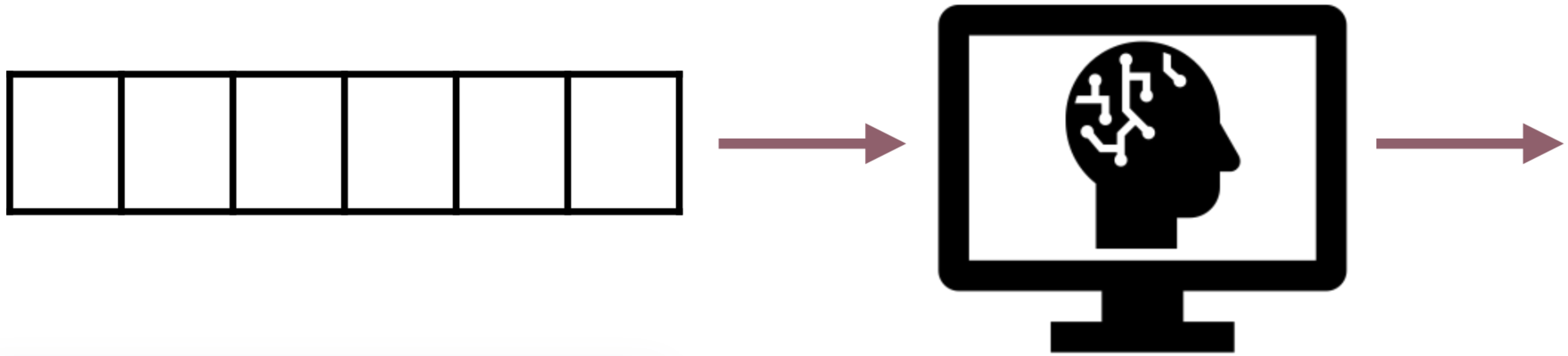
Nicholas C Spies ✉, Christopher W Farnsworth, Sarah Wheeler, Christopher R McCudden

Clinical Chemistry, hvae126, https://doi.org/10.1093/clinchem/hvae126

Published: 10 September 2024 **Article history** ▾

Defining A Machine Learning Pipeline

Inputs **Model** **Output**

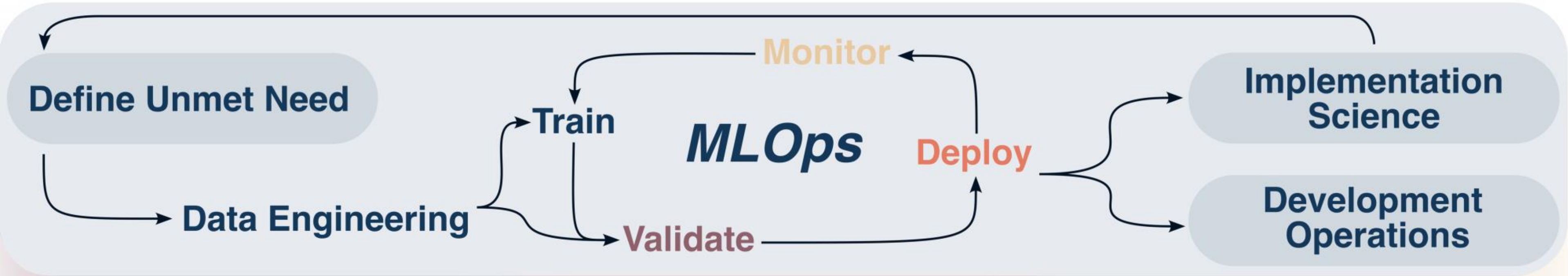


Output: *0.97*

Label: ***Positive***

Applicable: ✓

Explanation: ✓



Building The Model

```r
# Load required libraries
library(tidymodels)
library(arrow)

# Load the data
train <-
  read_feather("https://figshare.com/ndownloader/files/45407401") |>
  select(contam_comment, bun:sodium) |>
  mutate(contam_comment = factor(contam_comment))

# Define the feature recipe
recipe <- recipe(contam_comment ~ ., data = train)

# Define the model
model <- boost_tree(mode = "classification") |> set_engine("xgboost")

# Create the workflow
workflow <- workflow() |> add_recipe(recipe) |> add_model(model)

# Fit the model
fit <- workflow |> fit(data = train)
```

Testing The Model

Sodium	147 !
Potassium	3.5
Chloride	119 !
CO2	17 !
Creatinine	0.9
BUN	22
Calcium	6.6 !
Glucose	86

Model

Output

Output: *0.97*

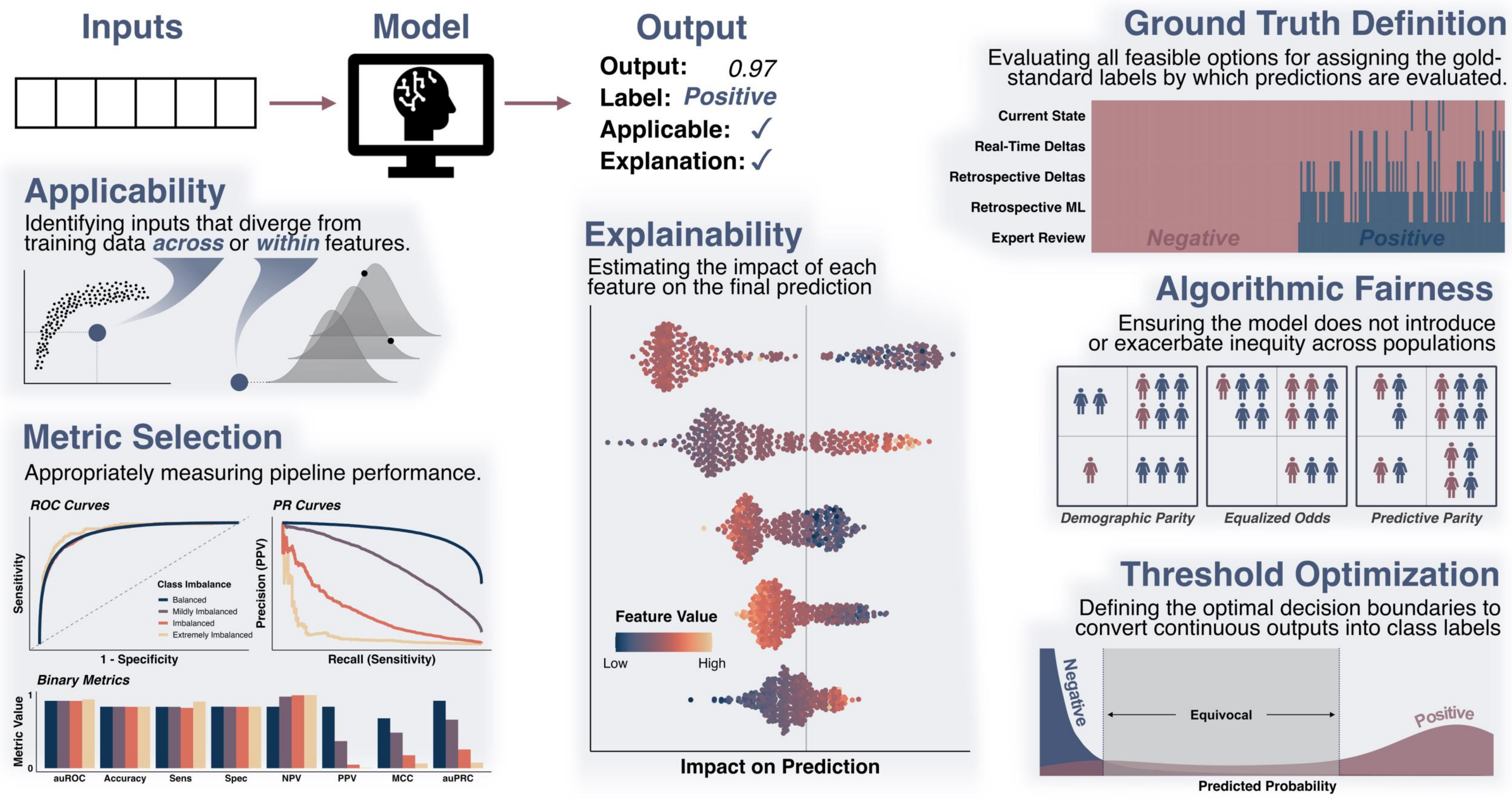
Label: ***Positive***

Applicable: ✓

Explanation: ✓

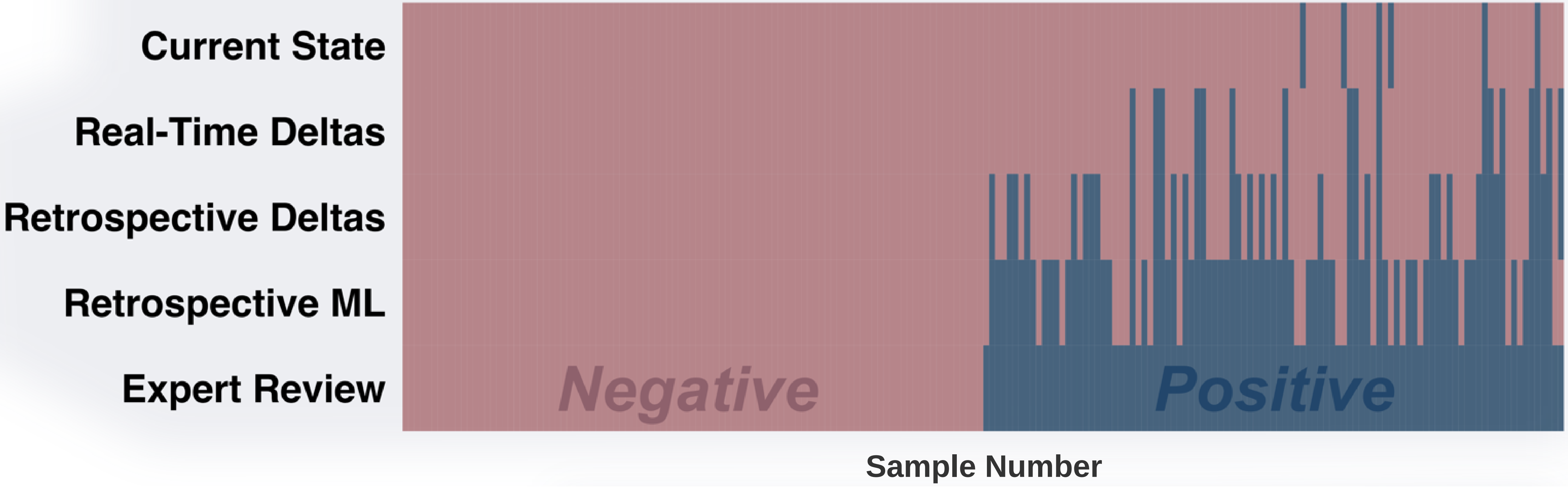
How can we *validate* a developed model?

Successful *Validation* of Machine Learning Pipelines

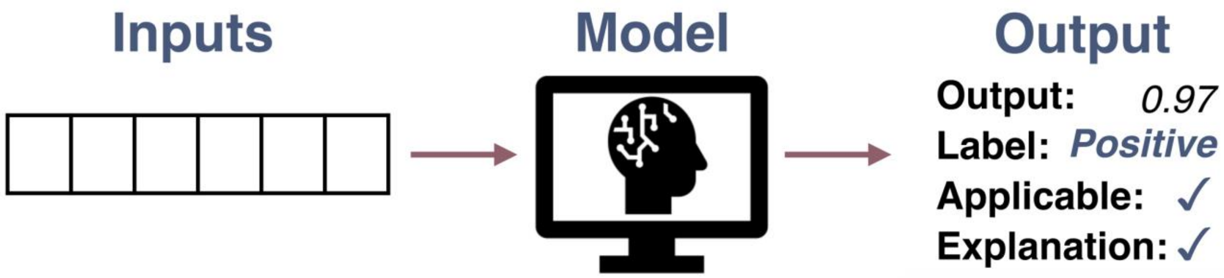


Ground Truth Definition

Evaluating all feasible options for assigning the gold-standard labels by which predictions are evaluated.



