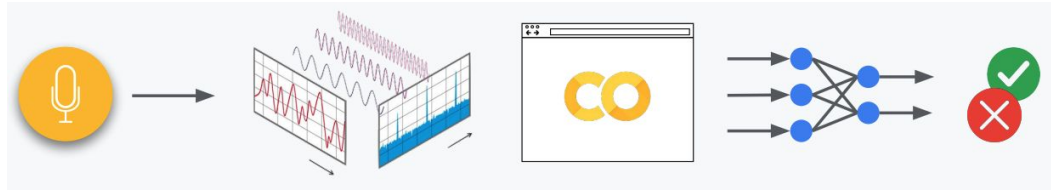


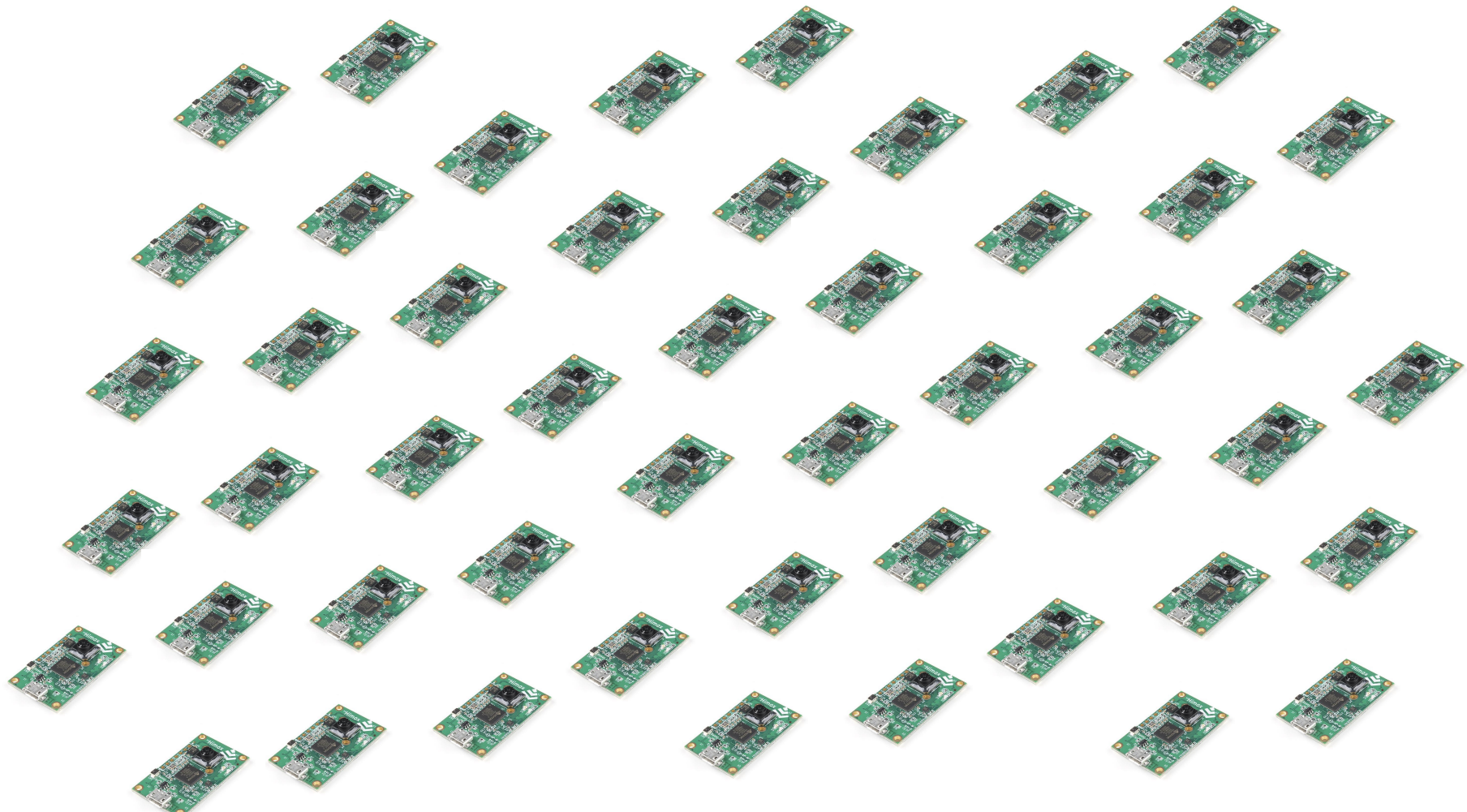
TinyMLOps: Scaling Deployments

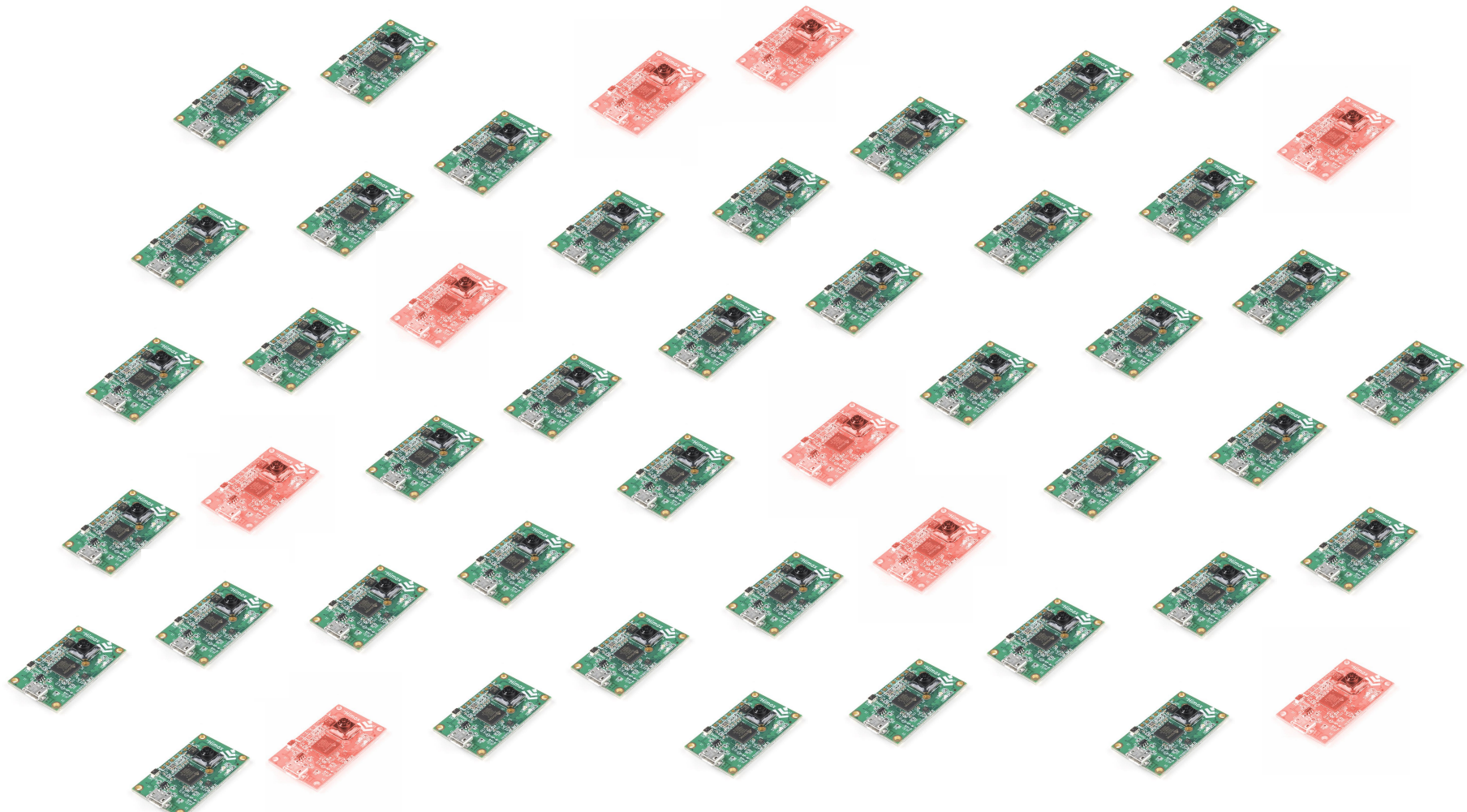


Colby Banbury
