



# Caterpillar Big Data Infrastructure

Big Data, Data Analytics,  
and Machine Learning





Caterpillar is the world's leading manufacturer of construction and mining equipment, industrial diesel engines and gas turbines, and diesel-electric locomotives.



# CATERPILLAR® Solutions

## Autonomy and Operator Assistance



Autonomous Haul Trucks



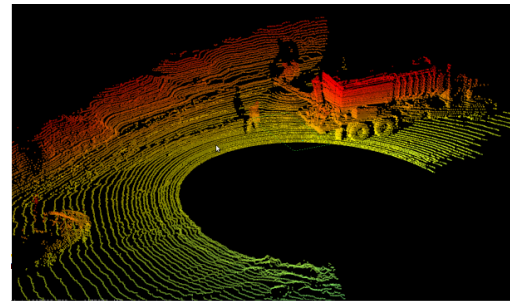
Non-Line of Sight  
Remote Semi-Autonomy



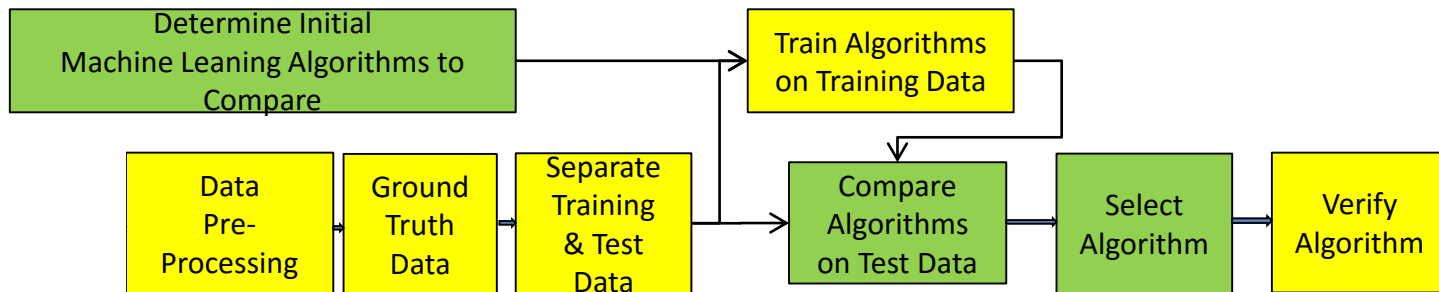
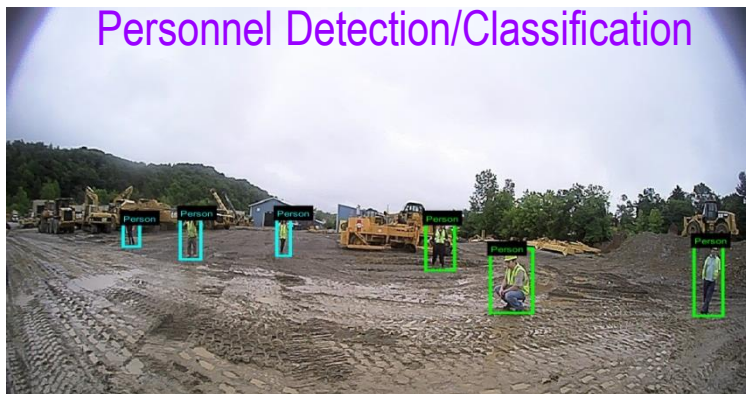
Operator Assistance