Successful *Validation* of Machine Learning Pipelines

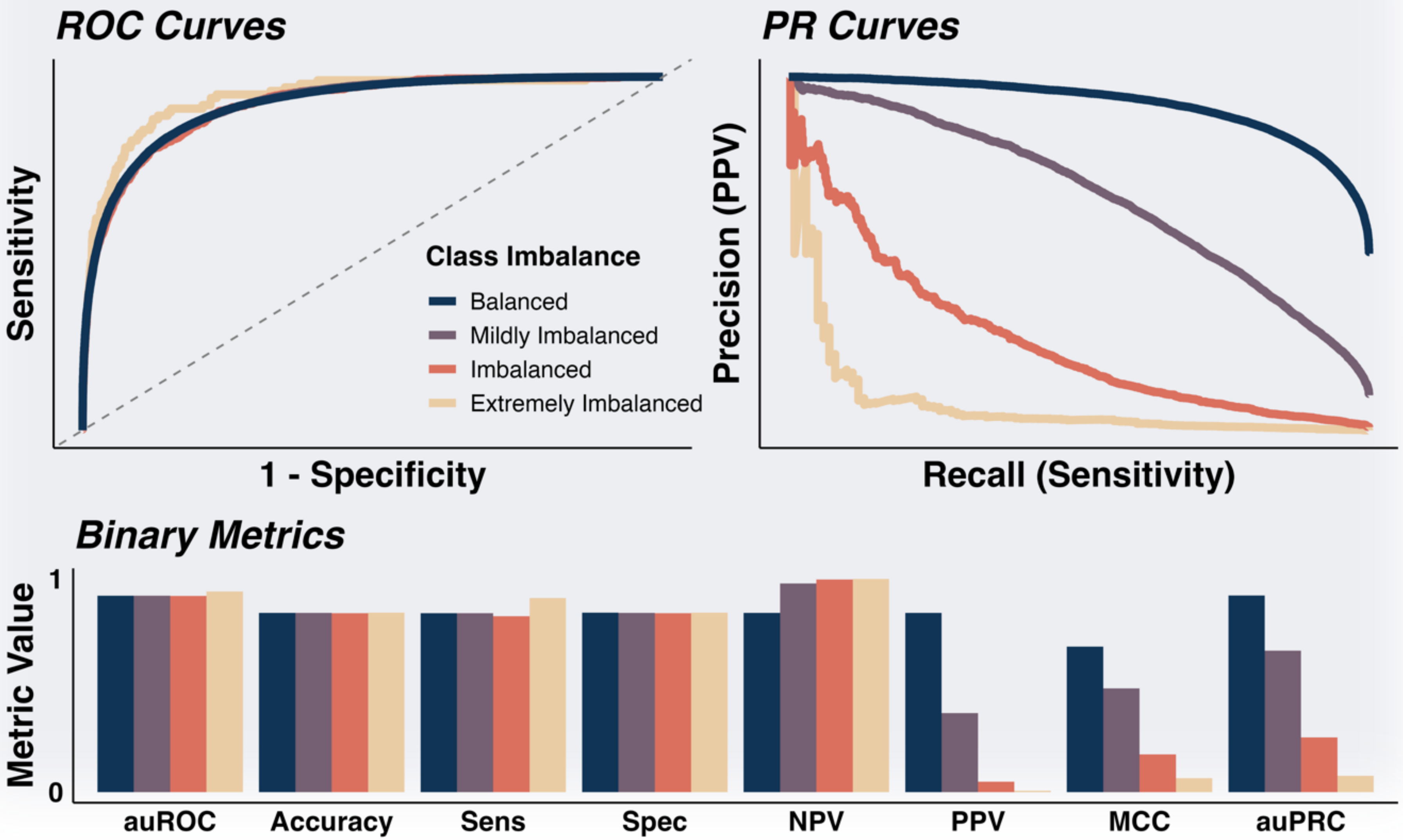


Ground Truth Definition

Evaluating all feasible options for assigning the gold-standard labels by which predictions are evaluated.

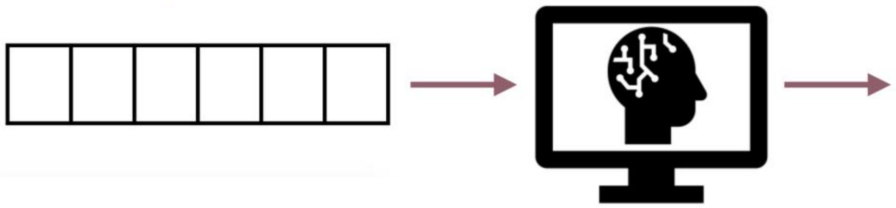
Metric Selection

Appropriately measuring pipeline performance.



Successful *Validation* of Machine Learning Pipelines

Inputs   Model   Output



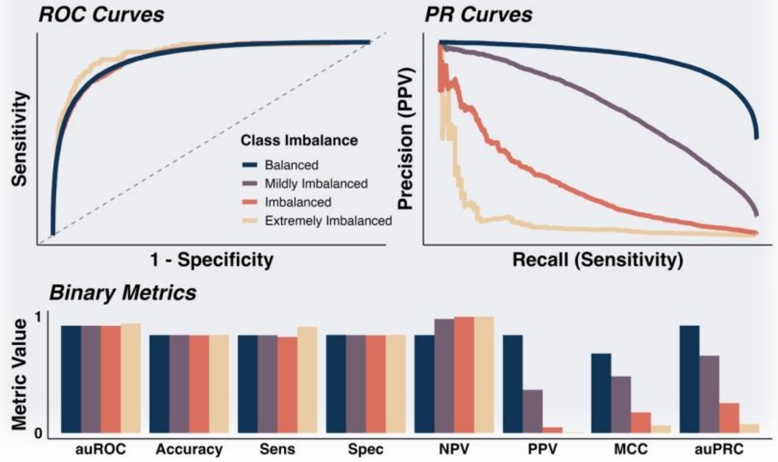
Output: *0.97*
Label: ***Positive***
Applicable: ✓
Explanation: ✓

Ground Truth Definition

Evaluating all feasible options for assigning the gold-standard labels by which predictions are evaluated.

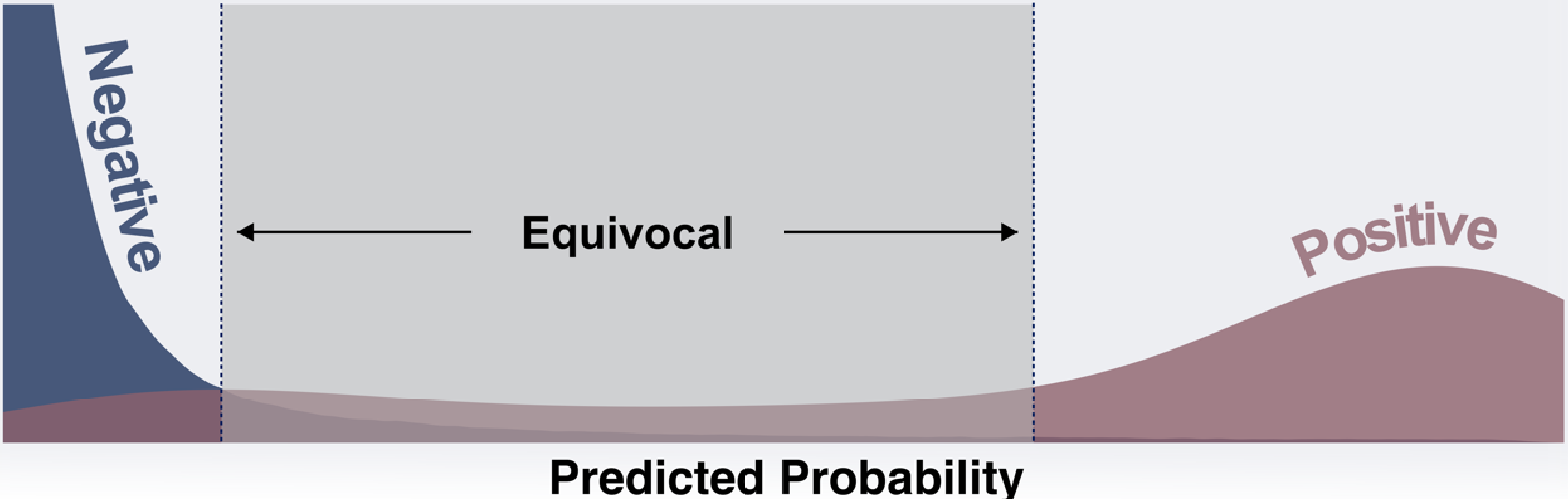
Metric Selection

Appropriately measuring pipeline performance.

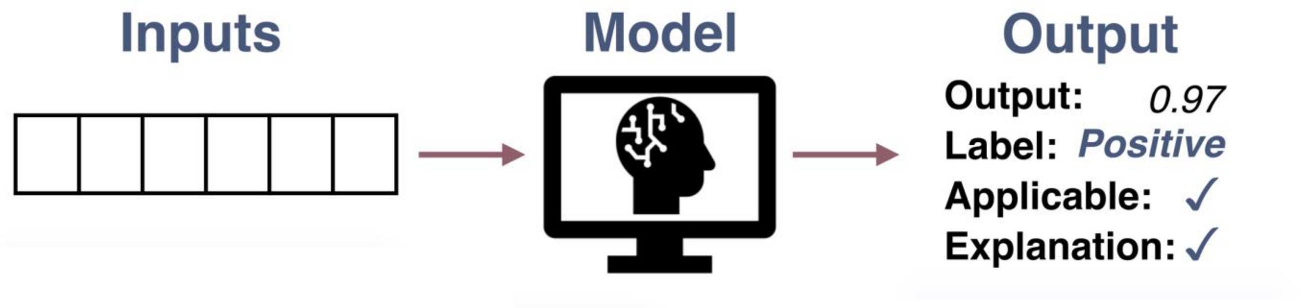


Threshold Optimization

Defining the optimal decision boundaries to convert continuous outputs into class labels



*Successful **Validation** of Machine Learning Pipelines*

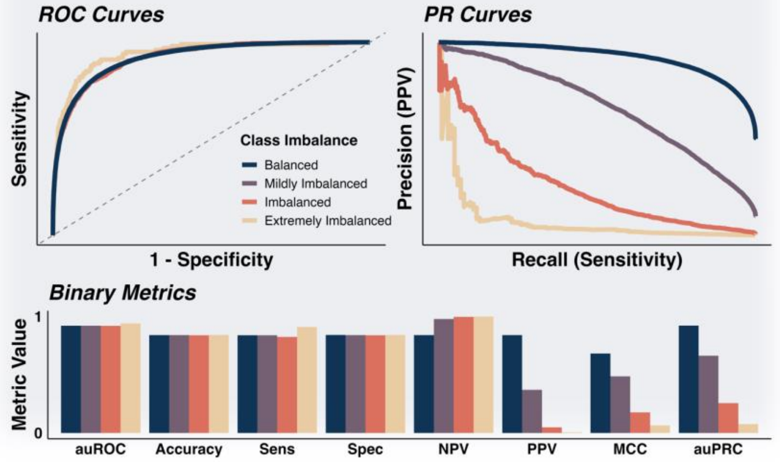


Ground Truth Definition

Evaluating all feasible options for assigning the gold-standard labels by which predictions are evaluated.

