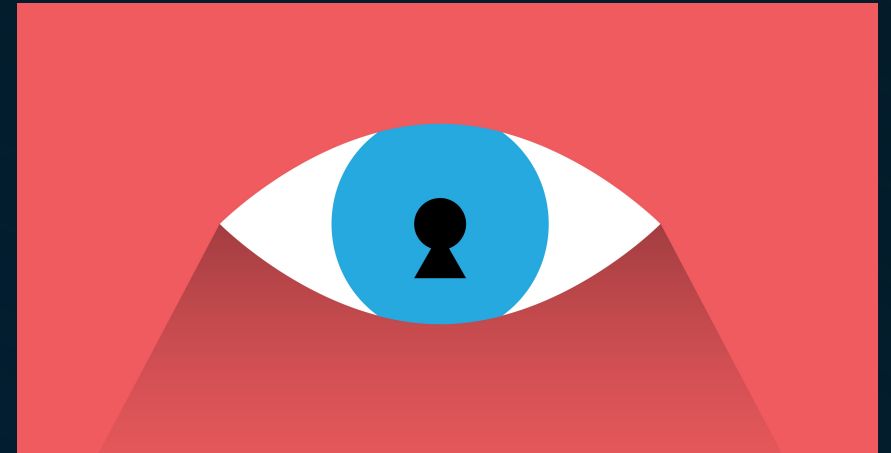


XAI: Interpretability of Artificial Intelligence to ensure Ethics in Drug Development