PhD Student
Edge Computing Lab, Harvard University







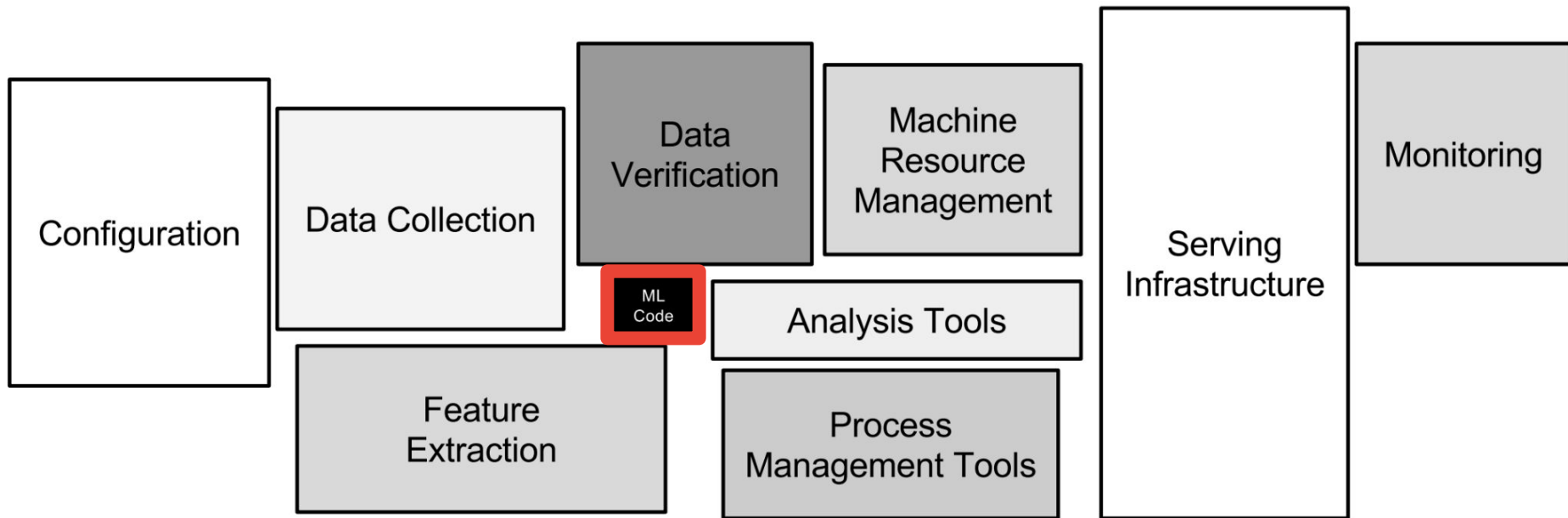


“Hidden Technical Debt in ML Systems”

ML
Code

Source: D. Scully, 2015