Machine Learning on Advanced Sensor Data

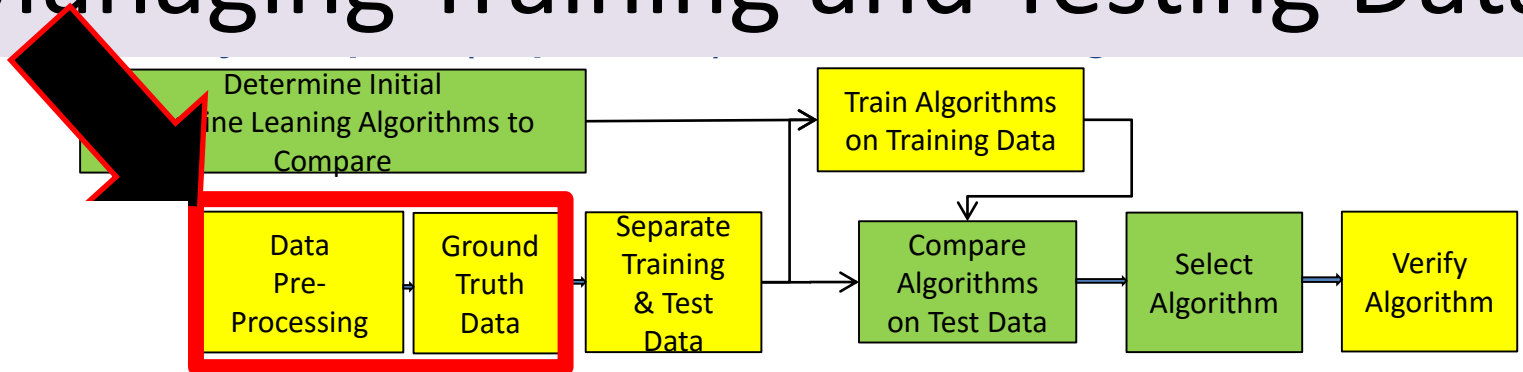


# Why Do We Need a Big Data Infrastructure?



**Example: Machine Learning Flow**

We Were Spending  
Too Much Time  
On Ground Truth  
and Managing Training and Testing Data