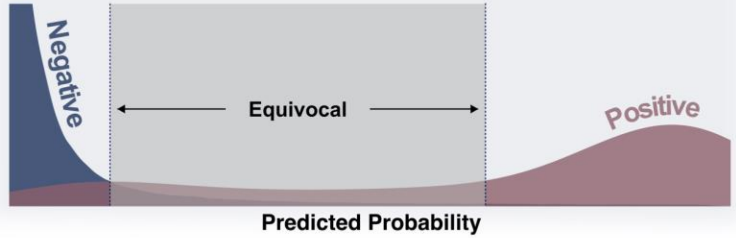
Metric Selection

Appropriately measuring pipeline performance.



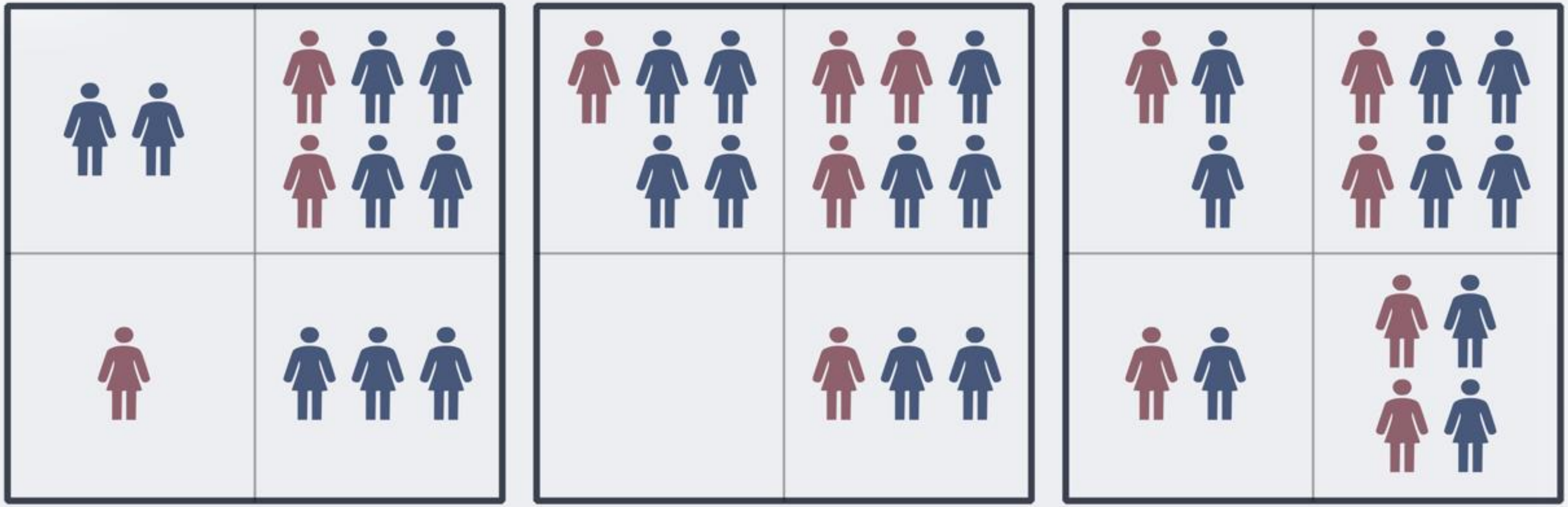
Threshold Optimization

Defining the optimal decision boundaries to convert continuous outputs into class labels



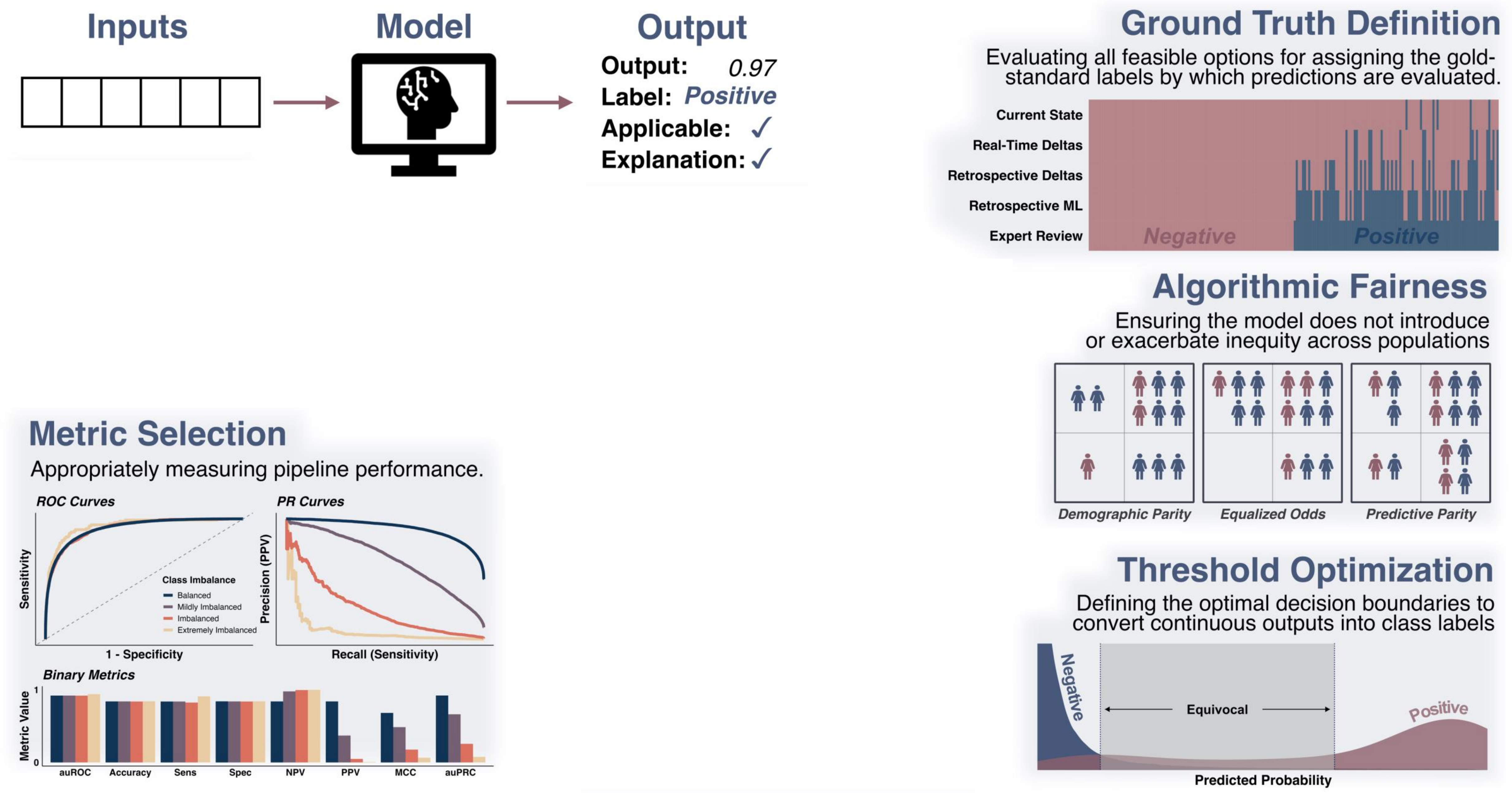
Algorithmic Fairness

Ensuring the model does not introduce or exacerbate inequity across populations



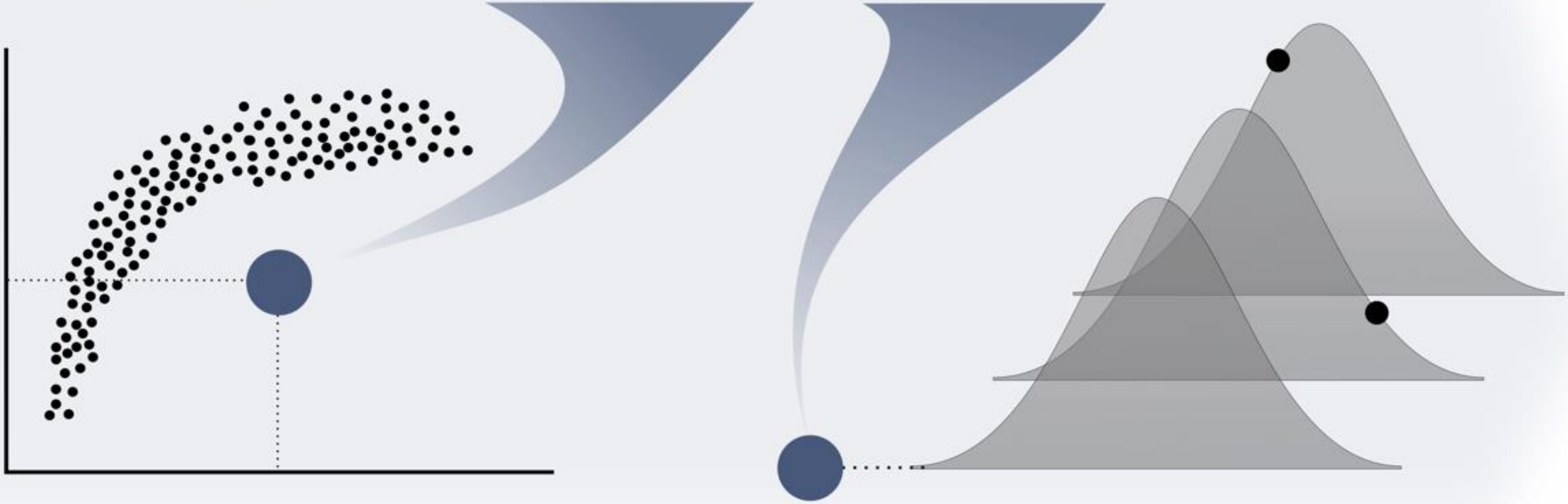
Demographic Parity *Equalized Odds* *Predictive Parity*