“Hidden Technical Debt in ML Systems”

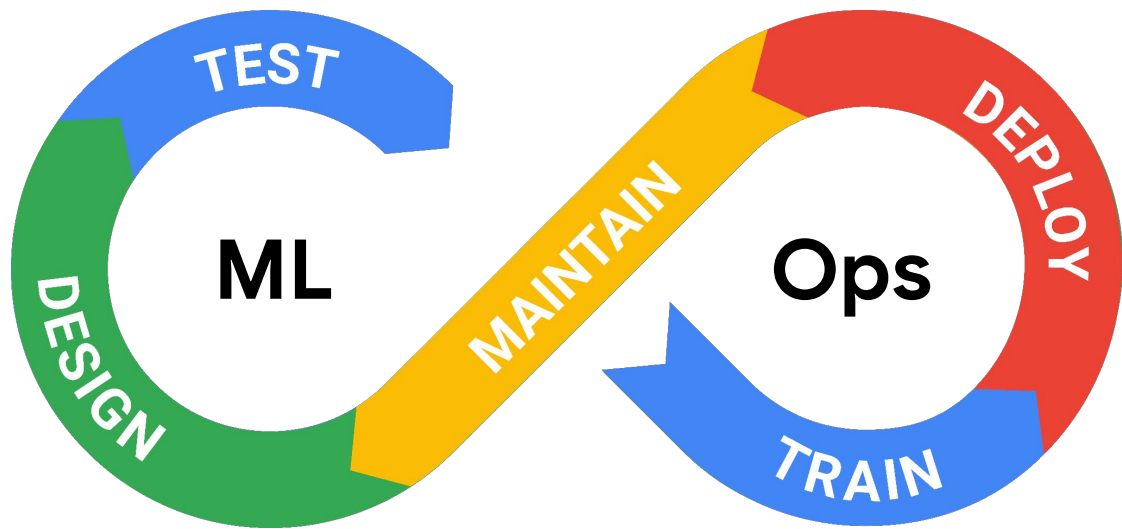


Source: D. Scully, 2015



Close The Loop





MLOps = ML Workflow + Automation

MLOps means...

