Towards Smart Detectors and Experiments: ML in Nuclear Physics

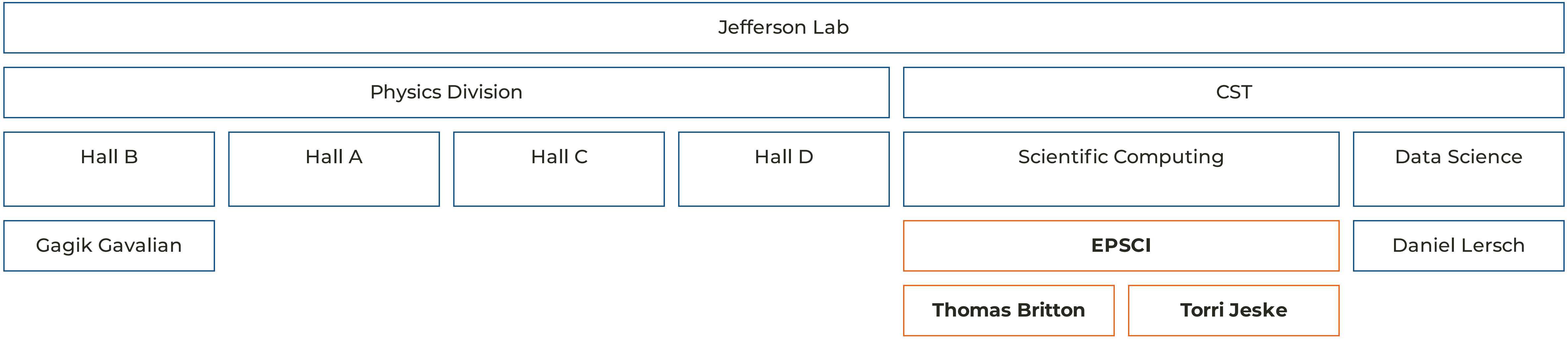
6/6/2024

Torri Jeske

roark@jlab.org

AI/ML at JLab

Multiple groups are working to implement AI/ML at each step of NP workflows.

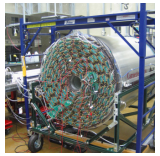


For some projects, EPSCI collaborates with the Data Science Department and Physics Division

EPSCI at JLab

Experimental Physics Software and Computing Infrastructure

Mission Statement: Identify, develop, implement, and maintain software and computing technologies in support of the Jefferson Lab Science Program



AI for Experimental Controls

Dynamically control and calibrate a detector in near real time

Hydra

Framework for training and managing AI for real time Data Quality Monitoring

JIRIAF

Test relocation of computing workflow from resources close to the experiment to remote data center for near real time data quality, calibration, and/or alignment.

EJFAT

Program FPGA for efficient network data routing of UDP packets, enabling low-latency, balanced distribution to compute farm endpoints.

PHASM

AI-based surrogate models of scientific code

AIOP

Control and optimized polarization for polarized targets and photon beams

Streaming DAQ

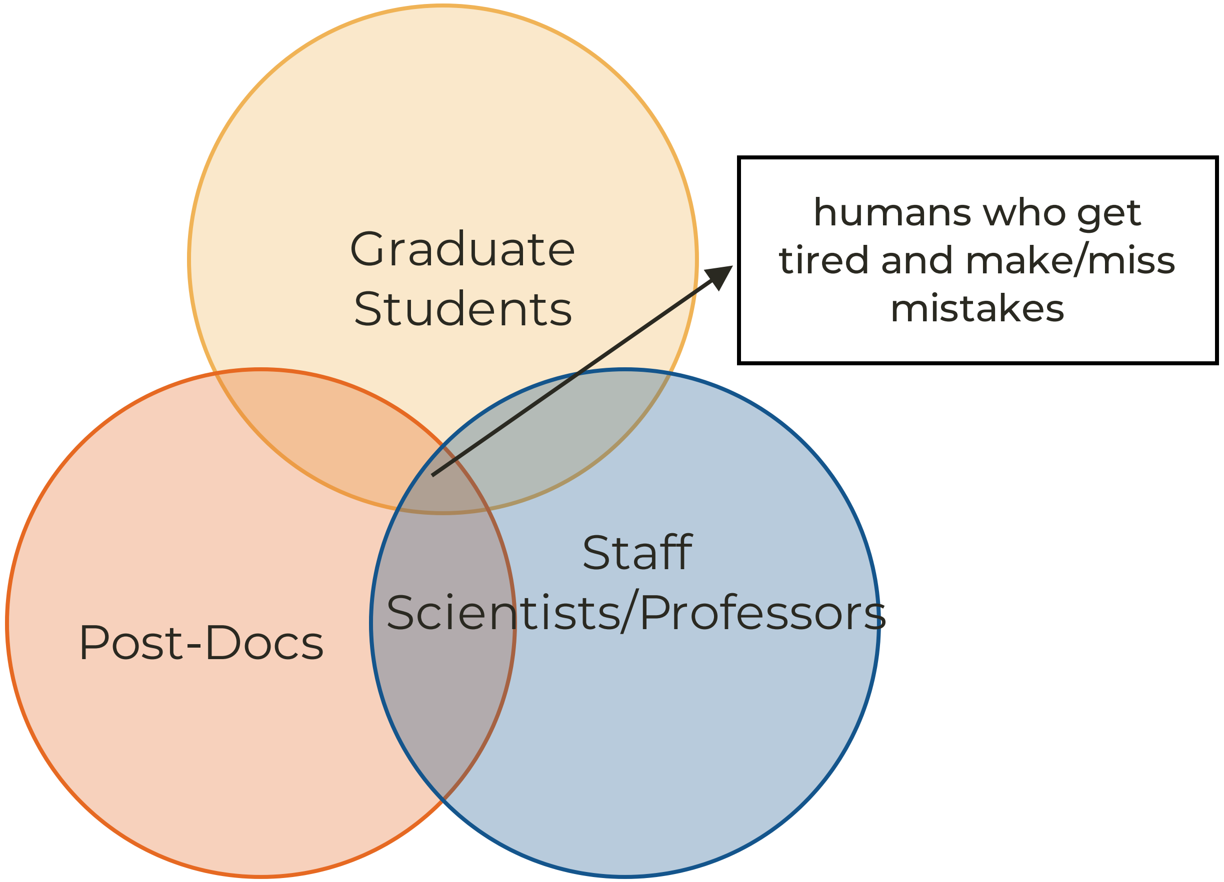
Developing the next generation of DAQ for Nuclear Physics

Experiment operations

1. **Design and Planning**
   Identify key questions and objectives for experiment
2. **Setup and Calibration**
   Installation, configuration, and alignment of detector systems
3. **Data Collection**
   Data acquisition, quality monitoring, etc
4. **Offline Tasks**
   Usually multiple rounds of calibration before proceeding with analysis
5. **Results**

Data quality monitoring: "On Shift" at JLab

Shifts are required as part of your membership to experimental hall collaboration



Responsible for

- DAQ operation
- Logging and responding to alarms
- **Data quality monitoring**
- Filling out BTA
- and much more