Successful Validation of Machine Learning Pipelines

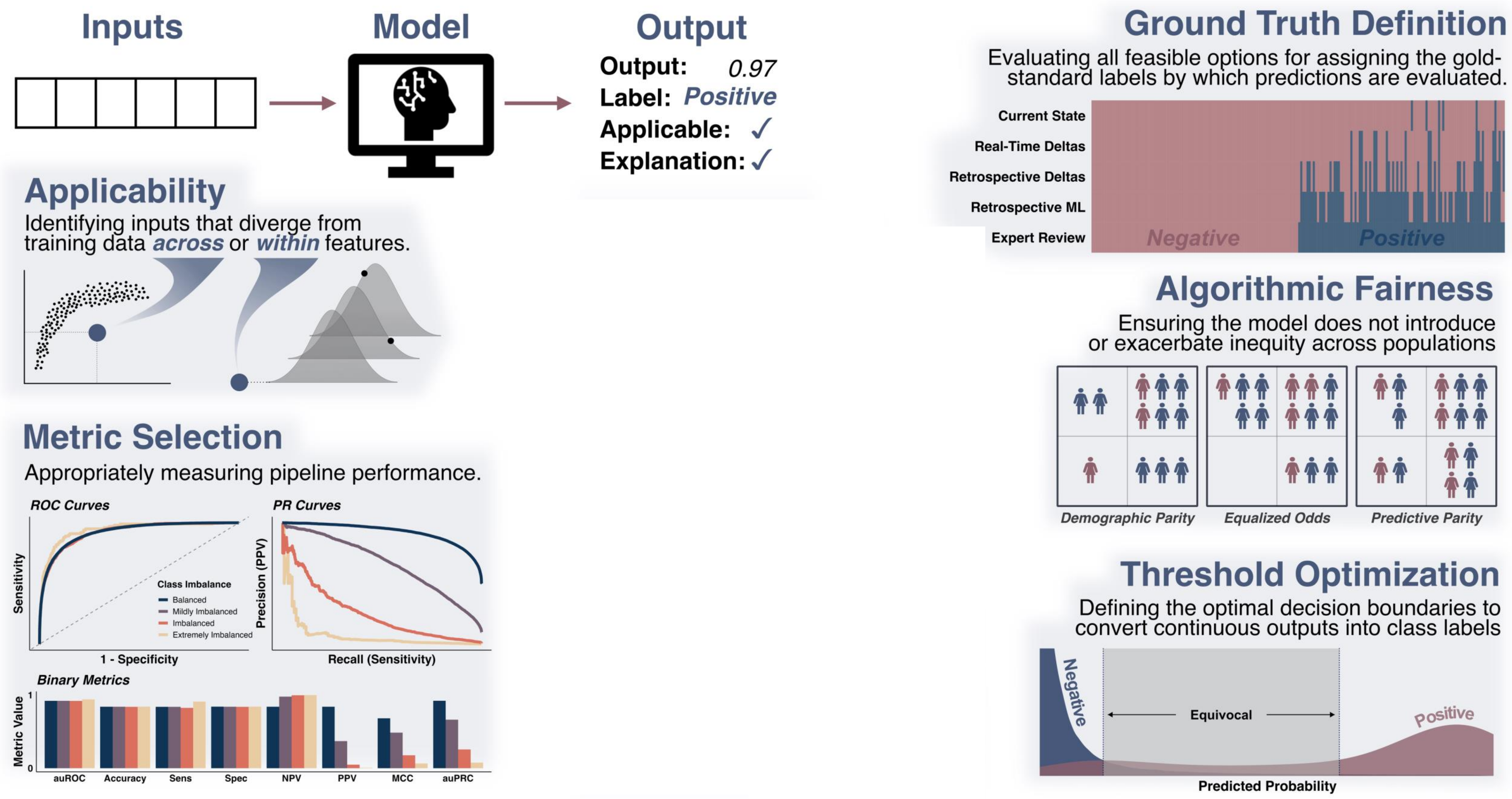


Applicability

Identifying inputs that diverge from training data ***across*** or ***within*** features.

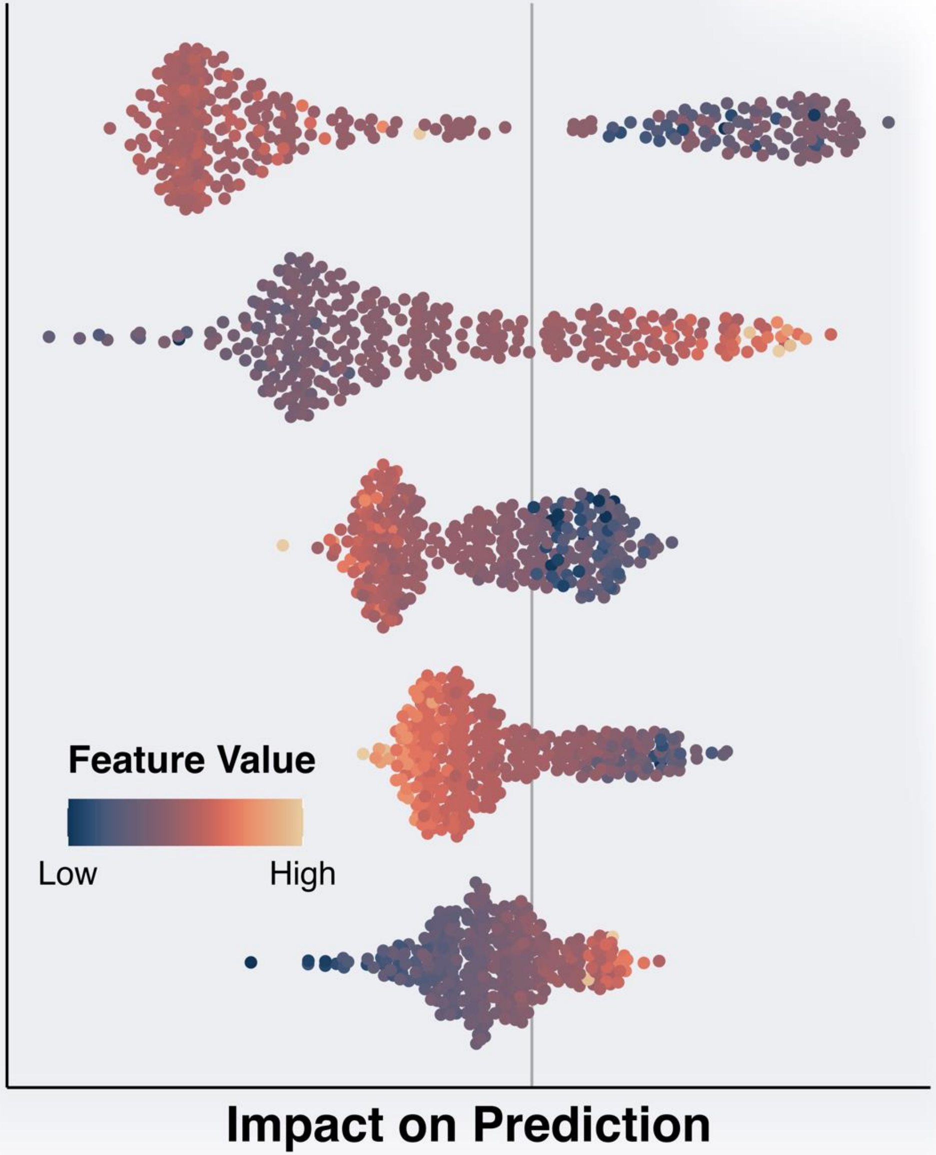


Successful *Validation* of Machine Learning Pipelines

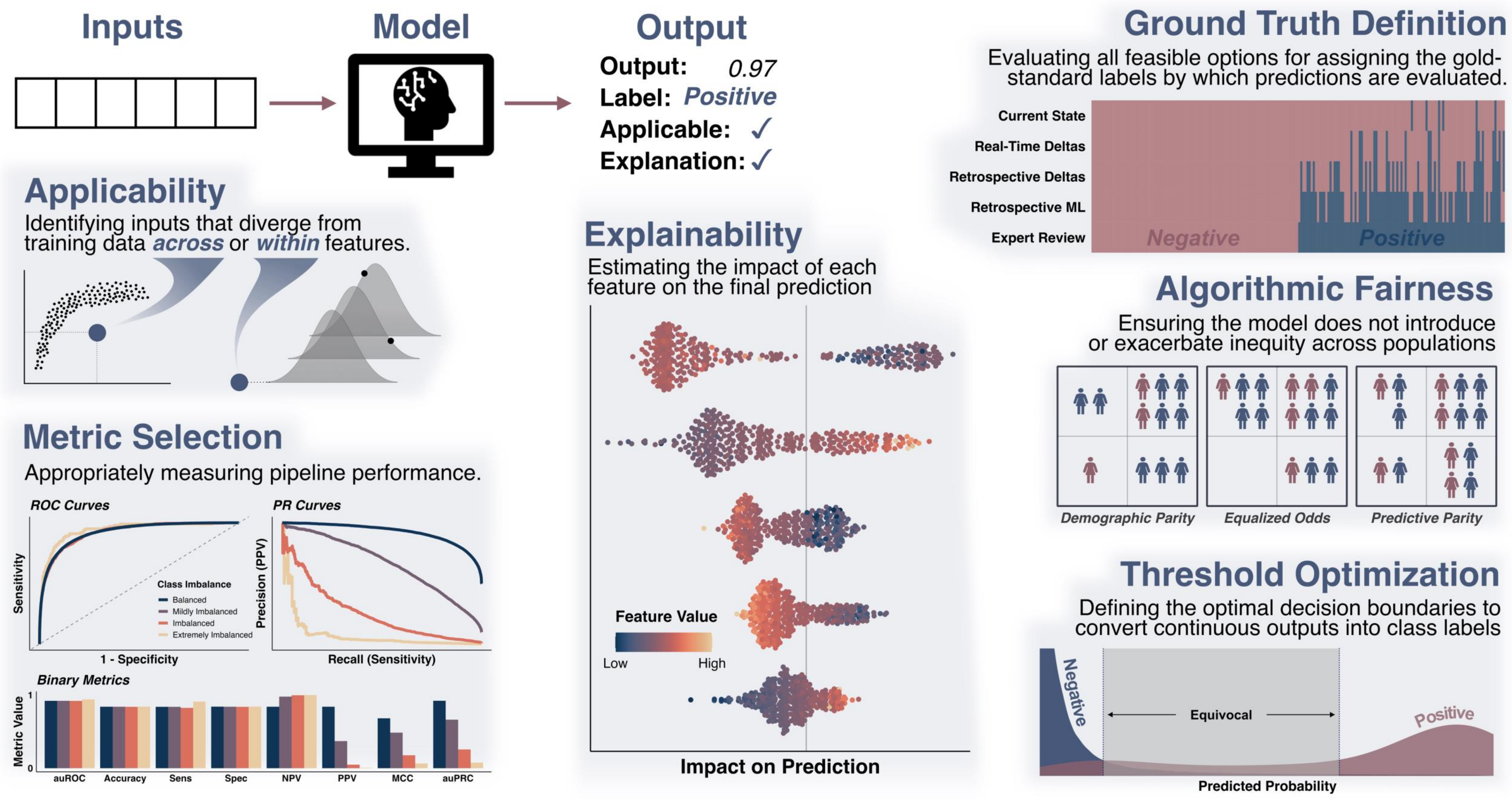


Explainability

Estimating the impact of each feature on the final prediction



*Successful **Validation** of Machine Learning Pipelines*



How do we ***implement*** a validated model?

*Successful **Implementation** of Machine Learning Pipelines*

Key Roles and Responsibilities

Subject Matter Experts

- Align implementation to fit unmet clinical need.
- Evaluate failure modes and off-target effects.

Successful *Implementation* of Machine Learning Pipelines

Key Roles and Responsibilities

Subject Matter Experts

- Align implementation to fit unmet clinical need.
- Evaluate failure modes and off-target effects.

Data Scientists & Data Engineers

- Build and evaluate models for deployment.
- Optimize storage and retrieval of input data.

Successful *Implementation* of Machine Learning Pipelines

Key Roles and Responsibilities

Subject Matter Experts

- Align implementation to fit unmet clinical need.
- Evaluate failure modes and off-target effects.

Data Scientists & Data Engineers

- Build and evaluate models for deployment.
- Optimize storage and retrieval of input data.

Software & ML Engineers

- Build robust and secure prediction pipelines.
- Implement best practices in DevOps/MLOps.

MLOps

The framework for **building**, **deploying**, and **monitoring** end-to-end ML solutions in live, **production environments** safely and effectively.

Successful *Implementation* of Machine Learning Pipelines

Key Roles and Responsibilities

Subject Matter Experts

- Align implementation to fit unmet clinical need.
- Evaluate failure modes and off-target effects.

Data Scientists & Data Engineers

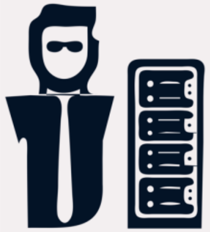
- Build and evaluate models for deployment.
- Optimize storage and retrieval of input data.

Software & ML Engineers

- Build robust and secure prediction pipelines.
- Implement best practices in DevOps/MLOps.

MLOps

The framework for **building**, **deploying**, and **monitoring** end-to-end ML solutions in live, **production environments** safely and effectively.



Information Technology & Systems

- Maintain interfaces for ML inputs and outputs.
- Develop infrastructure and allocate resources.