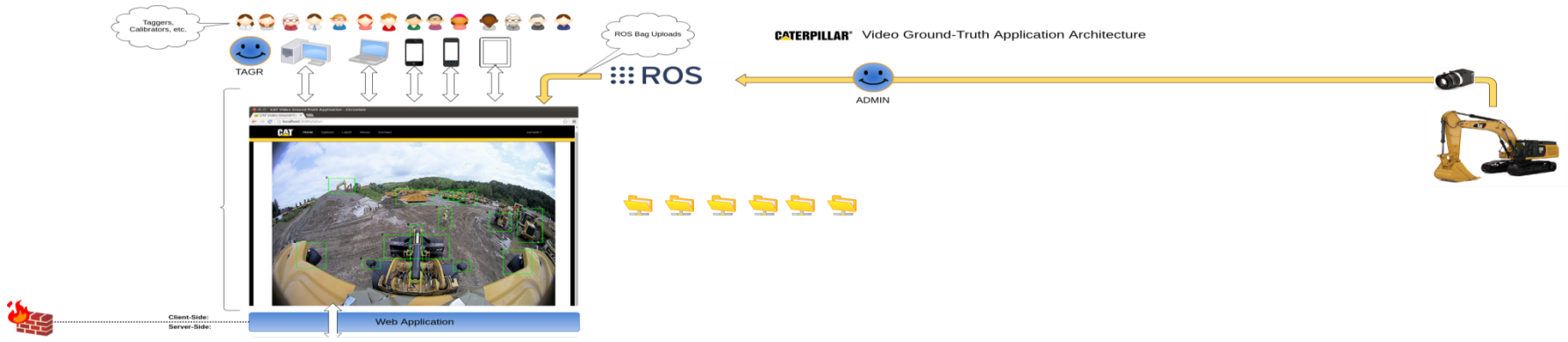


**Example: Machine Learning Flow**

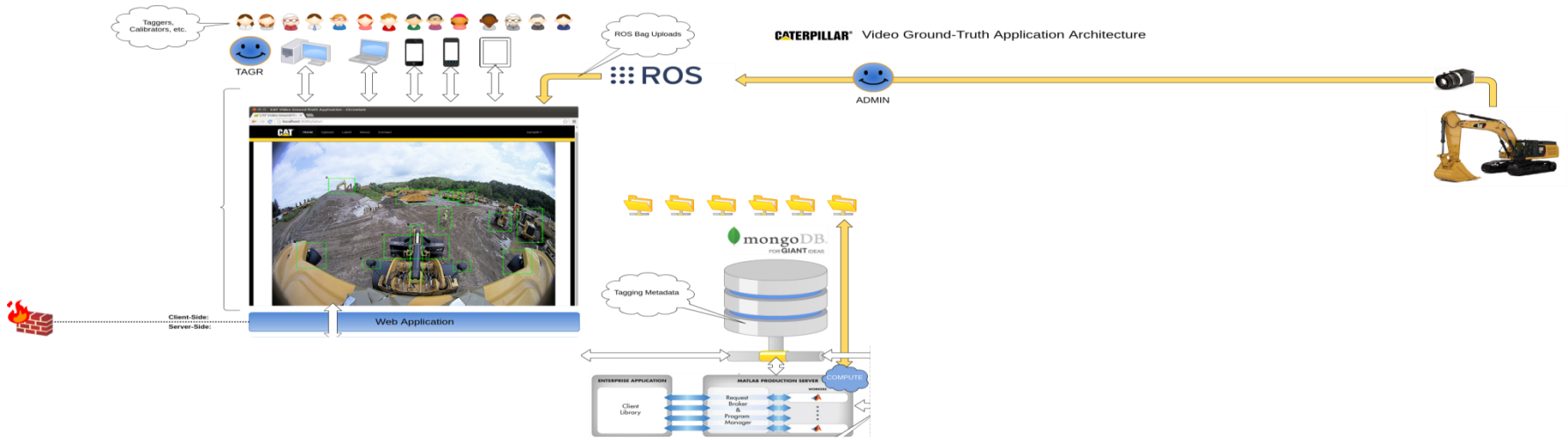
# CatBigDat – Field Data Collection



# CatBigDat – Web Based Ground Truth Tagging

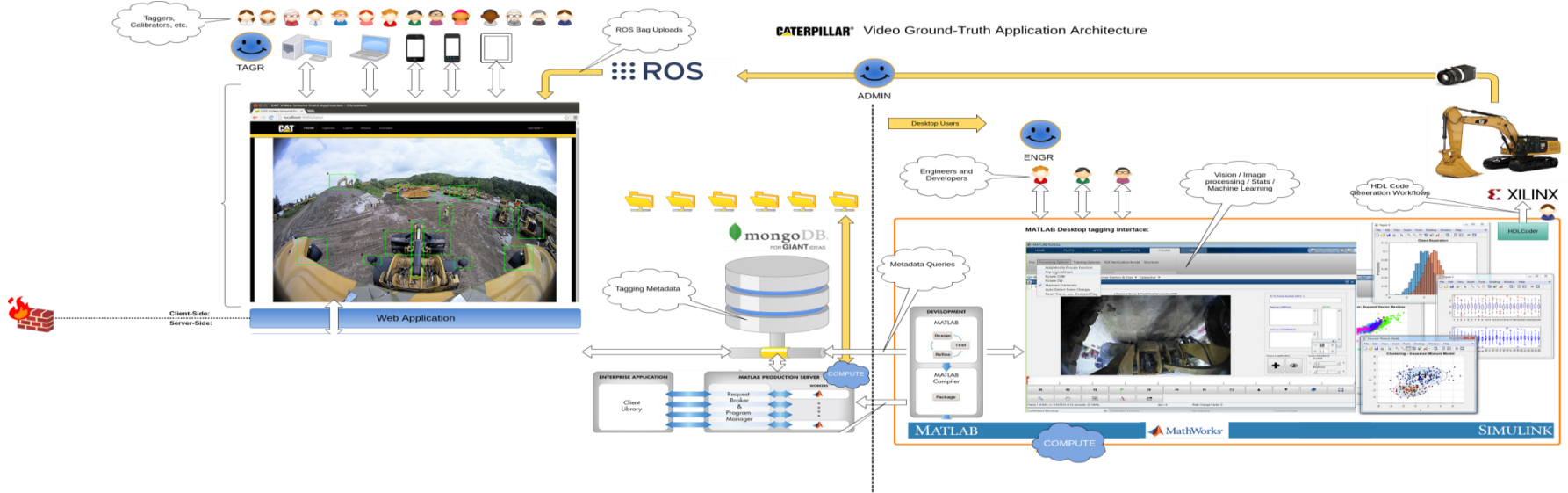


# CatBigDat – Ground Truth Metadata Database

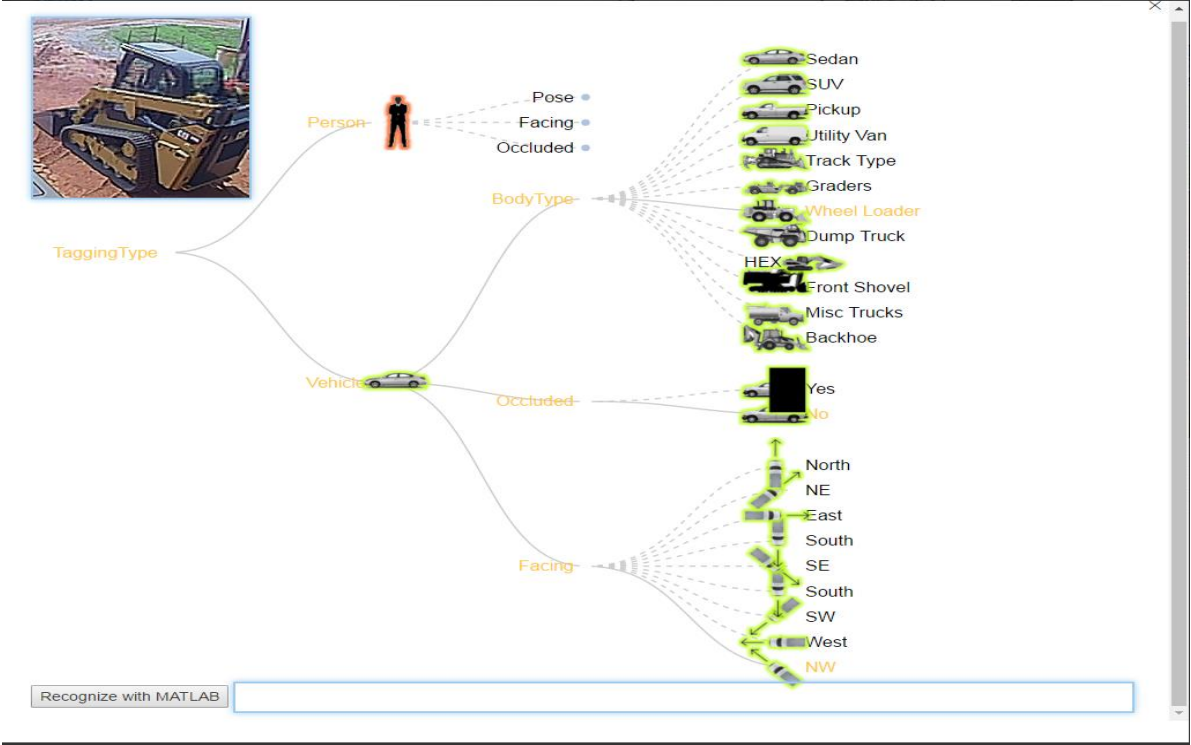




# CatBigDat – Engineering Interface Leverages Power of MATLAB



# Completely Flexible and Modifiable Ground Truth Label Hierarchy - Vehicle



# Completely Flexible and Modifiable Ground Truth Label Hierarchy - Personnel

x





Annotation Controls

Icon visibility:

Label visibility:

ROI type:

Box  Polyline  Segment



Restore pages? x

Chrome didn't shut down correctly.