PHUSE : EU Connect 2020

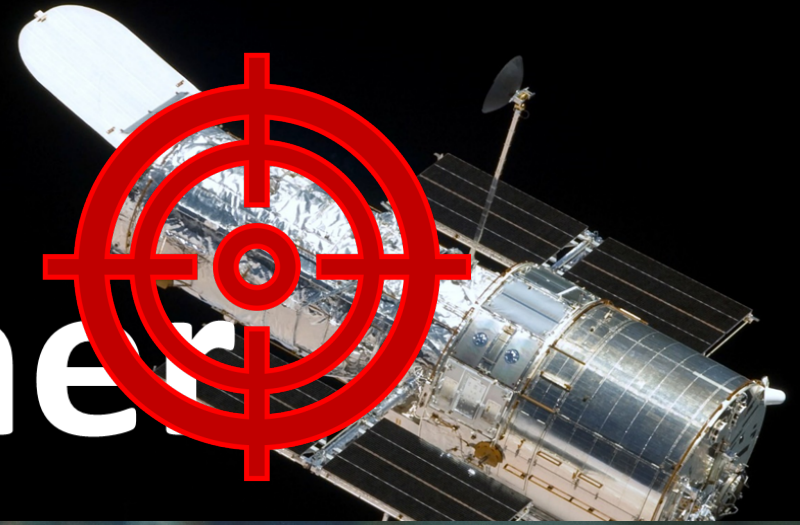
November 2020





See

Further

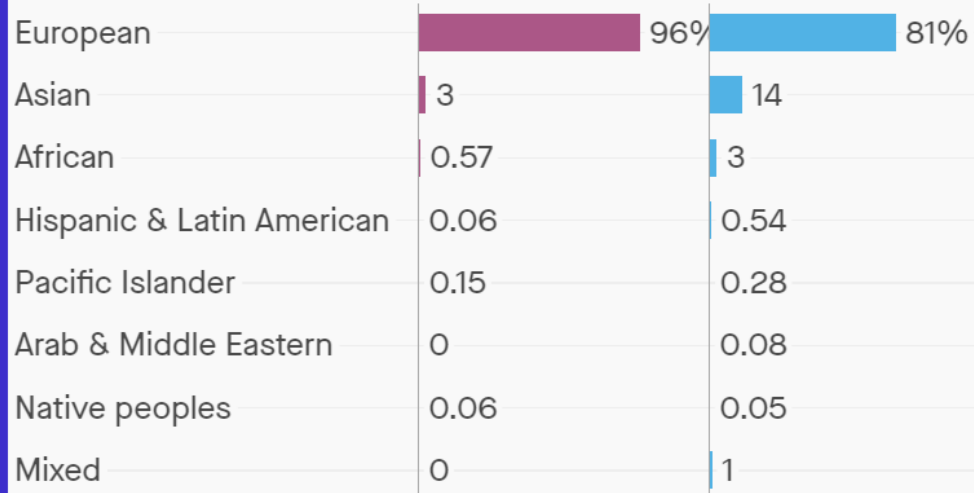


AI / Ethics & Pharma

Little progress is being made to improve diversity in genomics

Share of samples in genetic studies, by ancestry

■ 373 studies, up to 2009 ■ 2,511 studies, up to 2016



△ T L △ S | Data: Popejoy & Fullerton. Nature, 2016