5 Focus Areas

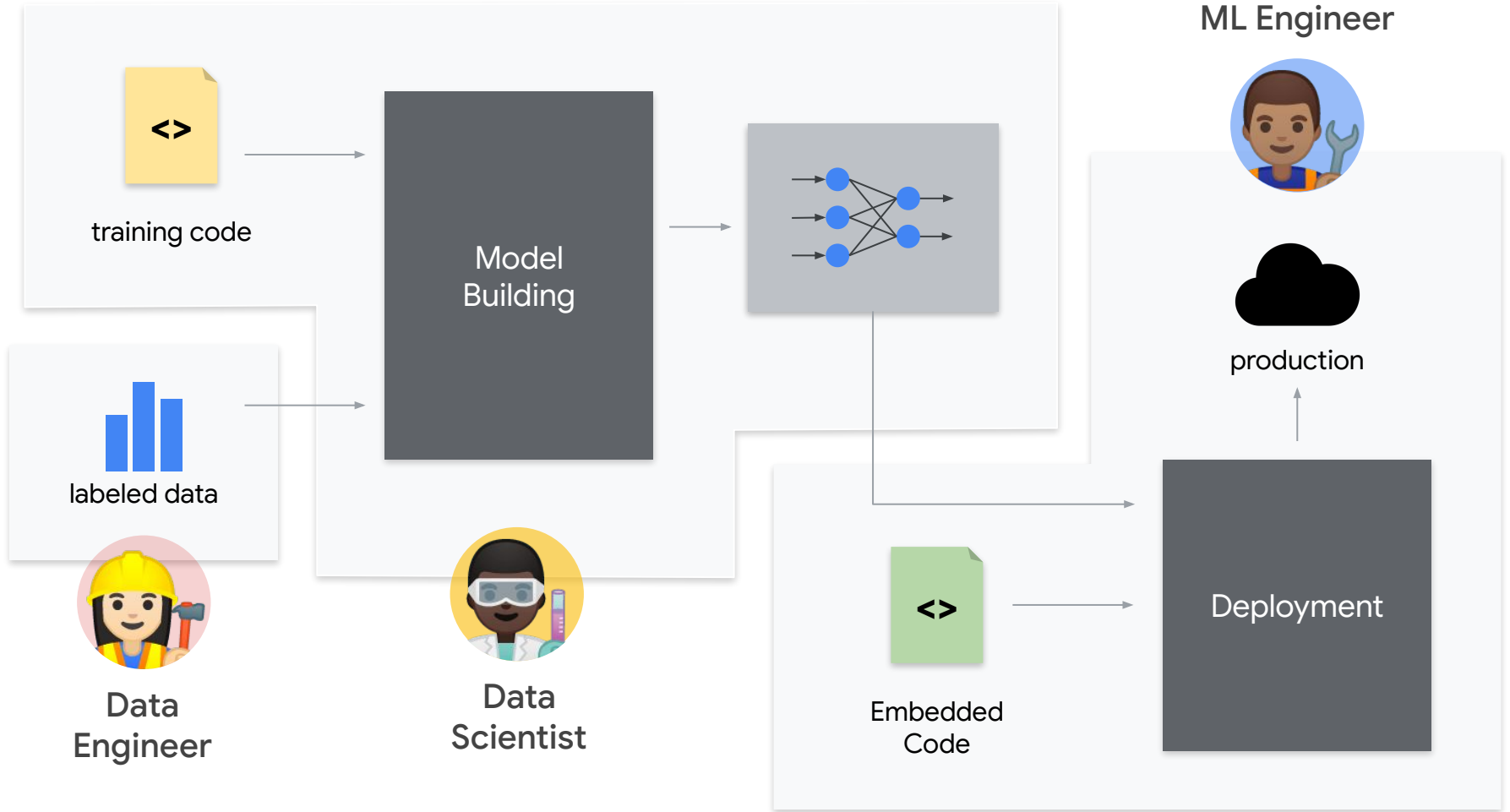
Running
end-to-end

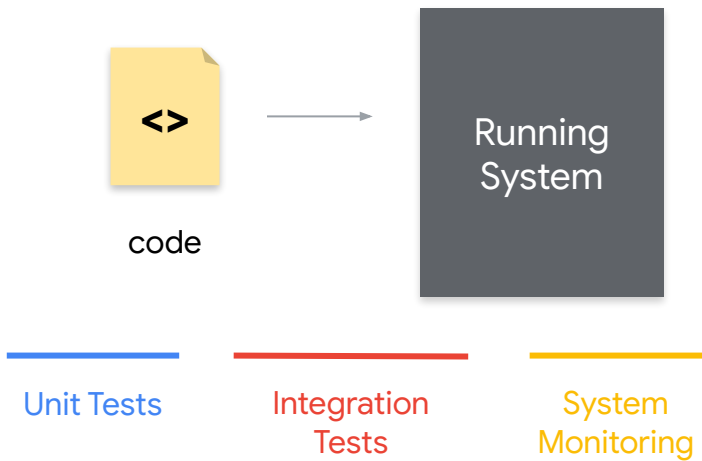
Managing
Complexity