Restore

Help make Google Chrome better by sending crash reports and usage statistics to Google

Saturation

Vibrance

Exposure

Hue

Sepia

Gamma

Noise

Clip

Sharpen

Blur

Independent control:  OFF

# General Additional Fields - Pick Lists

- Environmental Lighting
  - Sunny Day - Full day data, dawn to dusk on clear sunny day with mixed lighting (shadows and bright sunlight)
  - Cloudy Day - Full day data, dawn to dusk on cloudy day
  - Low Light
  - Night w/ Lights - Night data with vehicle lighting
  - Night w/ Lights and Incidental - Night data with vehicle and incidental lighting
- Background Environment (Construction Building, Construction Highway, Mine Surface, Commercial, Residential, Urban, Rural)
- Location (Indoor, Outdoor)
- Airborne obscurants (Dust, Fog, Smoke)
- Weather (Raining, Snowing)
- Ground Conditions (Mud/Dirt, Partial Snow, Majority Snow, On-Road, Off-Road, vegetation, gravel)
- Quality of Focus (Good, Poor, Lens Occlusion, Lens Damage)

# Example Queries w/ Example Results

- Standing, un-occluded people
- Crouching, un-occluded people
- Close range, occluded people
- Negative Data (e.g. Non-People)
- Hydraulic Excavator, Side View
- Hydraulic Excavator, Rear View
- Wheel Loader, Bucket in Air





```
Webix Player
MUTUAL-RESIZE
File Edit View Insert Tools Debug Window Help
...
c1=cat.annotation.find('QueryMAC1.json')
c1.show
c1.showChip
c2=cat.annotation.find('QueryMAC2.json')
c2.showChip
```

```
>>
c1 =

annotation with properties:

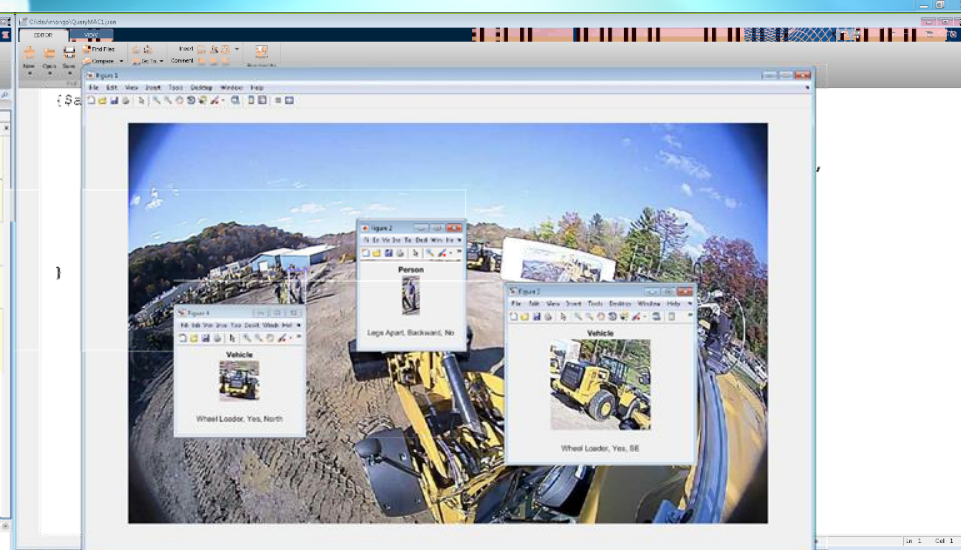
  id: 'a12e8cac-da0e-43ff-9862-5daf5fbc24c'
  image: [1x1 serializable]
  folder: '/srv/CAT/frames/2016-11-04-13-30-16_0_simulatedloading_frames/cont1108/'
  object: [1x1 serializable]
  description: []

>>
c2 =

1x2 annotation array with properties:

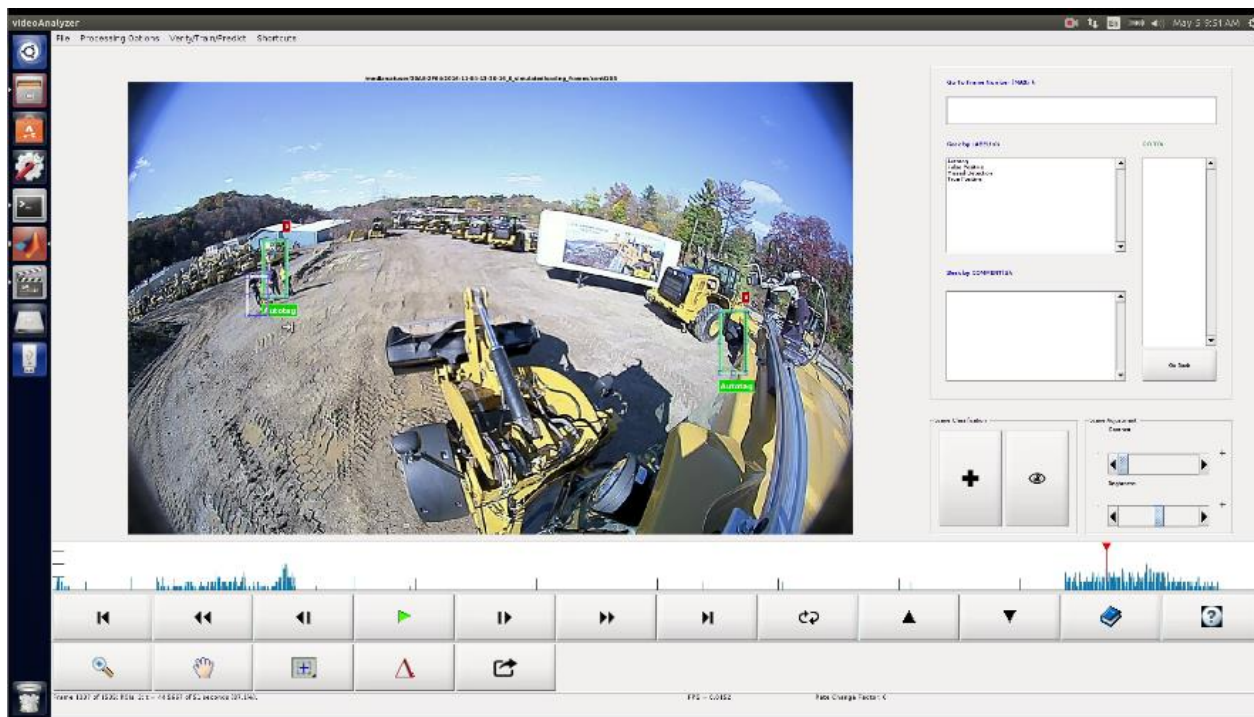
  id

fx>>
```



```
Webix Player
MUTUAL-RESIZE
File Edit View Insert Tools Debug Window Help
...
{$and:
  [
    ["object.hierarchy.children.name" : /vehicle/i],
    {$or: [{"object.hierarchy.children.children.name" : /north/i},
          {"object.hierarchy.children.children.name" : /se/i}],
    ["object.hierarchy.children.children.name" : /yes/i],
    {"object.hierarchy.children.children.name" : /Wheel Loader/i},
  ]
}
```