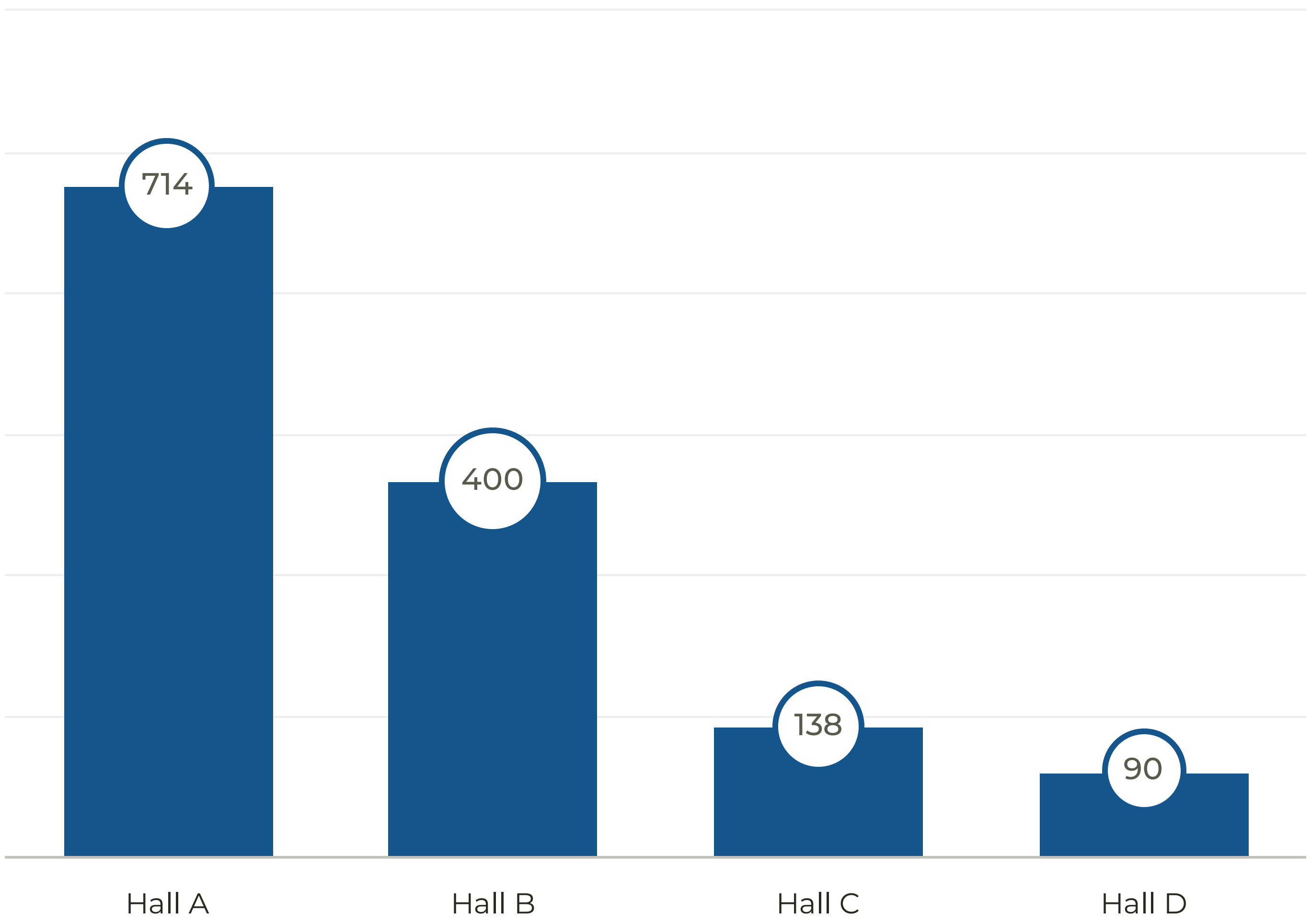
Online monitoring is **tedious**.

Varying levels of expertise

Inconsistent monitoring

Multiple plots per detector system

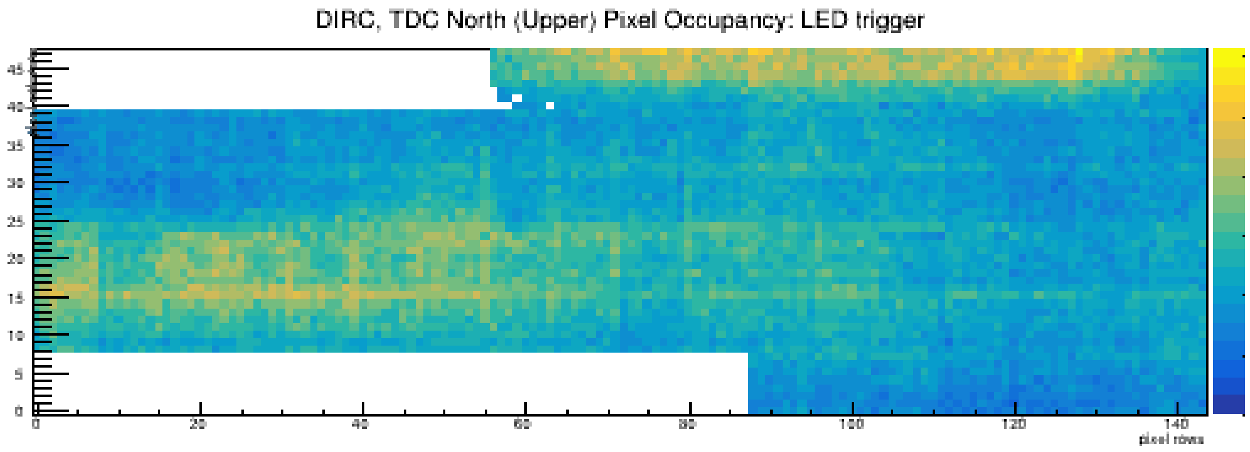
Too many plots to look at

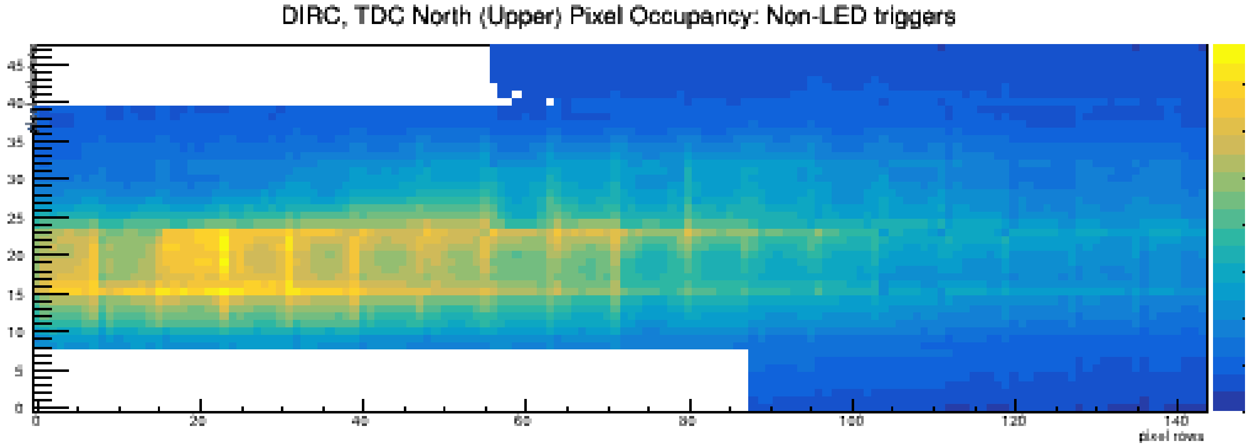


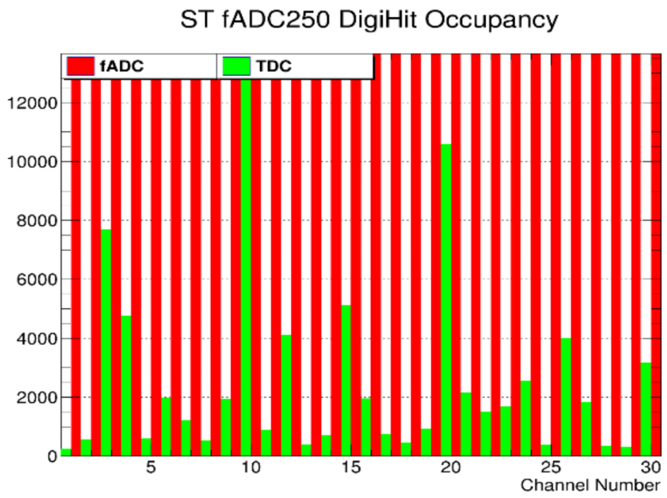
Approximate **number of individual histograms per experiment per run**, monitored by the shift crew for each experimental hall.

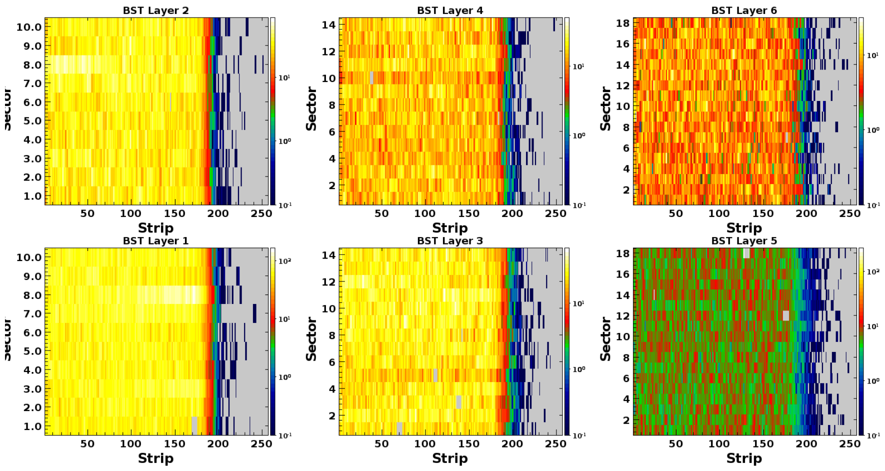
What makes these images bad?