*Successful **Implementation** of Machine Learning Pipelines*

Key Roles and Responsibilities

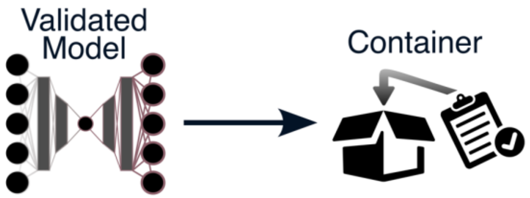
Subject Matter Experts

- Align implementation to fit unmet clinical need.
- Evaluate failure modes and off-target effects.

Data Scientists & Data Engineers

- Build and evaluate models for deployment.
- Optimize storage and retrieval of input data.

Terms and Technologies

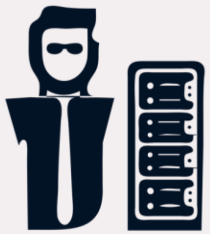


*Successful **Implementation** of Machine Learning Pipelines*

Key Roles and Responsibilities

Software & ML Engineers

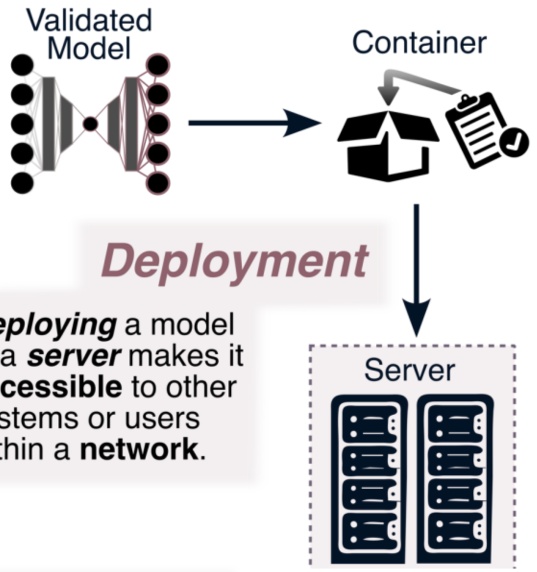
- Build robust and secure prediction pipelines.
- Implement best practices in DevOps/MLOps.



Information Technology & Systems

- Maintain interfaces for ML inputs and outputs.
- Develop infrastructure and allocate resources.

Terms and Technologies



Deployment

Deploying a model to a ***server*** makes it **accessible** to other systems or users within a **network**.

MLOps

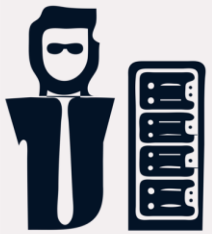
The framework for **building**, **deploying**, and **monitoring** end-to-end ML solutions in live, **production environments** safely and effectively.

Successful *Implementation* of Machine Learning Pipelines

Key Roles and Responsibilities

Software & ML Engineers

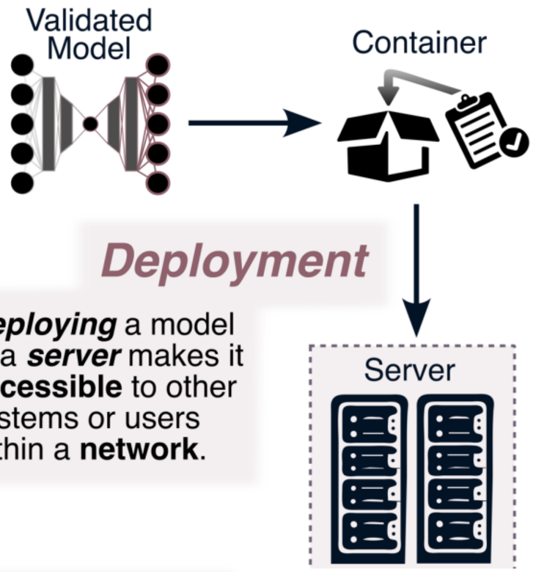
- Build robust and secure prediction pipelines.
- Implement best practices in DevOps/MLOps.



Information Technology & Systems

- Maintain interfaces for ML inputs and outputs.
- Develop infrastructure and allocate resources.

Terms and Technologies

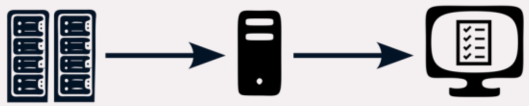


Deployment

Deploying a model to a ***server*** makes it **accessible** to other systems or users within a **network**.

Development Environment

Before deployment, the **full pipeline** should be **robustly tested** in an offline "**sandbox**" that is **completely isolated** from clinical workflows.



MLOps

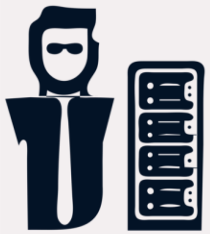
The framework for **building**, **deploying**, and **monitoring** end-to-end ML solutions in live, **production environments** safely and effectively.

Successful *Implementation* of Machine Learning Pipelines

Key Roles and Responsibilities

Software & ML Engineers

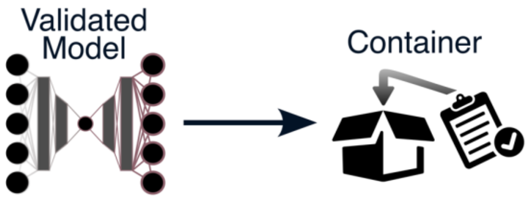
- Build robust and secure prediction pipelines.
- Implement best practices in DevOps/MLOps.



Information Technology & Systems

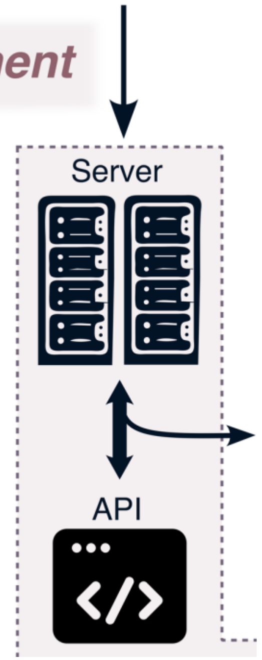
- Maintain interfaces for ML inputs and outputs.
- Develop infrastructure and allocate resources.

Terms and Technologies



Deployment

Deploying a model to a ***server*** makes it **accessible** to other systems or users within a **network**.

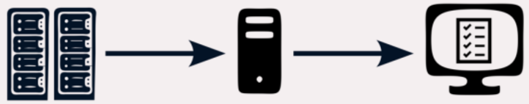


MLOps

The framework for **building**, **deploying**, and **monitoring** end-to-end ML solutions in live, **production environments** safely and effectively.

Development Environment

Before deployment, the **full pipeline** should be **robustly tested** in an offline "**sandbox**" that is **completely isolated** from clinical workflows.



Logging Metrics

Latency: *Turn-around time for predictions.*

Uptime: *% of time a prediction can be made.*

Scalability: *Change in latency/uptime with increased volume.*

Successful *Implementation* of Machine Learning Pipelines

Key Roles and Responsibilities

Data Scientists & Data Engineers

- Build and evaluate models for deployment.
- Optimize storage and retrieval of input data.

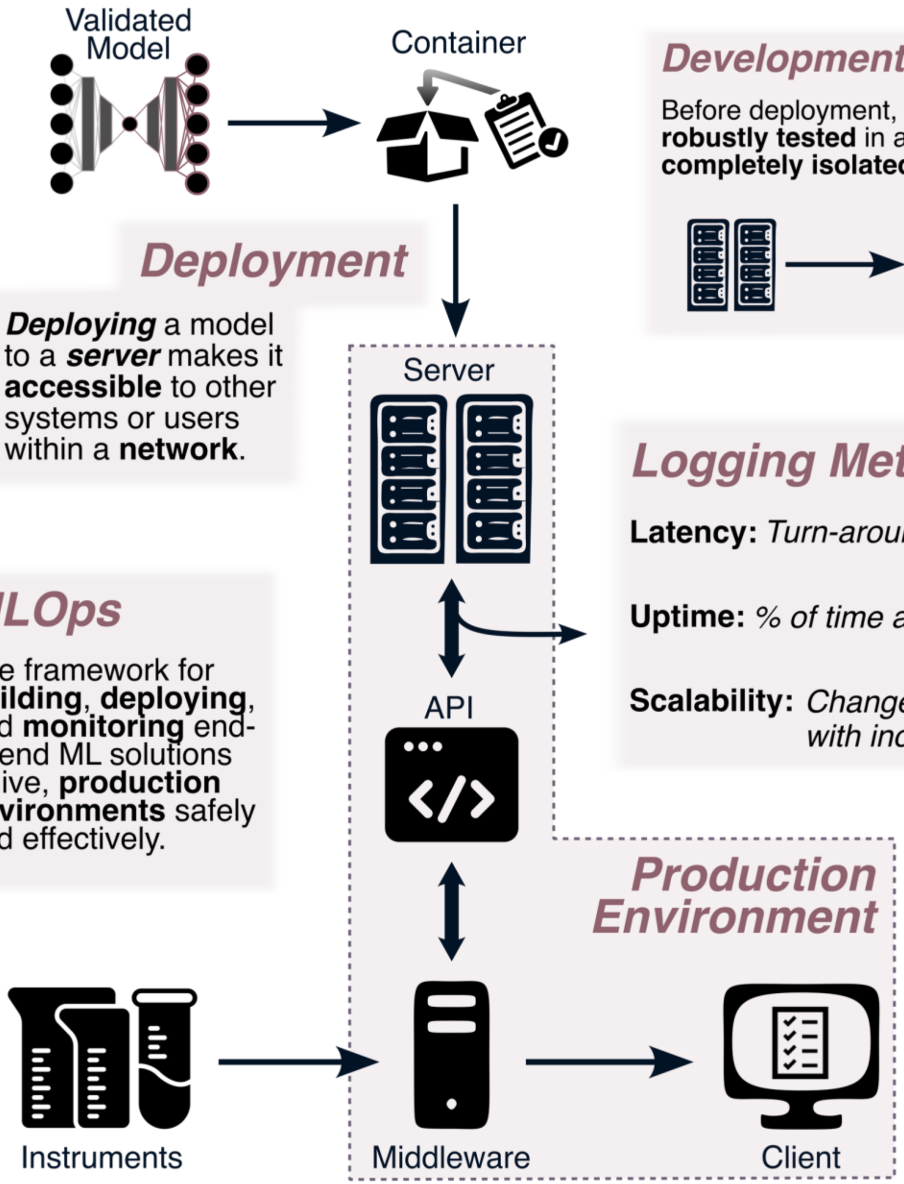
Software & ML Engineers

- Build robust and secure prediction pipelines.
- Implement best practices in DevOps/MLOps.

Information Technology & Systems

- Maintain interfaces for ML inputs and outputs.
- Develop infrastructure and allocate resources.

Terms and Technologies



Development Environment

Before deployment, the **full pipeline** should be **robustly tested** in an offline "**sandbox**" that is **completely isolated** from clinical workflows.

Deployment

Deploying a model to a ***server*** makes it **accessible** to other systems or users within a **network**.

MLOps

The framework for **building**, **deploying**, and **monitoring** end-to-end ML solutions in live, **production environments** safely and effectively.

Logging Metrics

Latency: *Turn-around time for predictions.*

Uptime: *% of time a prediction can be made.*

Scalability: *Change in latency/uptime with increased volume.*

Successful *Implementation* of Machine Learning Pipelines

Key Roles and Responsibilities

Subject Matter Experts

- Align implementation to fit unmet clinical need.
- Evaluate failure modes and off-target effects.

Data Scientists & Data Engineers

- Build and evaluate models for deployment.
- Optimize storage and retrieval of input data.