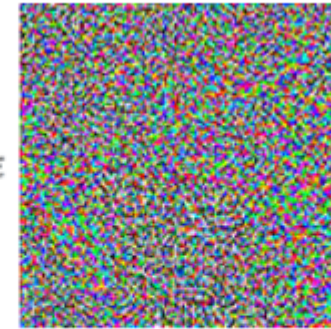
Share



"panda"

57.7% confidence

+ ϵ



=



"gibbon"

99.3% confidence

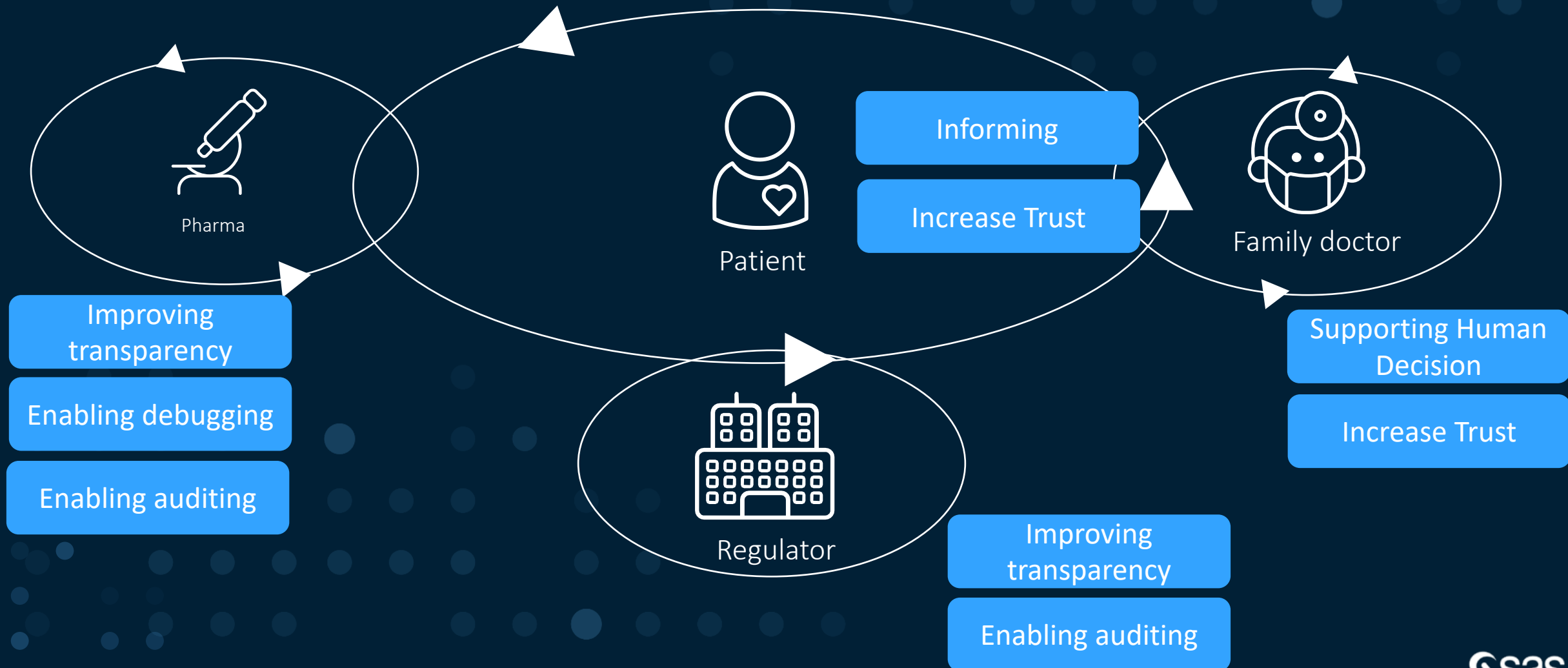
Dr. Nathwani added that he hoped to soon see AI systems that can help explain other AI systems—to solve the “black box” problem and produce explainable AI, or XAI. Both doctors and patients deserve a natural, human explanation for a machine-made decision, he said.