It's hard to tell right away if an image is bad!



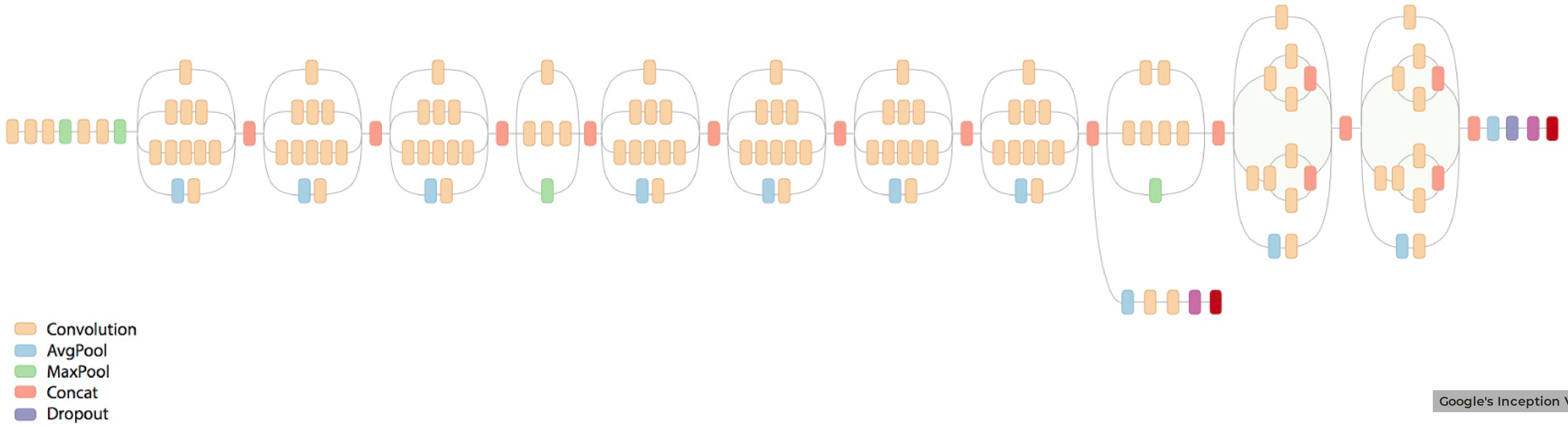






Can we utilize computer vision for data quality monitoring?

Thomas Britton's idea as a post doc in GlueX

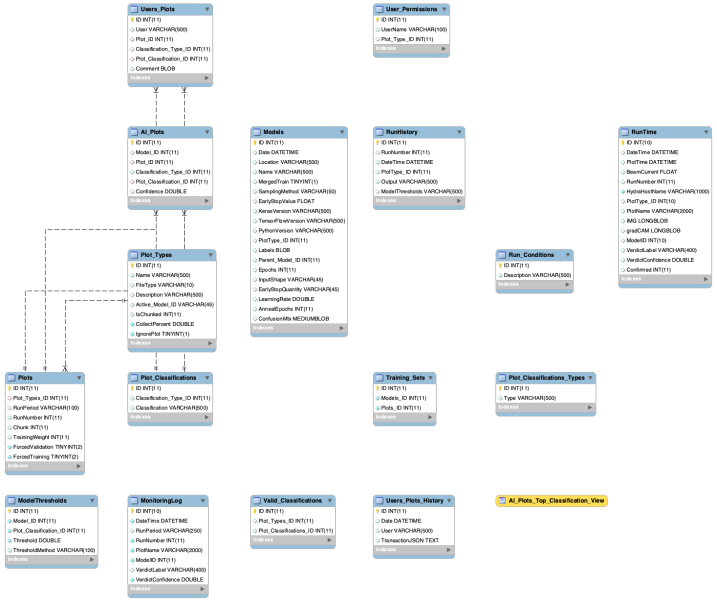


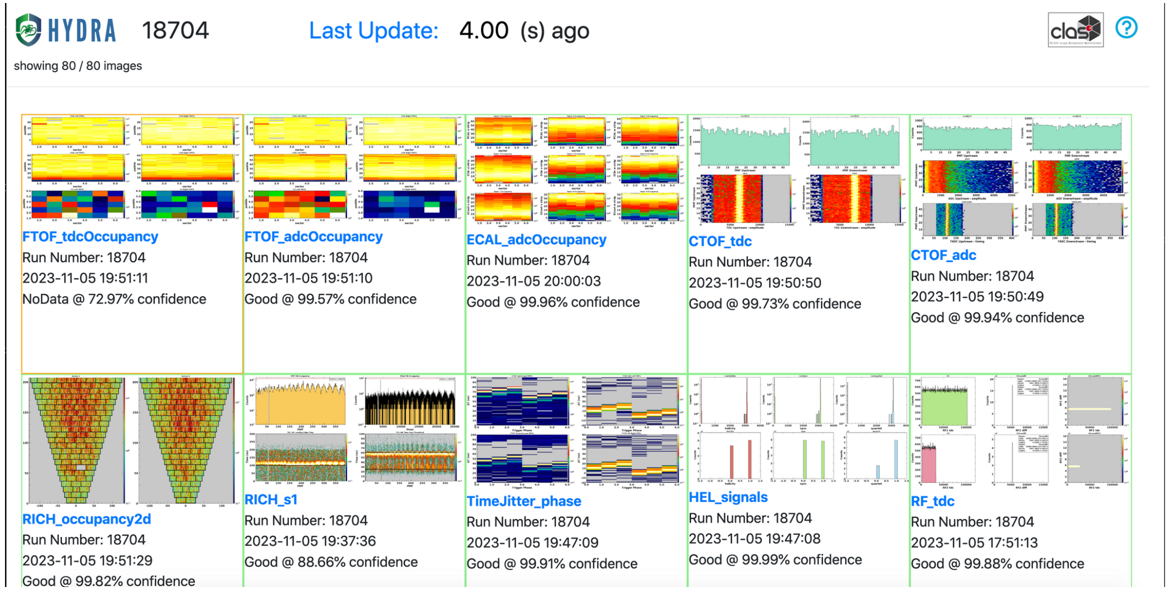
Google's Inception V3

Why InceptionV3? It was available and had good performance on standard image data sets.

HYDRA

An extensible framework for training, managing, and evaluating AI for real time data quality monitoring.





1/ MySQL back end

Unique plot identifiers, model training parameters, classifications, user permissions, labels, and more are all stored in MySQL database.