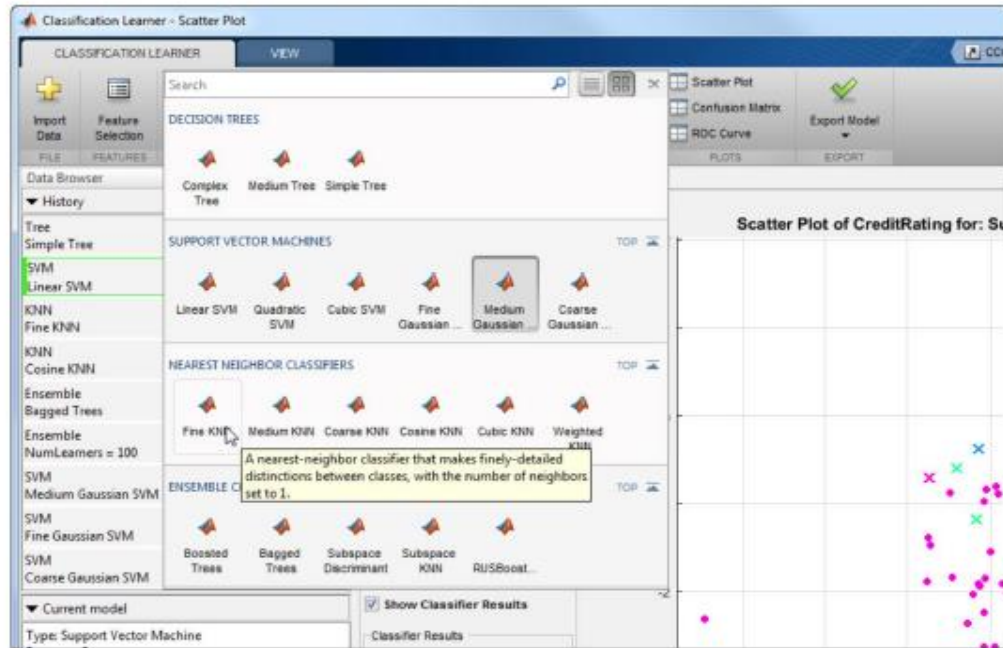
# Automatic Labeling of Data





# Tight integration with MATLAB Classification Learner App

- Simple queries into Caterpillar labeled data to import multi-class positive and negative data for training.
- Tight integration with MATLAB Machine Learning Backend (Classification Learner and Command Line)