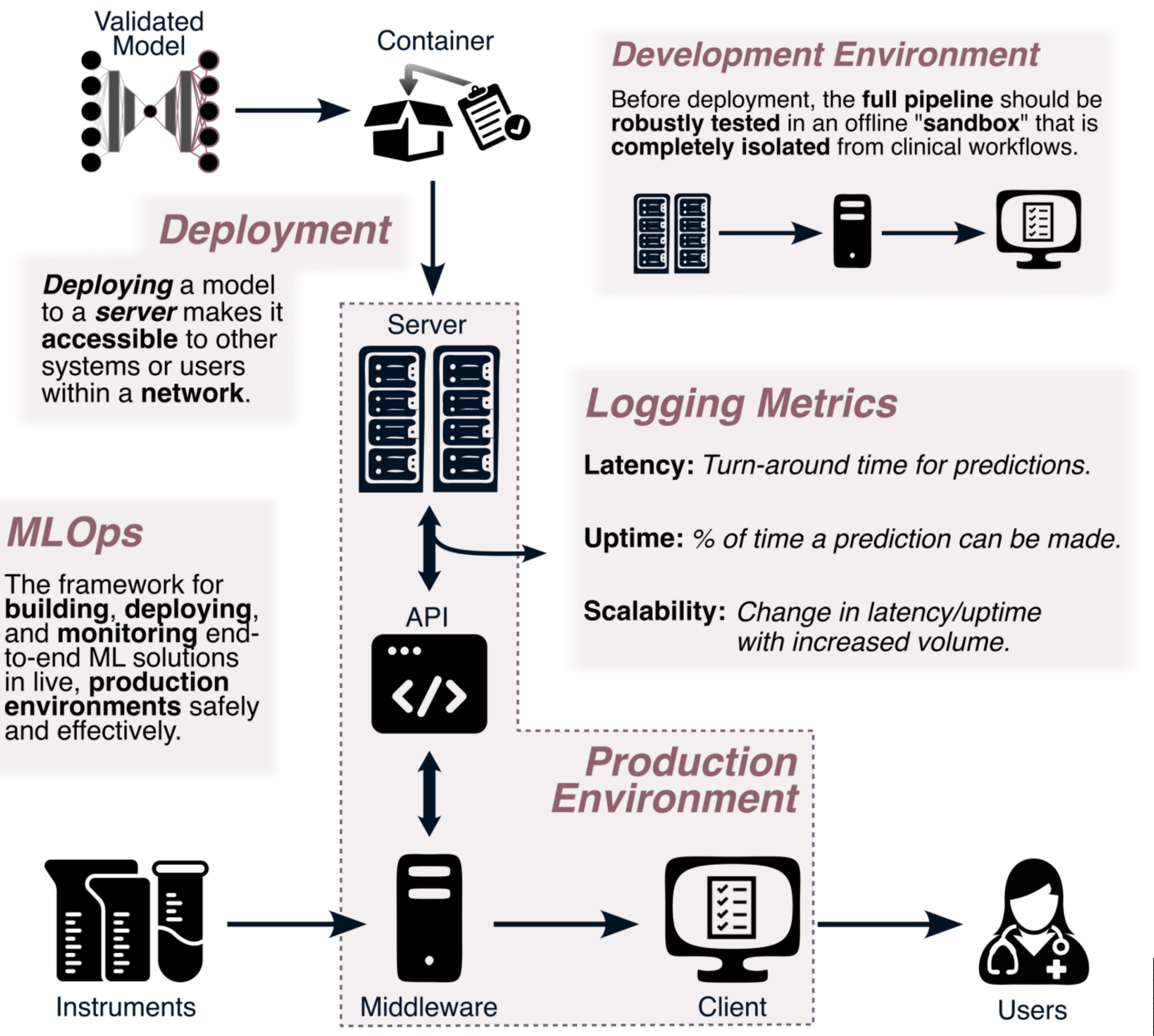
Software & ML Engineers

- Build robust and secure prediction pipelines.
- Implement best practices in DevOps/MLOps.

Information Technology & Systems

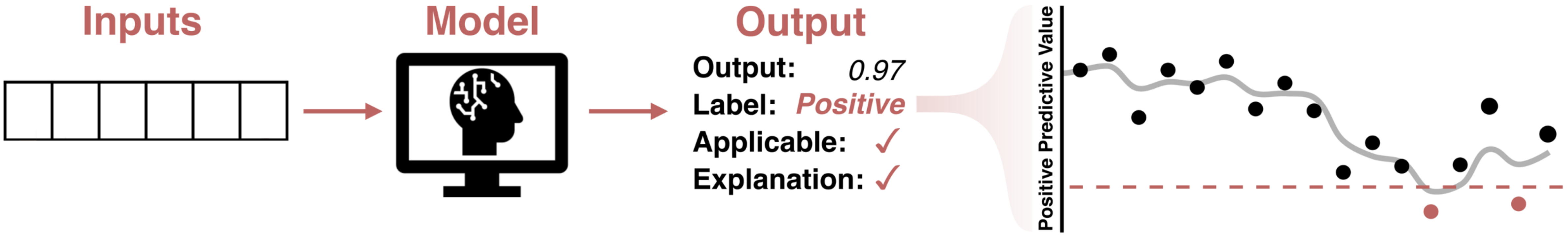
- Maintain interfaces for ML inputs and outputs.
- Develop infrastructure and allocate resources.

Terms and Technologies

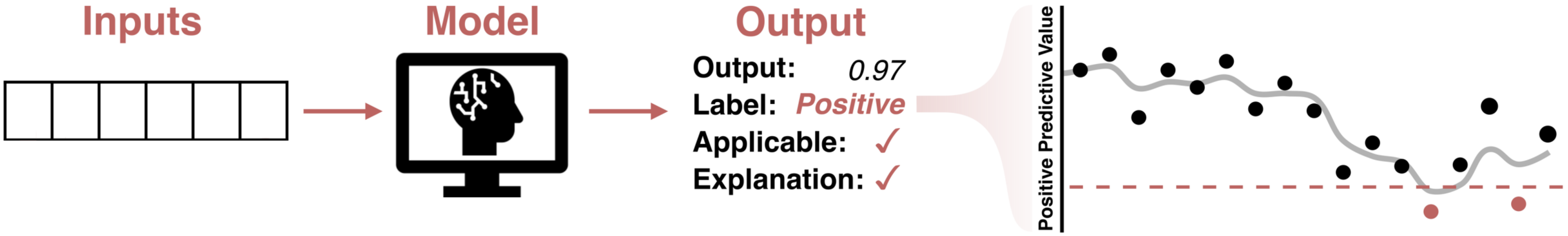


How do we ***monitor*** an implemented model?

*Successful **Monitoring** of Machine Learning Pipelines*



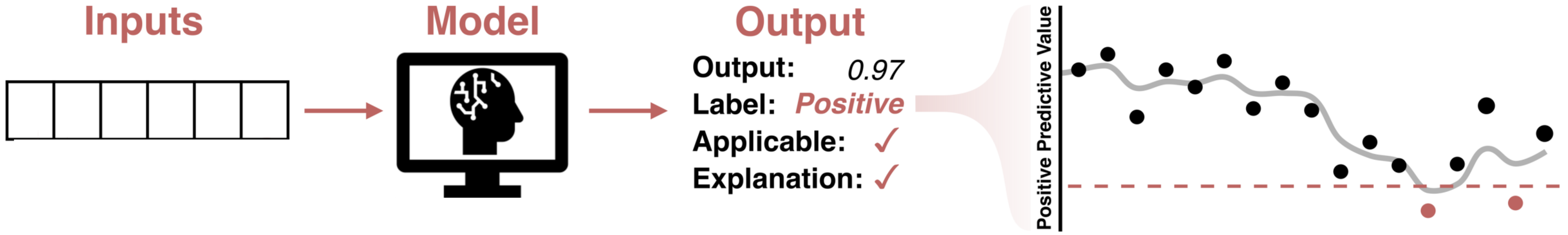
*Successful **Monitoring** of Machine Learning Pipelines*



Performance Drift

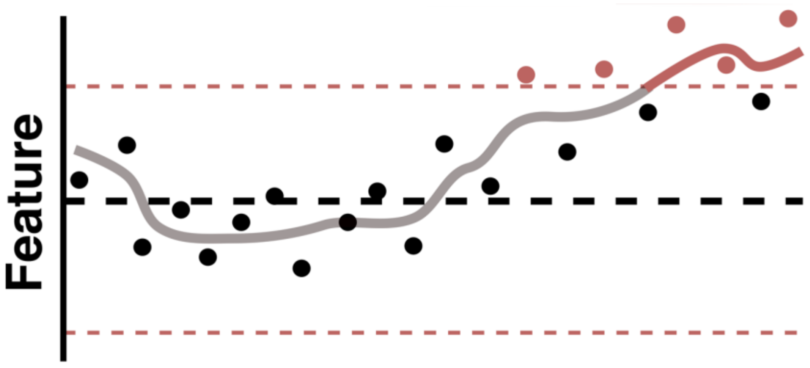
Data drift or ***concept drift*** causes ML pipelines to lose performance. Closed-loop systems are crucial for identifying when this occurs.

Successful *Monitoring* of Machine Learning Pipelines



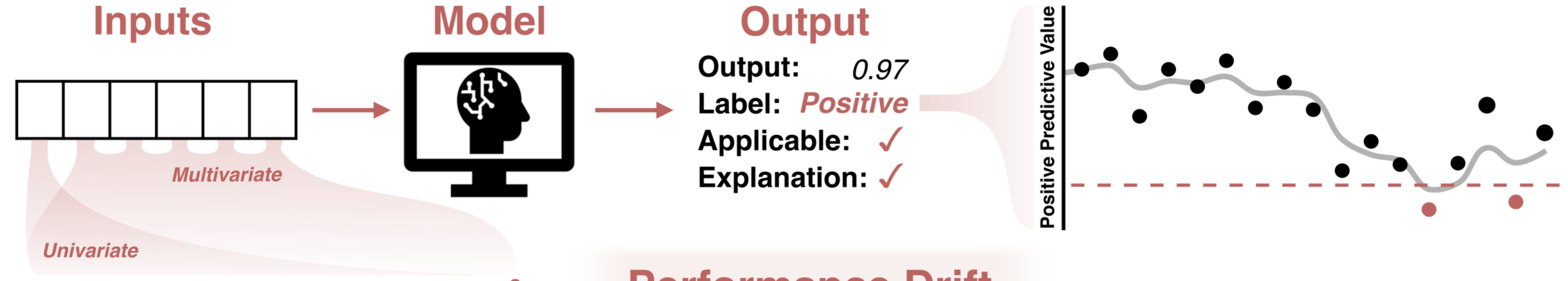
Performance Drift

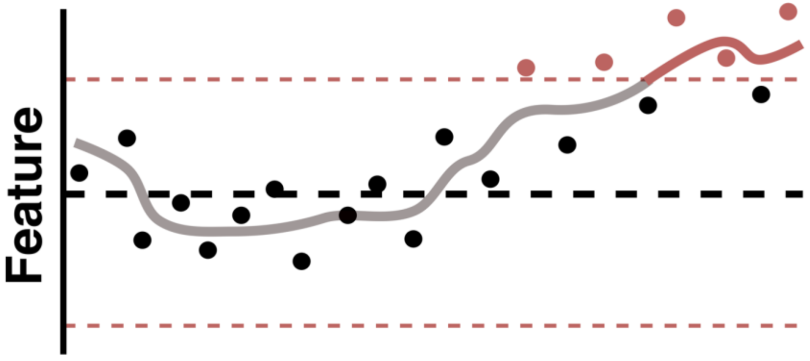
Data drift or ***concept drift*** causes ML pipelines to lose performance. Closed-loop systems are crucial for identifying when this occurs.