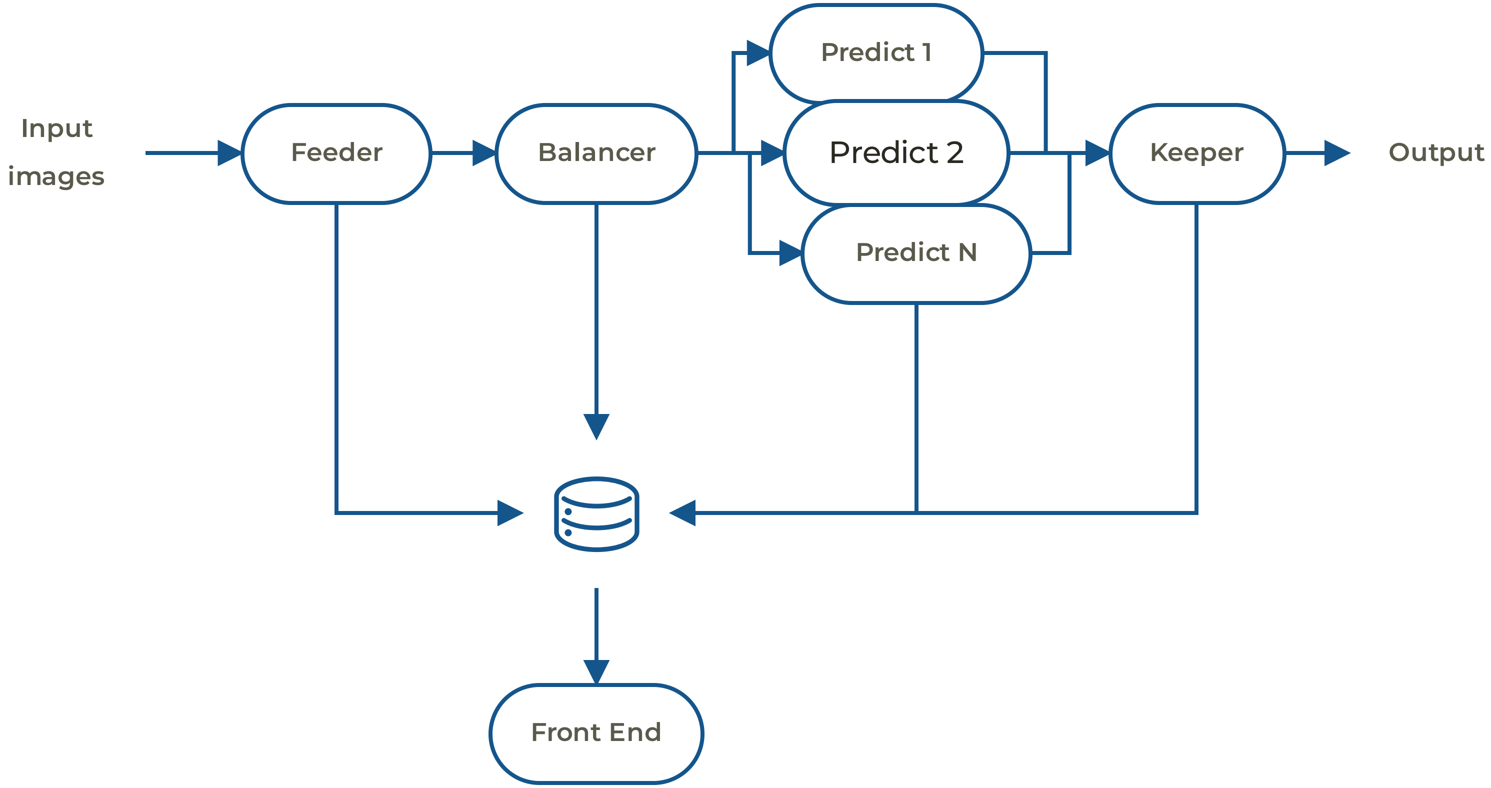
2/ Web based front end

Web based front end for labeling, monitoring, and model validation.

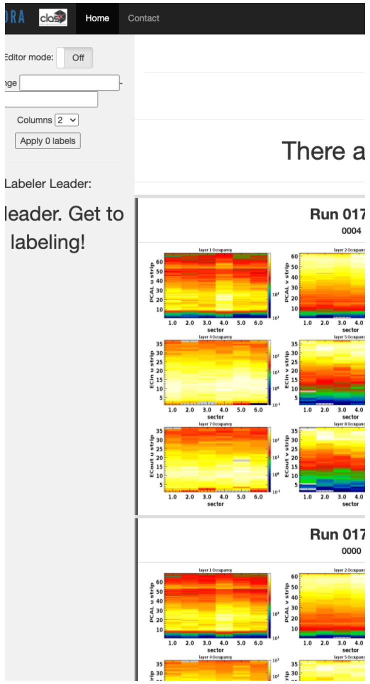
EPSCI develops and maintains Hydra in all Halls in collaboration with Users

HYDRA: Back End

- Hydra system
  - Written in python
  - Images move through Hydra via 0MQ
  - Can utilize multiple predicts if necessary
- MySQL database
  - Supplies information needed for UI
  - Stores all relevant information for every image it sees