What are we talking about

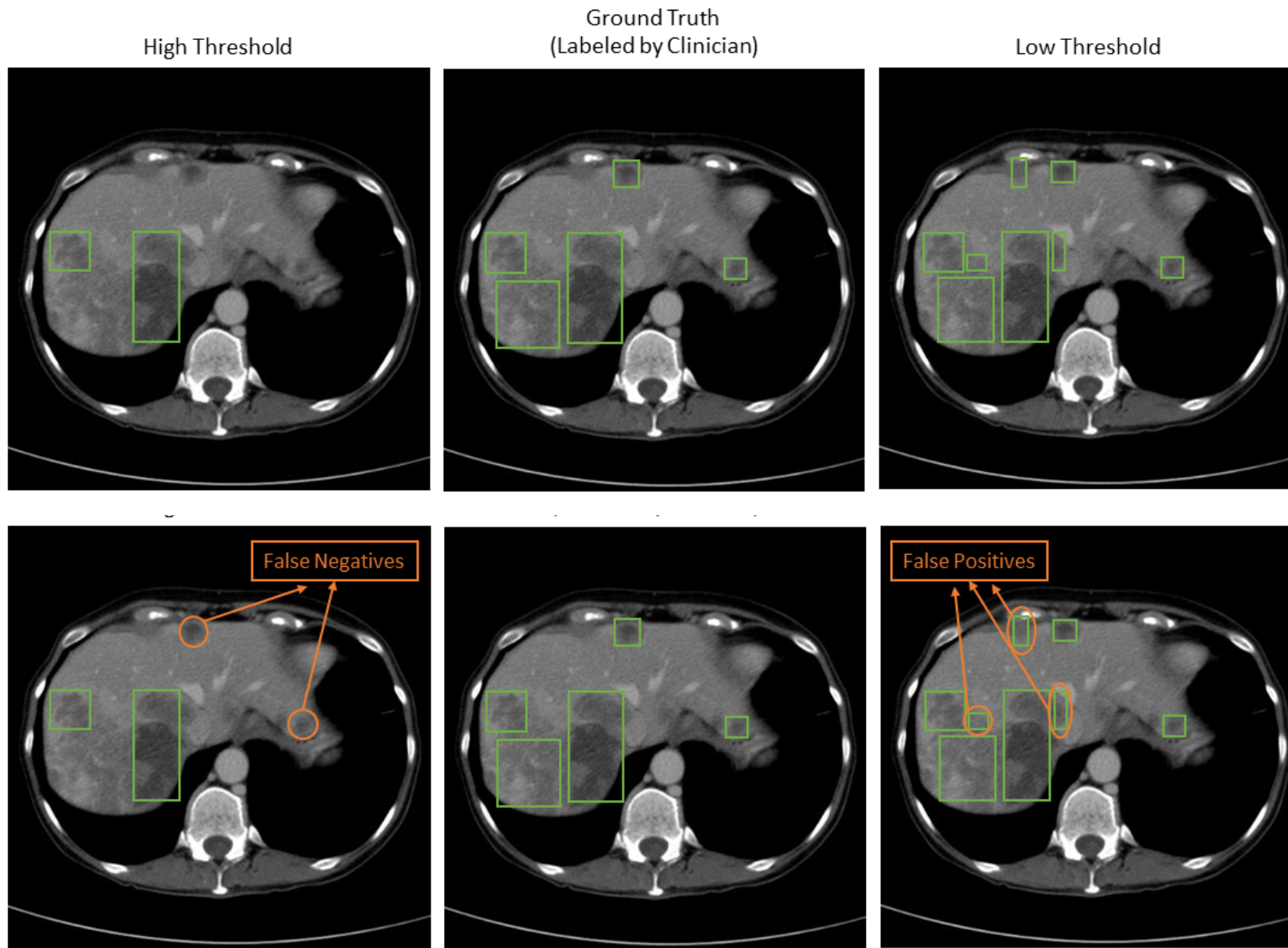
Terminology Clarification

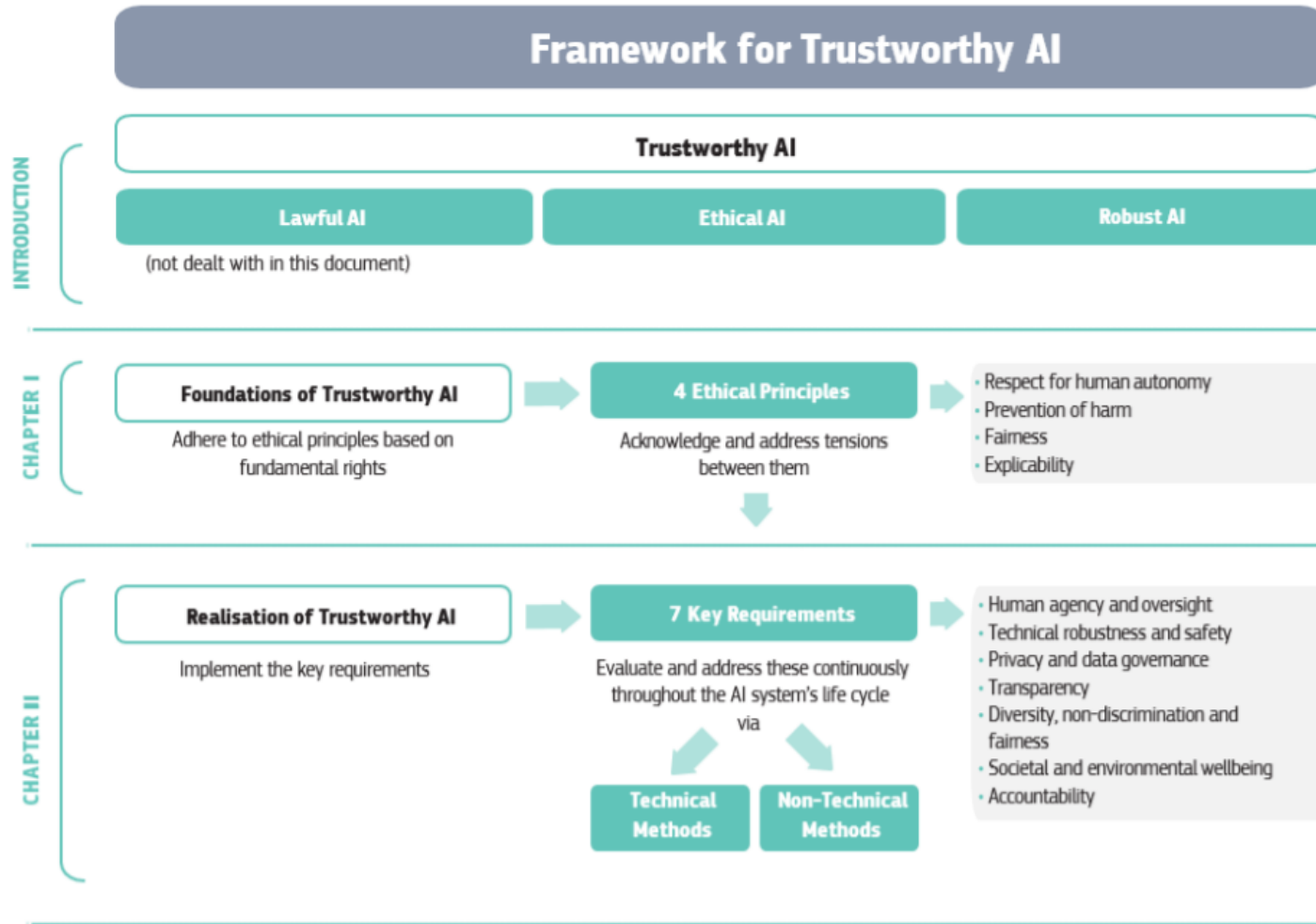
- Understandability denotes the characteristic of a model to make a human understand how the model works without explaining its internal structure or the algorithmic means by which the model processes data internally
- Comprehensibility: When conceived for ML models, refers to the ability of a learning algorithm to represent its learned knowledge in a human understandable fashion. (...) Given its difficult quantification, comprehensibility is normally tied to the evaluation of the model complexity.
- Interpretability: It is defined as the ability to explain or to provide the meaning in understandable terms to a human.
- Explainability: is associated with the notion of explanation as an interface between humans and a decision maker that is, at the same time, both an accurate proxy of the decision maker and comprehensible to humans.
- Transparency: A model is considered to be transparent if by itself it is understandable. (...) Transparent models (...) are divided into three categories: *simulatable models, decomposable models and algorithmically transparent models.*