Data Drift

*Successful **Monitoring** of Machine Learning Pipelines*



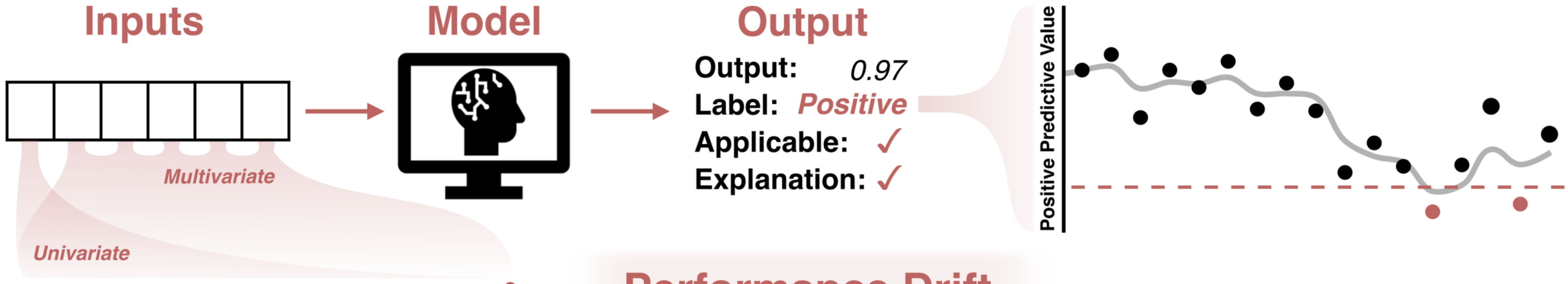


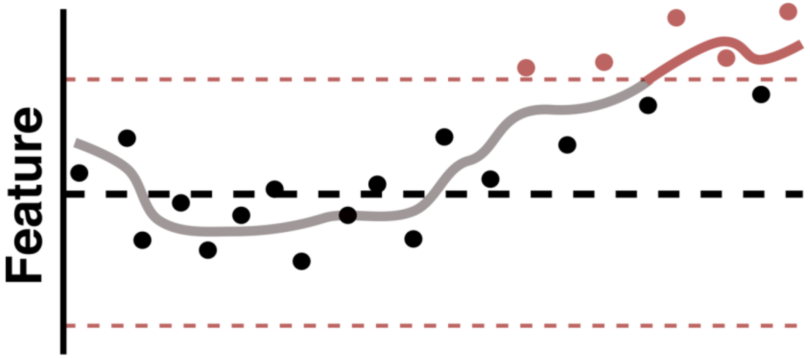
Performance Drift

Data drift or ***concept drift*** causes ML pipelines to lose performance. Closed-loop systems are crucial for identifying when this occurs.

Data Drift

Successful *Monitoring* of Machine Learning Pipelines





Performance Drift

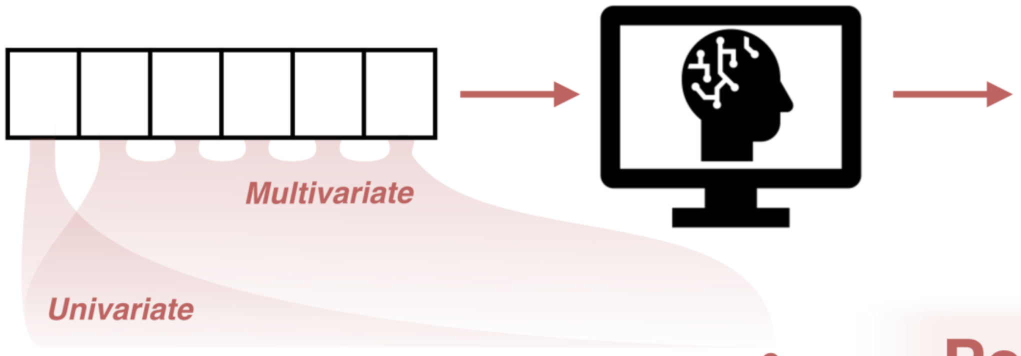
Data drift or ***concept drift*** causes ML pipelines to lose performance. Closed-loop systems are crucial for identifying when this occurs.

Data Drift

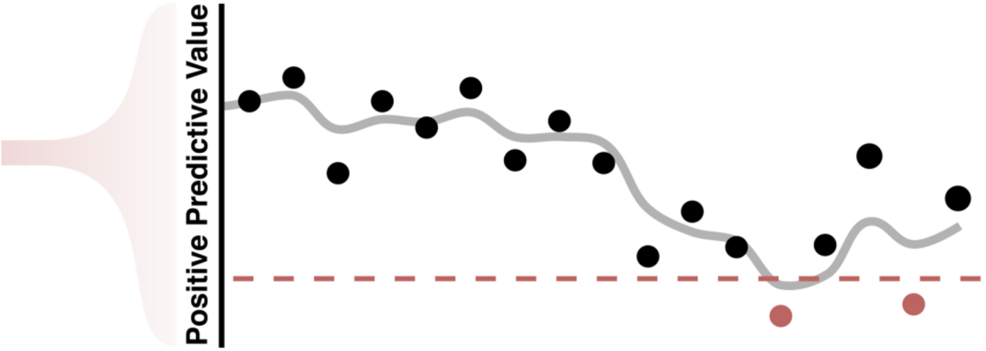
	Detection	*Correction*
Uni-	- Threshold Flags - Moving Averages	- Input Preprocessing - Analyzer Recalibration
Multi-	- Principal Components - Mahalanobis Distance	- Input Transformation - Model Retraining

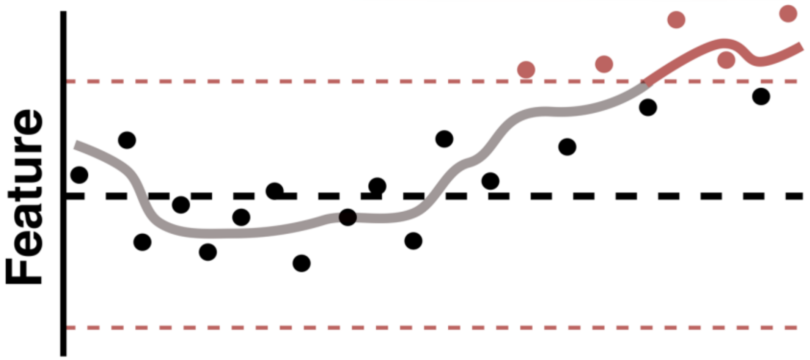
*Successful **Monitoring** of Machine Learning Pipelines*

Inputs **Model** **Output**



Output: *0.97*
Label: ***Positive***
Applicable: ✓
Explanation: ✓





Performance Drift

Data drift or ***concept drift*** causes ML pipelines to lose performance.
Closed-loop systems are crucial for identifying when this occurs.

Concept Drift

Occurs when **real-world labels diverge** from **training** labels.
Difficult to detect and correct without input from **subject-matter experts**.

Data Drift

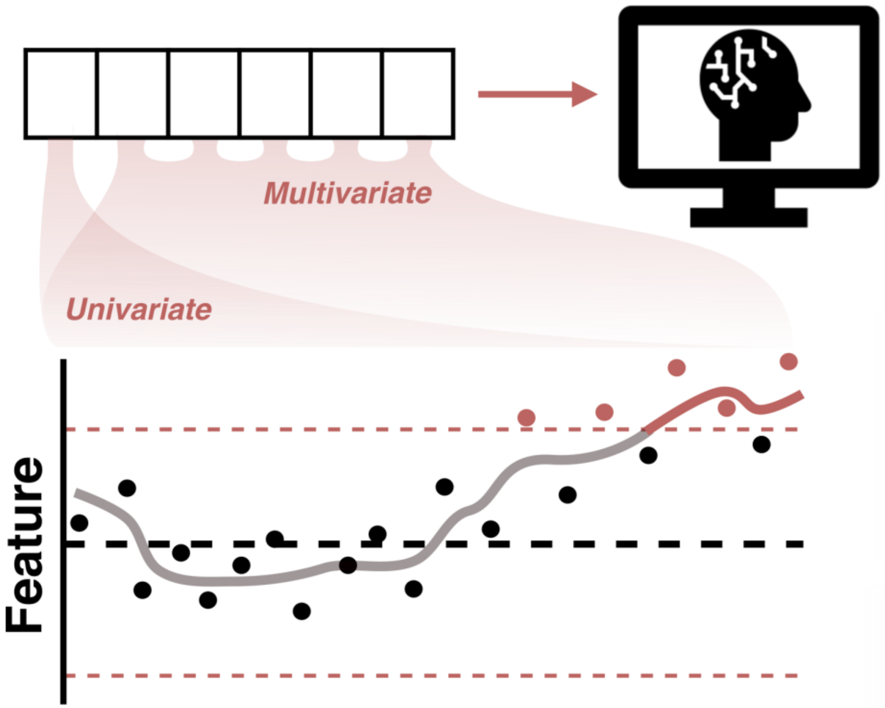
	Detection	*Correction*
Uni-	- Threshold Flags - Moving Averages	- Input Preprocessing - Analyzer Recalibration
Multi-	- Principal Components - Mahalanobis Distance	- Input Transformation - Model Retraining

Successful *Monitoring* of Machine Learning Pipelines

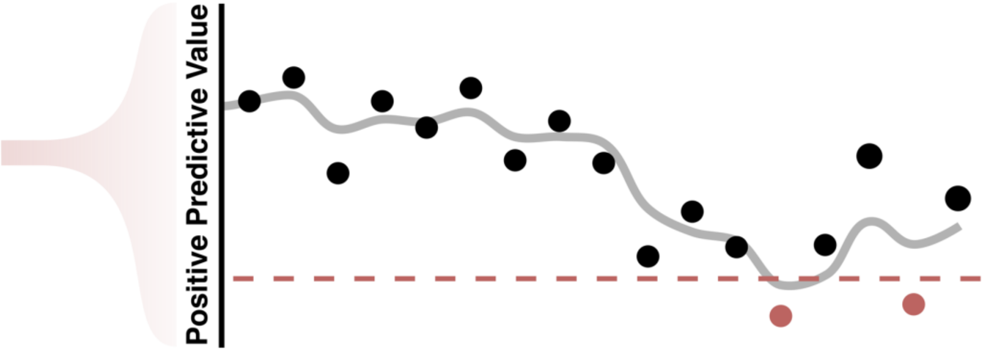
Inputs

Model

Output



Output: *0.97*
Label: *Positive*
Applicable: ✓
Explanation: ✓



Performance Drift

Data drift or ***concept drift*** causes ML pipelines to lose performance.
Closed-loop systems are crucial for identifying when this occurs.

Concept Drift

Occurs when **real-world labels diverge** from **training** labels.
Difficult to detect and correct without input from **subject-matter experts**.