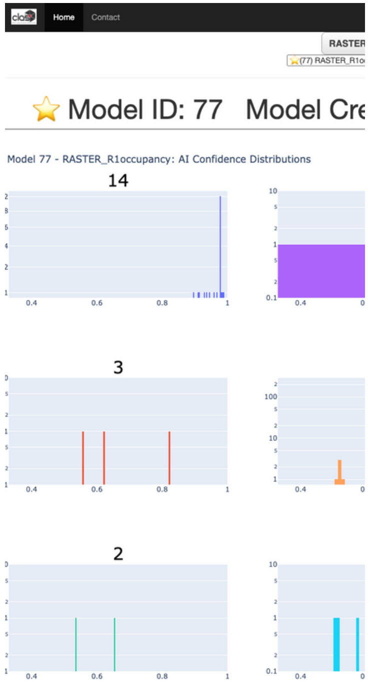


HYDRA: Front End



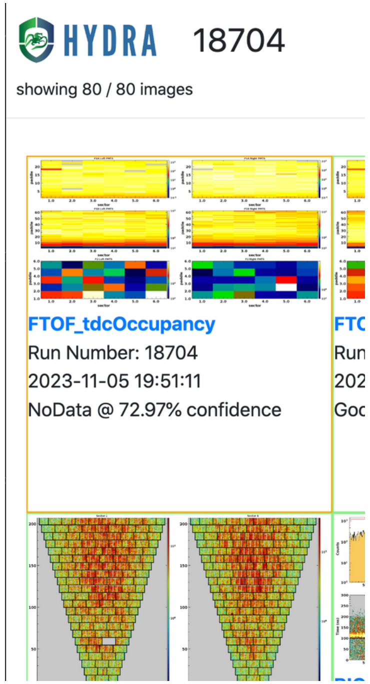
Data Labeler

Efficiently label hundreds (thousands) of images



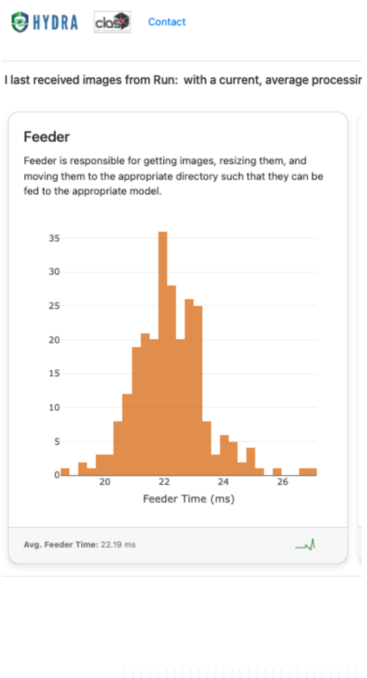
Library

Contains enhanced confusion matrix, thresholds, active model designations



Run

See predictions in real time

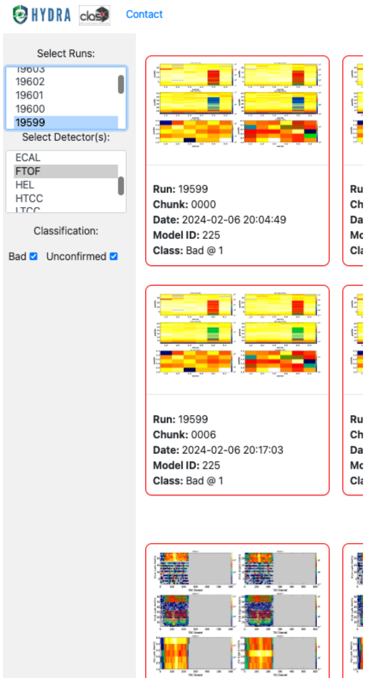


Status

Monitor heartbeats for back end processes and image processing time

Grafana

Dashboard displays all predictions over time

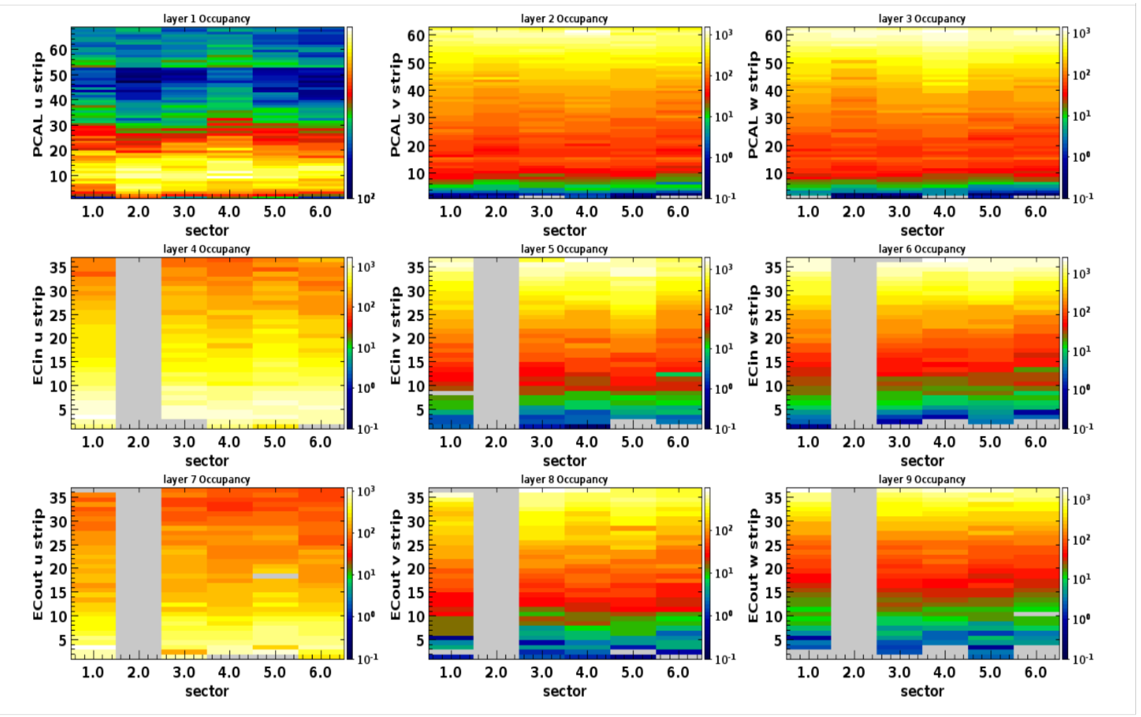


Log

Display concerning plots sorted by detector from previous day

Can the model tell us what about the image is bad?

Sometimes! With Gradient-Weighted Class Activation Maps



Original image

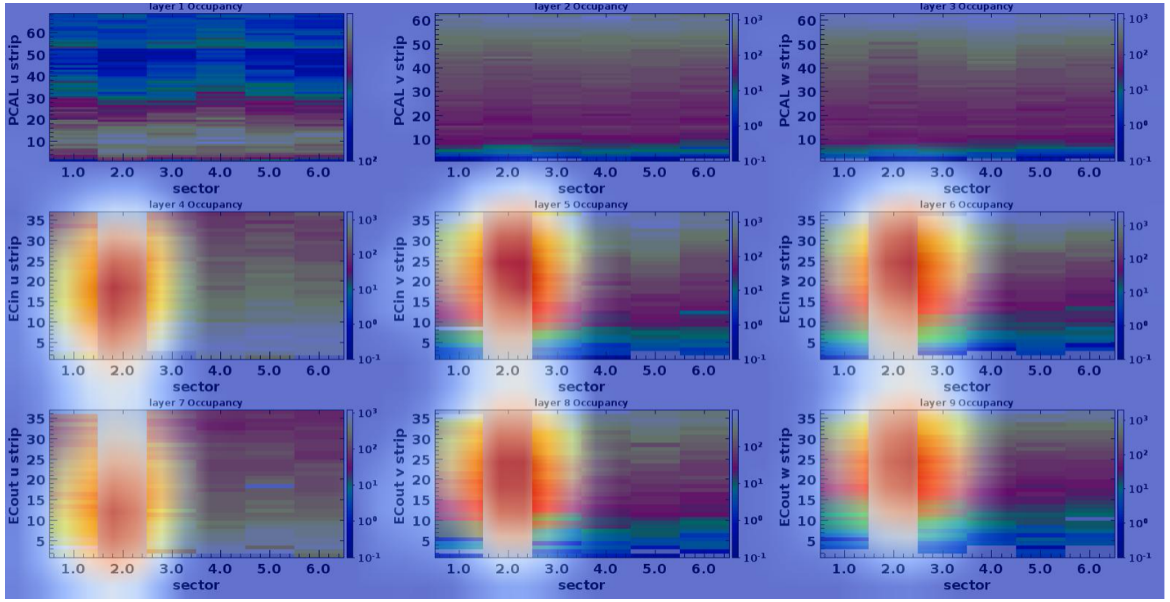
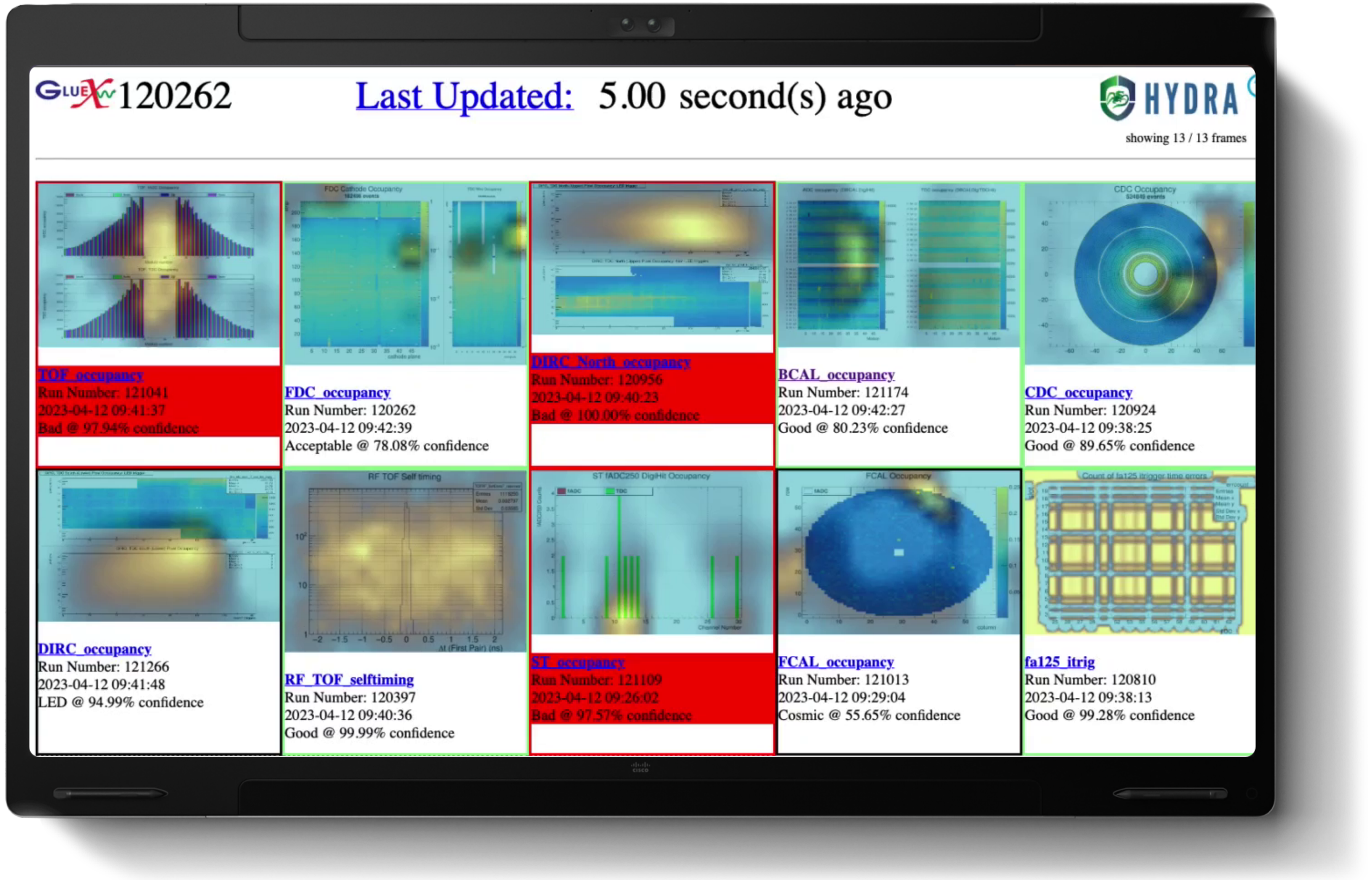


Image + GradCAM heatmap

The generated heat maps may not always coincide with what we expect.

Hydra Run + GradCAM

What you might see in your counting house :)



Dynamic Ordering

'Bad' images are moved to the top.

GradCAM

Where is Hydra looking when it makes its prediction?