XAI in the context of Patient Centricity



XAI in the Pharma Context





European
Commission

Strategy Shaping
Europe's digital

April 2019
Ethics guidelines
for trustworthy AI

Food & Drug Administration

Proposed Regulatory Framework for Modifications to AI/ML for SaMD

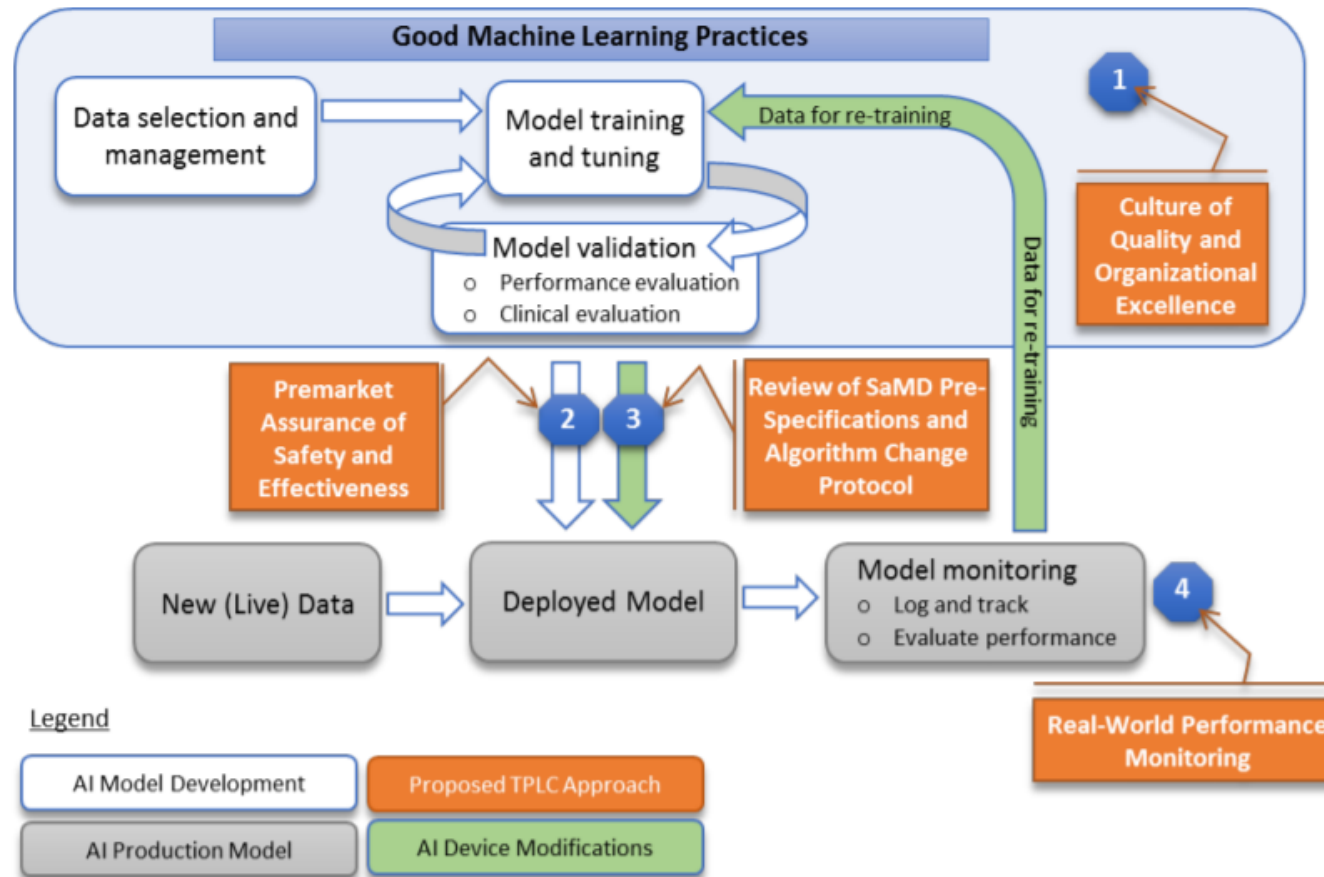


Figure 2: Overlay of FDA's TPLC approach on AI/ML workflow

AI & Interpretability

Questions

- *What are the main variables?*
- *How those variables impact prediction?*
- *How to explain a particular prediction?*

Techniques



Variable Importance



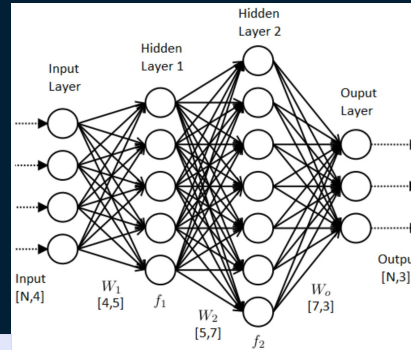
**Partial Dependence (PD)
Individual Conditional Expectation (ICE)**



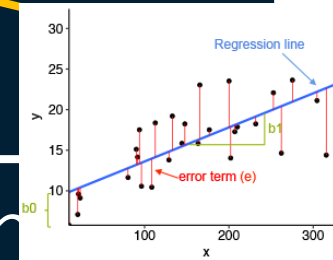
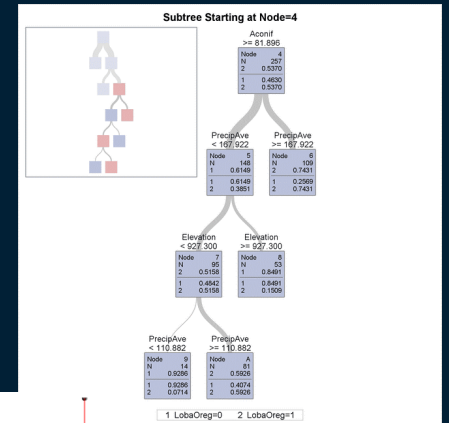
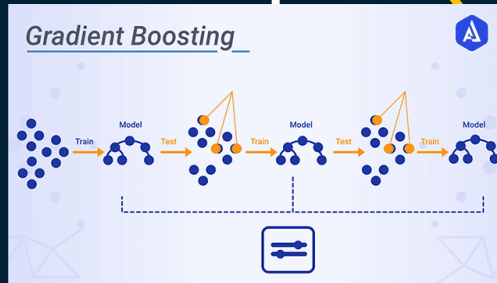
**Local Interpretable Model-agnostic
Explanation, Kernel SHAP, HyperSHAP**