Data Drift

	Detection	*Correction*
Uni-	- Threshold Flags - Moving Averages	- Input Preprocessing - Analyzer Recalibration
Multi-	- Principal Components - Mahalanobis Distance	- Input Transformation - Model Retraining

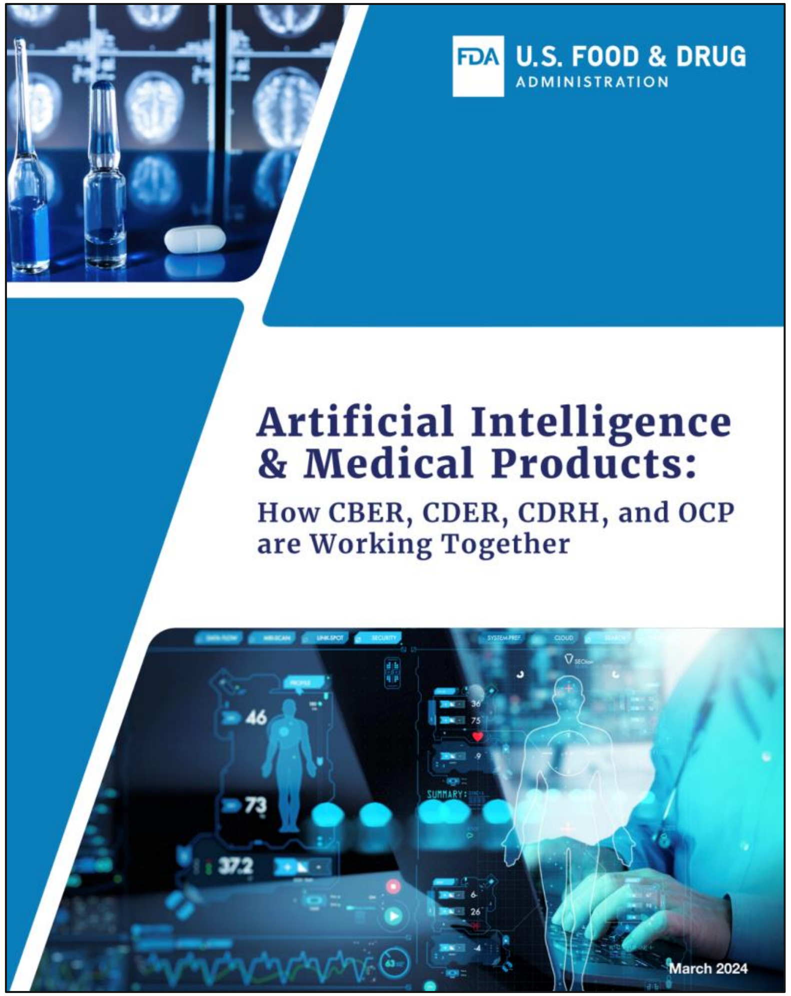
Updating Models

Champion

Replacement models can be **continuously retrained and evaluated** to replace deteriorating models before they impact live workflows.

Challengers

A Note On Regulatory Guidance



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

Predetermined Change Control Plans for Machine Learning-Enabled Medical Devices: Guiding Principles

October 2023

In 2021, the U.S. Food and Drug Administration (FDA), Health Canada, and the U.K.'s Medicines and Healthcare products Regulatory Agency (MHRA) jointly identified 10 guiding principles that can inform the development of Good Machine Learning Practice (GMLP). GMLP supports the development of safe, effective, and high-quality artificial intelligence/machine learning technologies that can learn from real-world use and, in some cases, improve device performance.

In this document, FDA, Health Canada, and MHRA jointly identified 5 guiding principles for predetermined change control plans. These principles draw upon the overarching GMLP guiding principles, in particular principle 10, which states that deployed models are monitored for performance and re-training risks are managed.

Advancements in digital health technologies include artificial intelligence/machine learning-enabled medical devices (MLMD). Regulatory expectations that are aligned with best practices for development and change management, such as those described in the GMLP Guiding Principles, can help to support the quality of such devices. Ultimately, this can lead to patient benefits such as earlier access to innovative technologies or more accurate diagnoses.

The change management process helps to ensure the ongoing safety and effectiveness of devices in the face of change throughout the device's total product lifecycle (TPLC). However, certain changes to MLMDs, such as changes to a model or algorithm, may be substantive or significant. For this reason, they can require regulatory oversight, such as additional premarket review. Such regulatory expectations may not always coincide with the rapid pace of MLMD development.

Internationally, the medical device community is discussing the use of predetermined change control plans (PCCPs) as a way of managing certain device changes where regulatory authorization before marketing is typically required. PCCPs can be used to help:

- align regulatory processes with the rapid and ongoing approach to change management in MLMDs
- manage risks in a timely and ongoing fashion through monitoring, maintenance, and/or improving device performance
- uphold high regulatory standards to ensure device safety and effectiveness.

For this document, the term PCCP describes a plan, proposed by a manufacturer, that specifies:

- certain planned modifications to a device
- the protocol for implementing and controlling those modifications and
- the assessment of impacts from modifications.

PCCPs may be developed and implemented in different ways in different regulatory jurisdictions.

One key objective of the 5 Guiding Principles for PCCPs for MLMD is to provide foundational considerations that highlight the characteristics of robust PCCPs. Another objective of this document is to facilitate and foster ongoing engagement and collaboration among stakeholders on the PCCP concept for MLMD. As with the GMLP Guiding Principles, this document intends to lay a foundation for PCCPs and encourages international harmonization.

International harmonization and stakeholder consensus on the core concepts of PCCPs will help support the advancement of responsible innovations in the digital health space.

We welcome your continued feedback through the FDA public docket (FDA-2019-N-1185) at Regulations.gov, and we look forward to engaging with you on these efforts. This work is being spearheaded by the Digital Health Center of Excellence for the FDA, the Medical Devices Directorate Digital Health Division at Health Canada and the software and AI team at the MHRA. Contact us directly at Digitalhealth@fda.hhs.gov, software@mhra.gov.uk, and mddpolicypolitiquesdim@hc-sc.gc.ca.

EUROPEAN COMMISSION

Brussels, 21.4.2021
COM(2021) 206 final

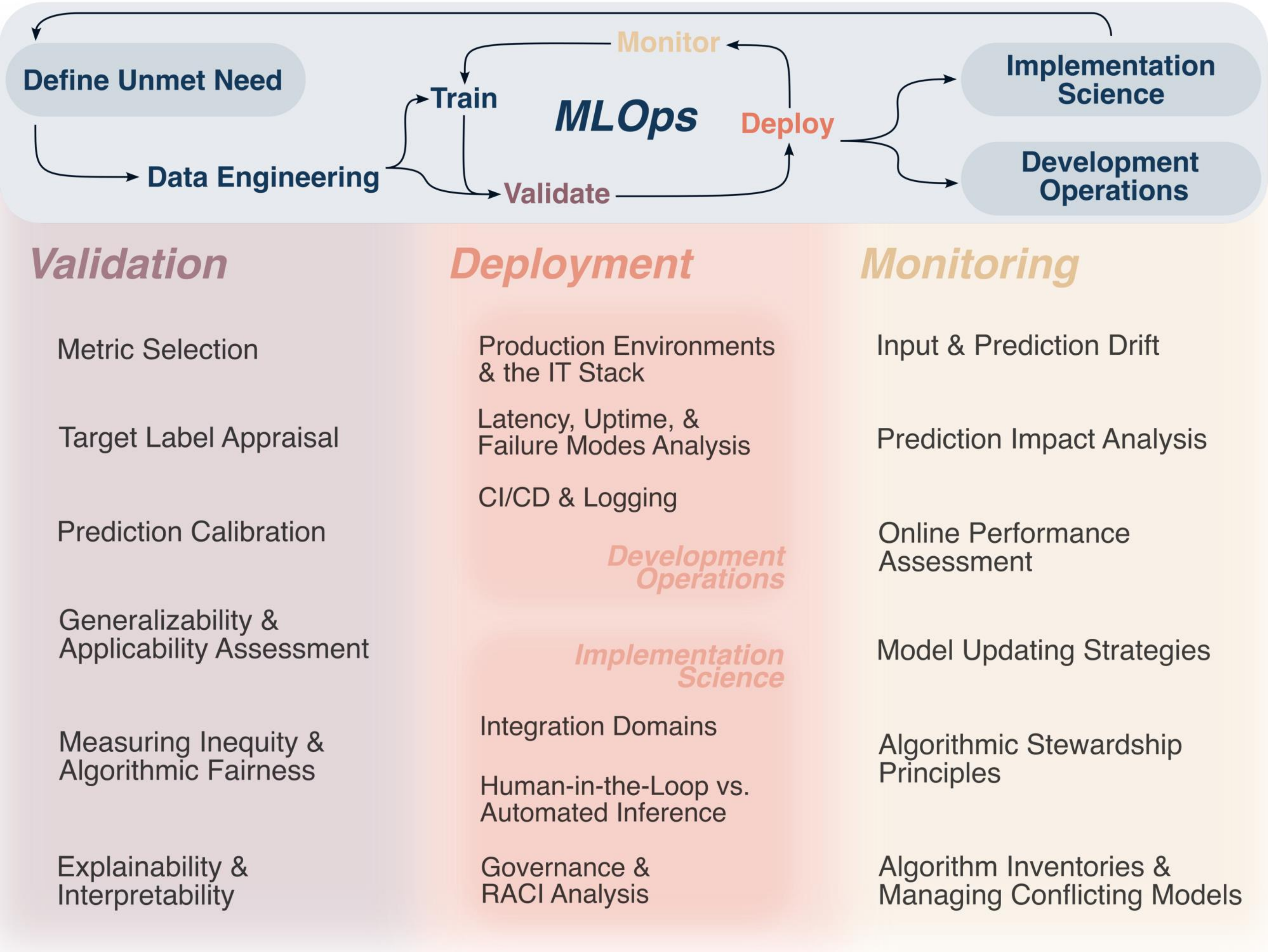
2021/0106 (COD)

Proposal for a

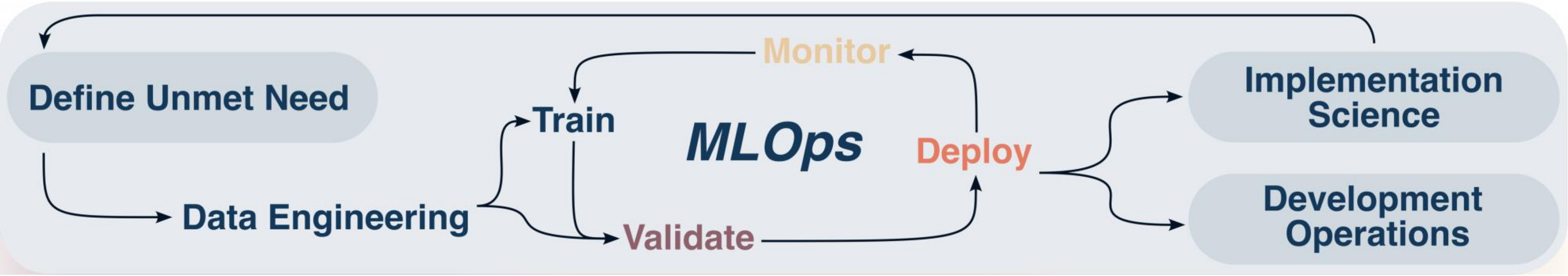
REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS

{SEC(2021) 167 final} - {SWD(2021) 84 final} - {SWD(2021) 85 final}



Define Unmet Need
Data Engineering
Train
Validate
Monitor
MLOps
Deploy
Implementation Science
Development Operations
Validation
Deployment
Monitoring
Metric Selection
Target Label Appraisal
Prediction Calibration
Generalizability & Applicability Assessment
Measuring Inequity & Algorithmic Fairness
Explainability & Interpretability
Production Environments & the IT Stack
Latency, Uptime, & Failure Modes Analysis
CI/CD & Logging
Development Operations
Implementation Science
Integration Domains
Human-in-the-Loop vs. Automated Inference
Governance & RACI Analysis
Input & Prediction Drift
Prediction Impact Analysis
Online Performance Assessment
Model Updating Strategies
Algorithmic Stewardship Principles
Algorithm Inventories & Managing Conflicting Models



JOURNAL ARTICLE

Validating, Implementing, and Monitoring Machine Learning Solutions in the Clinical Laboratory Safely and Effectively

Nicholas C Spies ✉, Christopher W Farnsworth, Sarah Wheeler, Christopher R McCudden

Clinical Chemistry, hvae126, https://doi.org/10.1093/clinchem/hvae126

Published: 10 September 2024 **Article history** ▾

Questions?

Email: nick.spies@aruplab.com