Optional overlay*

Transferable skills with **HYDRA**

sometimes it's nice to take a break from ROOT histograms ;)

- System written entirely in python
  Tensorflow, pandas, SciKit Learn, etc
- Web based front end
  Javascript, HTML, CSS
- MySQL database used for back end
  PHP, SQL

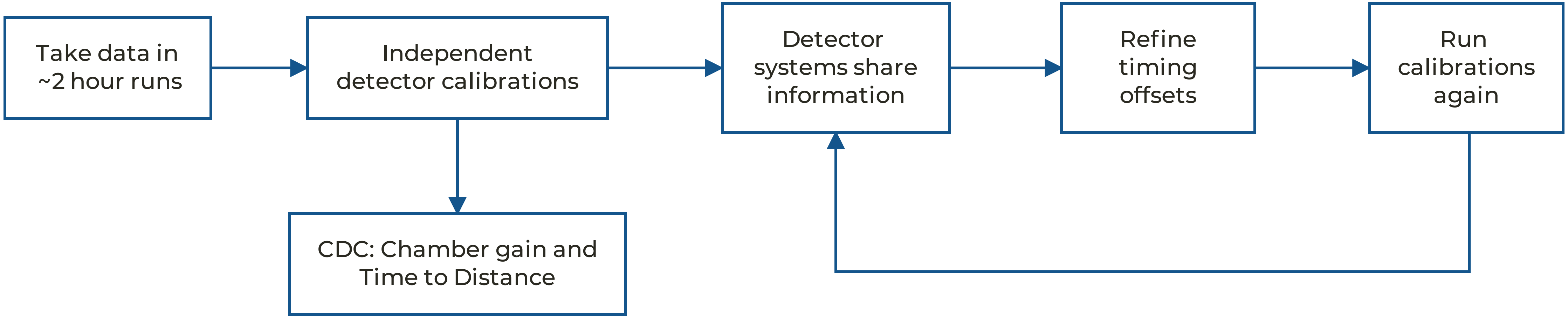
There is Hydra related service work in Hall B for those interested.

Experiment operations

1. **Design and Planning** — Identify key questions and objectives for experiment
2. **Setup and Calibration** — Installation, configuration, and alignment of detector systems
3. **Data Collection** — Data acquisition, quality monitoring, etc
4. **Offline Tasks** — Usually multiple rounds of calibration before proceeding with analysis
5. **Results**

Offline calibrations

Iterative and *time consuming*



Time scale for calibrations is on the order of months, increasing the time between data taking and publication.

GlueX Central Drift Chamber

Used to detect and track charged particles with momenta > 0.25 GeV/c

- ## Detector specs

  1.5 m long x 1.2 m diameter cylindrical, straw tube chamber

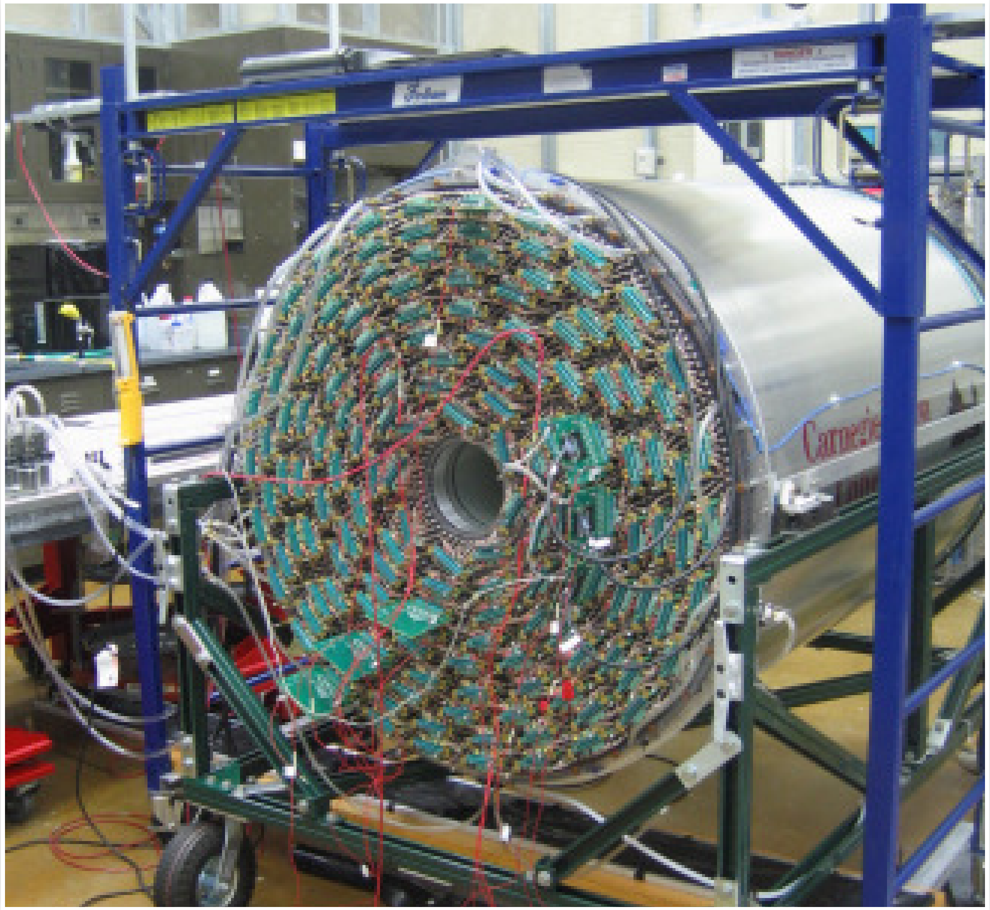
  3522 anode wires **traditionally held at 2125 V**

  50:50 Ar/CO_2 gas mixture

- ## Two main calibrations:

  Chamber gain and drift time to drift distance

  Affects PID selection in analysis via dE/dx

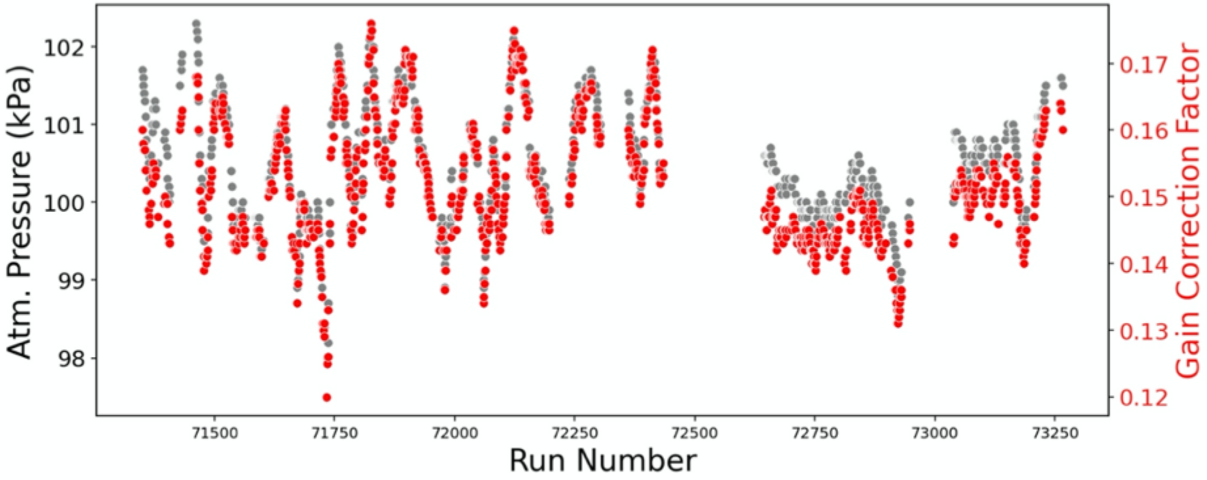


Standard operation

HV is set to 2125 V

Gain fluctuates *mostly* with atmospheric pressure

Can we train a model to predict the gain correction factor?



The gas gain fluctuates during data taking.

Challenges

There's a lot of them!