Accuracy vs Interpretability

Accuracy

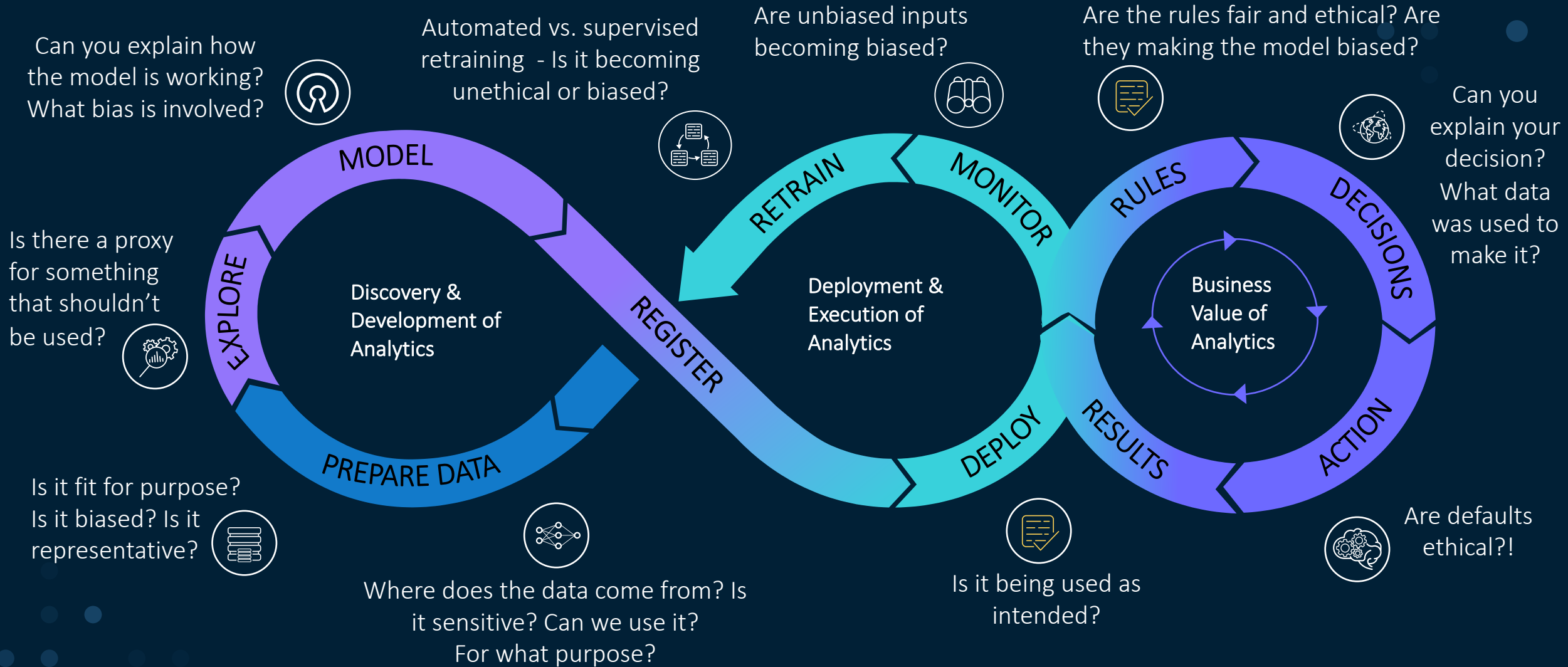


Nirvana



Interpretability

Bringing Ethics into the Analytic Lifecycle



From Black Box to Trusty AI



Why do you do that..

Why didn't you do differently..

When do you succeed..

When did you failed..

When do I trust you..

How do I correct an error..

I understand why..

I understand why not..

I know when you will succeed..

I know when you will fail..

I know when I can trust you..