# Integration with Auto-Coding Tools And 3<sup>rd</sup> Party Machine Learning

HDL Coder

Caffe

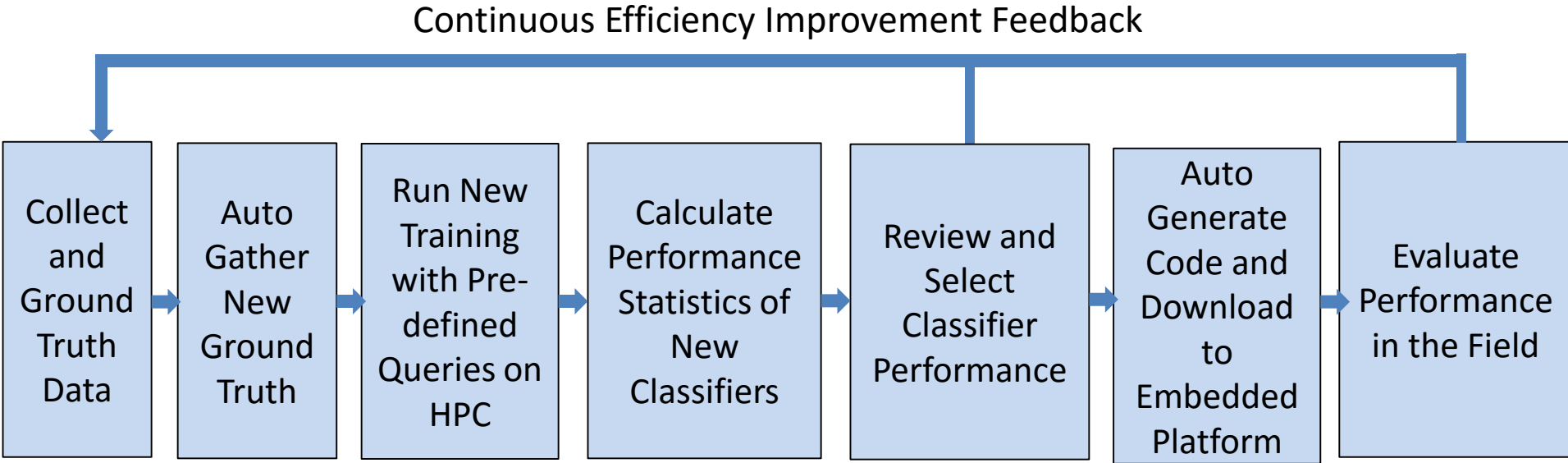
Deep learning framework

VIVADO.  
HLx Editions

- DESIGN
- SYSTEM
- WEBPACK

SDSoC<sup>™</sup>  
Environment

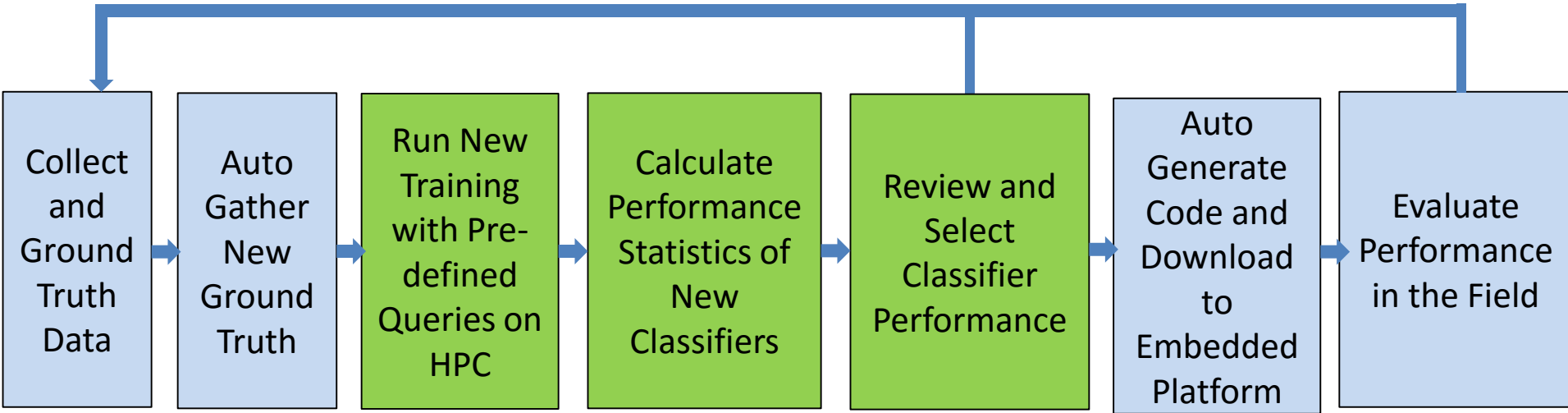
# Using MATLAB for Continuous Improvement in our Big Data, Data Analytics, and Machine/Deep Learning Infrastructure



Because it is MATLAB, development time is short

# Future Direction for the Infrastructure

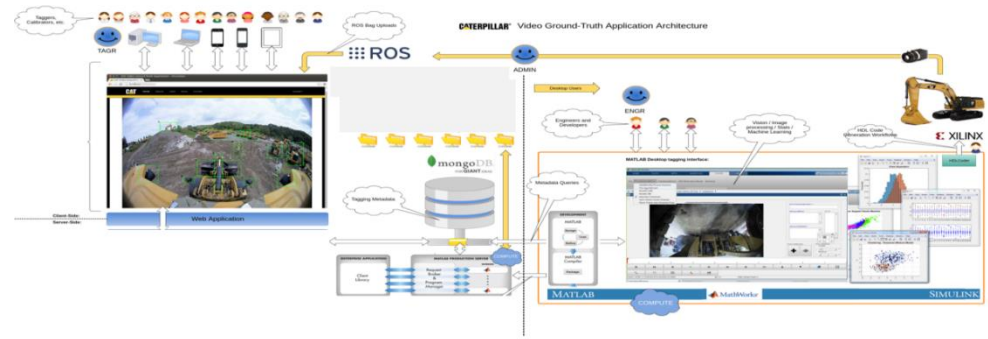
Continuous Efficiency Improvement Feedback



Make it Even Easier to Find Best Classifiers to Solve a Given Problem - More Science, Less Art

# Conclusions

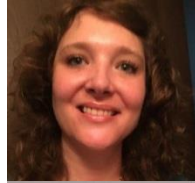
- Developed big data and machine/deep learning infrastructure
- Web based ground truth interface
- Automatic ground-truth -- limits need for human supervision, reducing development time
- Database for storing and querying meta-data
- Engineering interface with tight integration with MATLAB products for learning, visualization, verification
- Code generation - direct to embedded real-time platforms
- Scalable in number of users, amount of data, and compute power



Thank You!



Lisa Crosier



Amine El Helou



Gary Gunterman



Joe Forcash



Arvind Hosagrahara



Larry Mianzo



Steve Kuznicki



Dan Troniak



Brett Shoelson