Offline vs Online Calibrations

1. Safety constraints
2. Control policies
3. Trustworthiness
4. There's always a bug we didn't account for

User Interface/ Experience

1. Interpretable UI for shift takers
2. Easy control ON/OFF button
3. Physics based evaluation metric

Data Science

1. Quick training and inference time
2. Readily available input features
3. Robustness to out-of-domain inferences
4. Uncertainty Quantification

What features should we include?

somewhat determined by calculating Shapley values

- The Kitchen Sink

  Every possible input feature obtained from reconstruction, EPICS, other detectors, etc

- Input features that are readily available

  fetched from EPICS archive

- Input features that are closely related to the CDC

  In hindsight, probably the most obvious approach

Interpretability

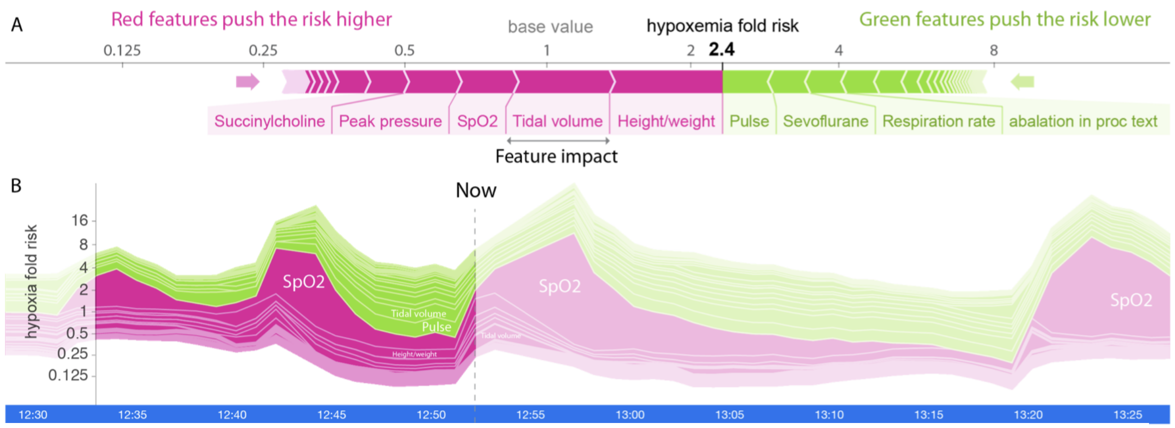
Shapley Values

Lloyd Shapley, Nobel Prize in Economics 2012

How does an individual input feature contribute to a prediction?

Based on 'fairness' properties from Game Theory

Model agnostic, but can be slow to evaluate



A) EXPLAINED RISK OF HYPOXEMIA IN THE NEXT 5 MINUTES DURING A SURGICAL PROCEDURE. B) EXPLAINED RISK EVOLVING OVER TIME.

Model Performance

Metric: Mean Absolute Percent Error

Used data from 2018 and 2020 run periods

	# Features	MAPE	MAX PE
Linear regression	11	1.3	19.1
MLP- 7 layers	122	1.8	11.4
GPR	5	1.5	9.1
RF	82	1.7	18.5
XGBoost -- corr > 0.2	82	1.44	11.8

Gaussian Processes

Supervised learning method used to solve regression and probabilistic classification problems

- ## Suited for small datasets

  ~430 training runs, 106 testing

  Existing calibrations used as target values

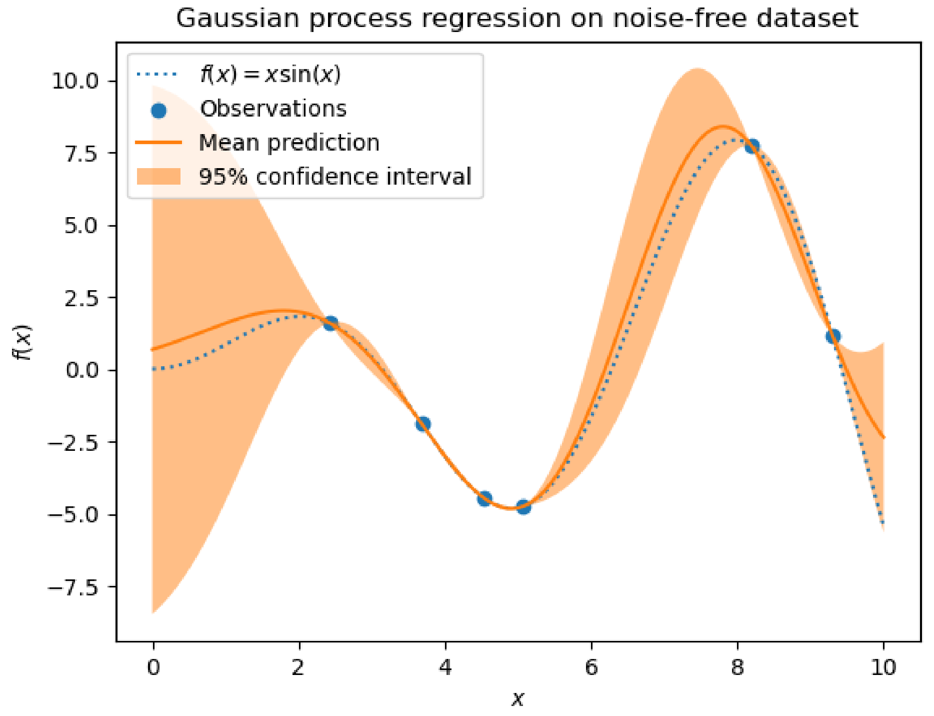
- ## Provide uncertainty quantification

  GPs give mean and standard deviation of the output when predicting

- ## Fast training and inference

  Inference is obtained in ~3 ms

Find a nice tutorial on GPs here

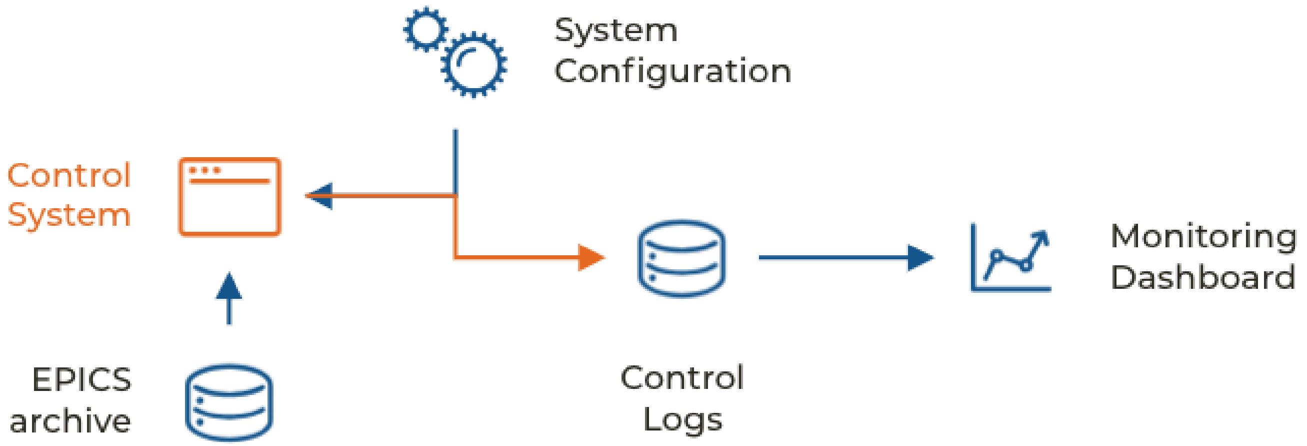


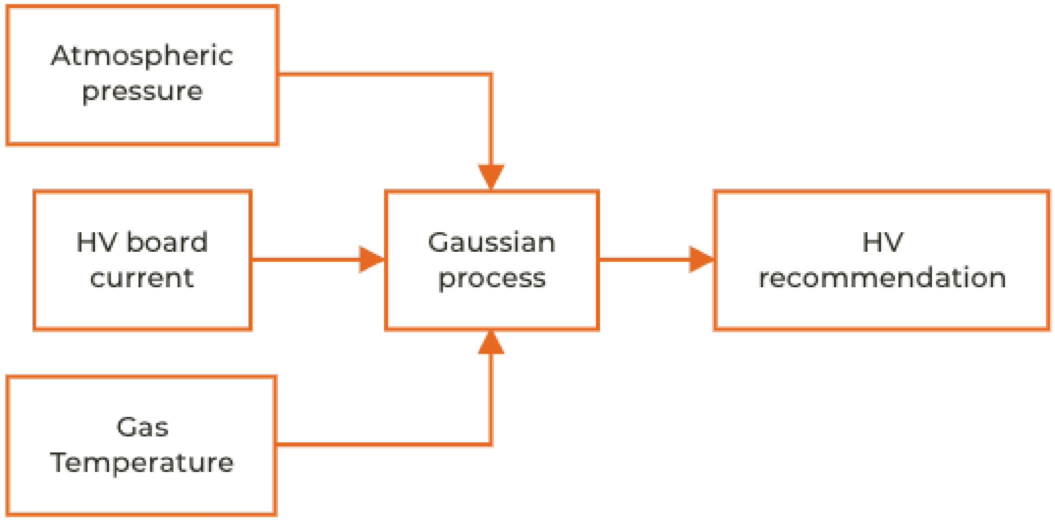
Online Calibration and Control

ML system to calibrate and control the GlueX Central Drift Chamber

Stabilize the response of the detector during the experiment

Successful collaboration of physicists and data scientists!