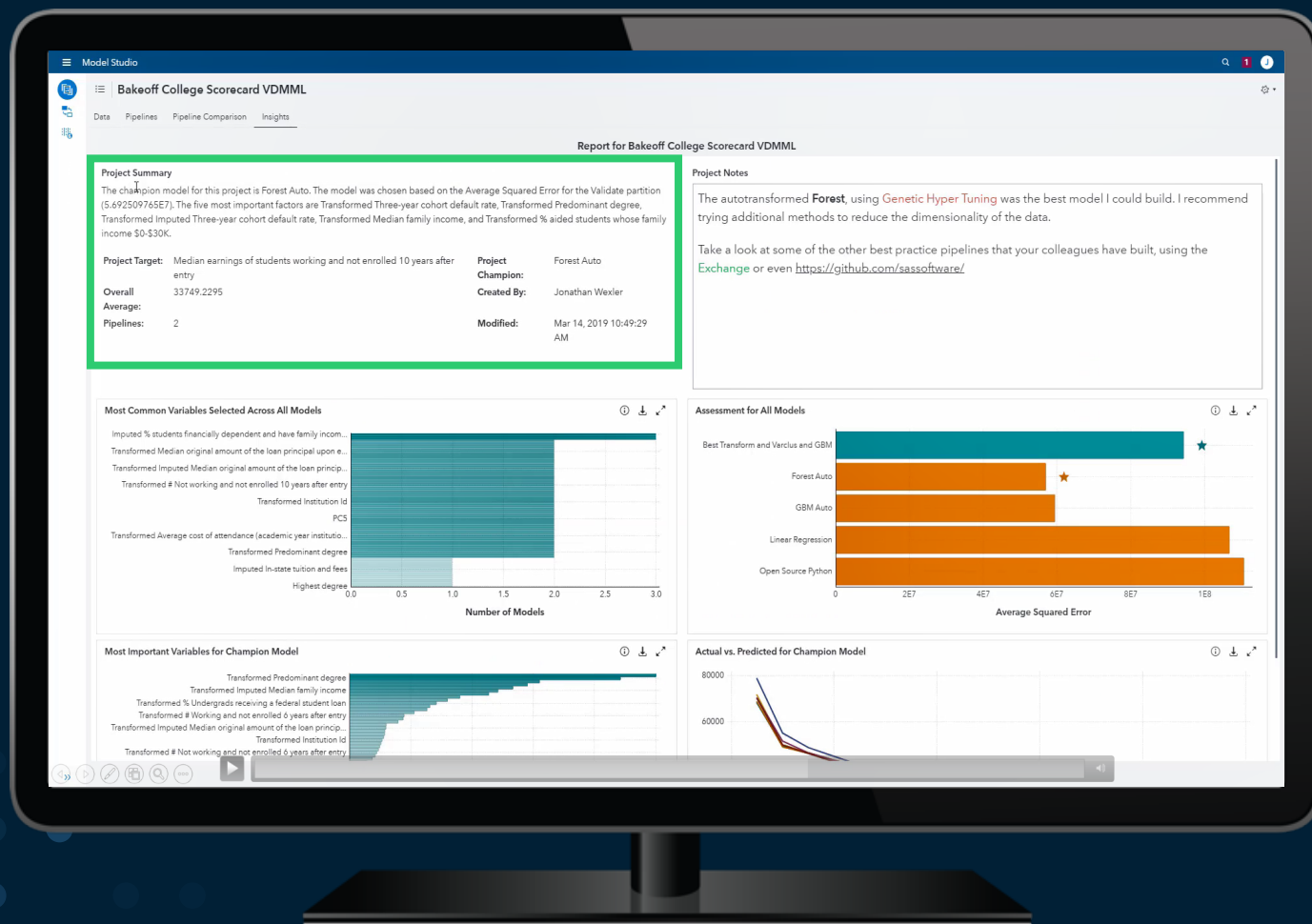
I know where you erred....

- Explainable Model
- Explanation interfaces

BLACK BOX / WEAK AI

WEALTHY / TRUSTED AI

Explainability & NLP Capabilities



- ✓ Model Interpretability Tab
- ✓ Explain models in plain language
- ✓ Project summaries using NLG
- ✓ PD – Partial Dependency
- ✓ LIME – Lcl. Interpret. Model-Ag. Expl.
- ✓ ICE – Individual Conditional Expl.
- ✓ SHAP – SHapley Additive exPlanations
- ✓ Collaboration w/ Project Notes