Control Policies

As they are now, not an exhaustive list

- ## Defined range of allowed HV settings

  Determined by detector expert

- ## High Uncertainty

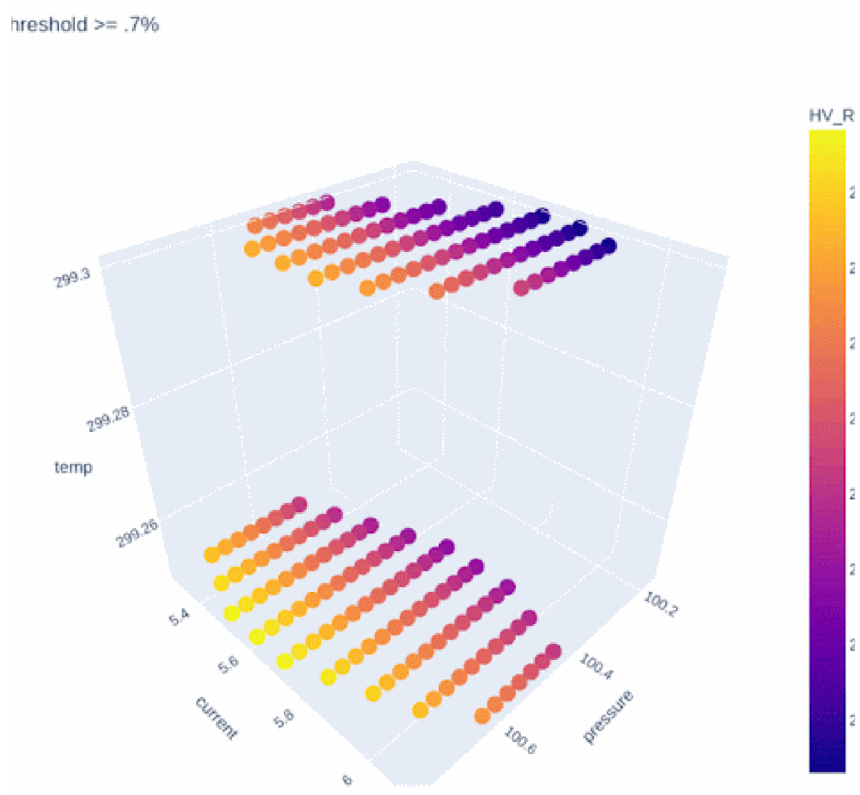
  Determine closest point to region of certainty, use that HV setting

  OR

  Revert to 2125 V and take more data

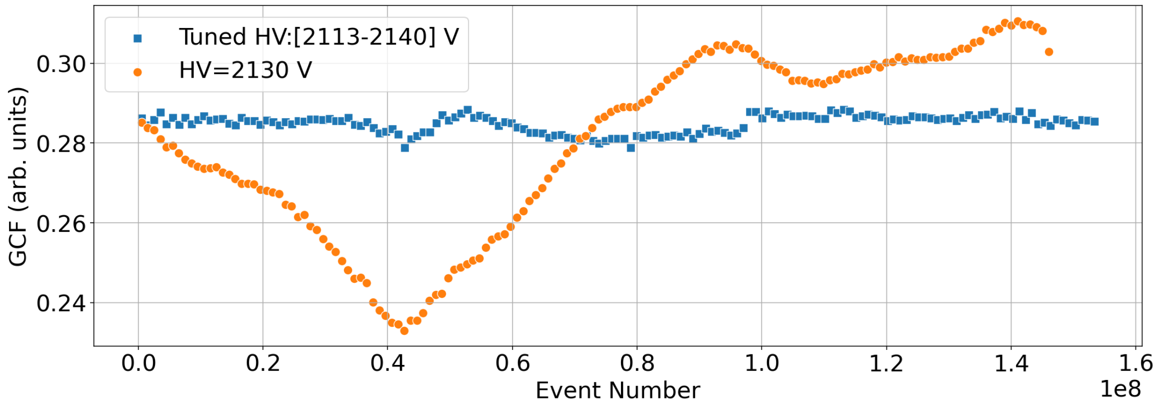
- ## "Trusting humans"

  If roboCDC detects a HV setting outside of our allowed range, we do nothing and assume there is something else going on (e.g., high current tests)



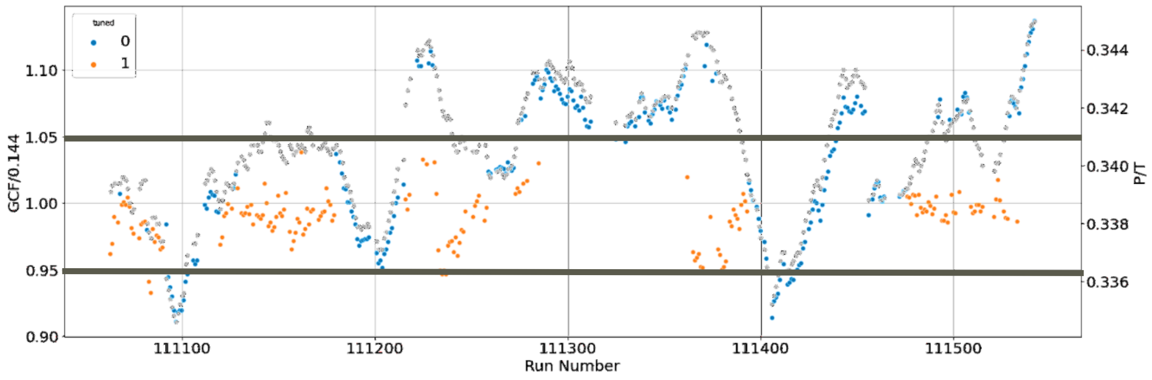
Results and Current Status

The CDC HV is set run-by-run based on the GP predicted gain correction factor.



Initial Cosmics test, 2021

Orange points indicate the gain correction factor using a **fixed** HV. **Blue** points indicate the gain correction factor while using the **tuned** HV setting.

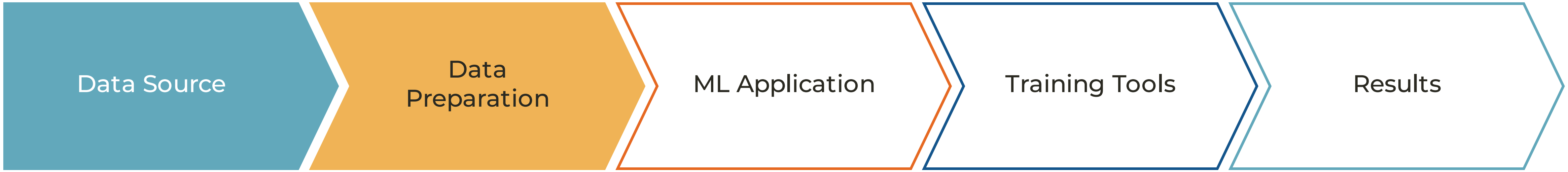


Primex-Eta Run Period, 2022-2023

Orange indicates tuned HV depending on environmental conditions. **Blue** indicates fixed HV. The gray points are the ratio of atmospheric pressure and gas temperature. The horizontal lines indicate our error tolerance as determined by the detector expert.

Questions

Data Science Pipeline



Real or synthetic
Quality
Dimensionality
Format
Size
Time-to-acquire

Data cleaning
Restructuring
Correlations
Visualization

Classification
Regression
Clustering
Feature Extraction

Cross validation
Hyperparameter optimization

Predictions
Confidence level
Explainability
Uncertainty Quantification