Evaluating
Results

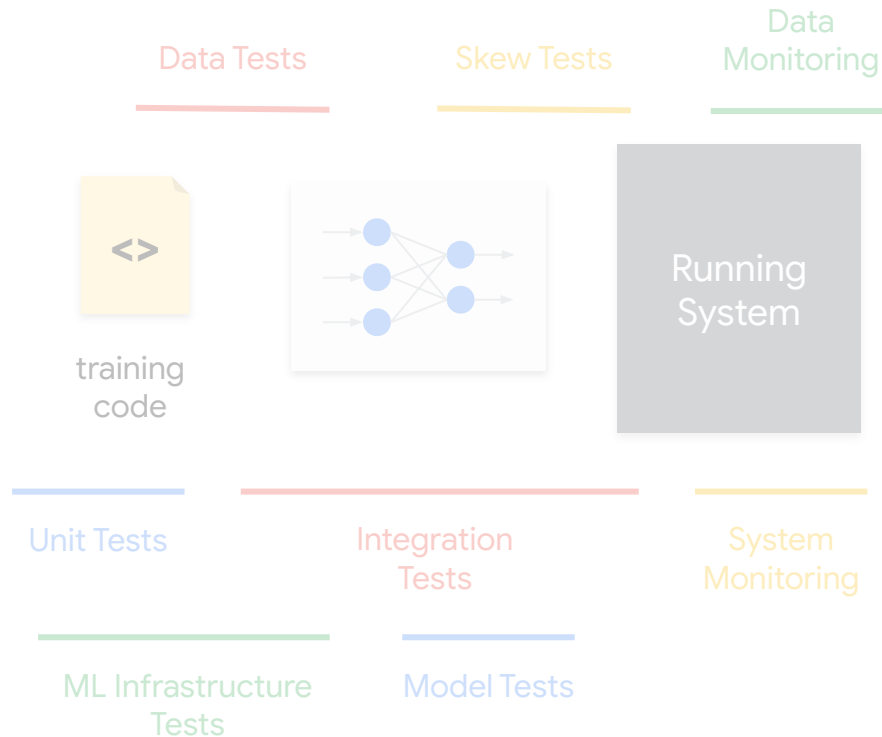
Improving
Models

Tracking
deployments

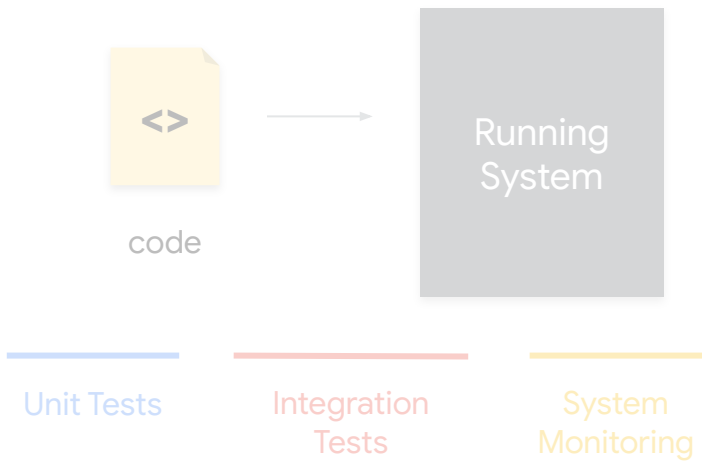




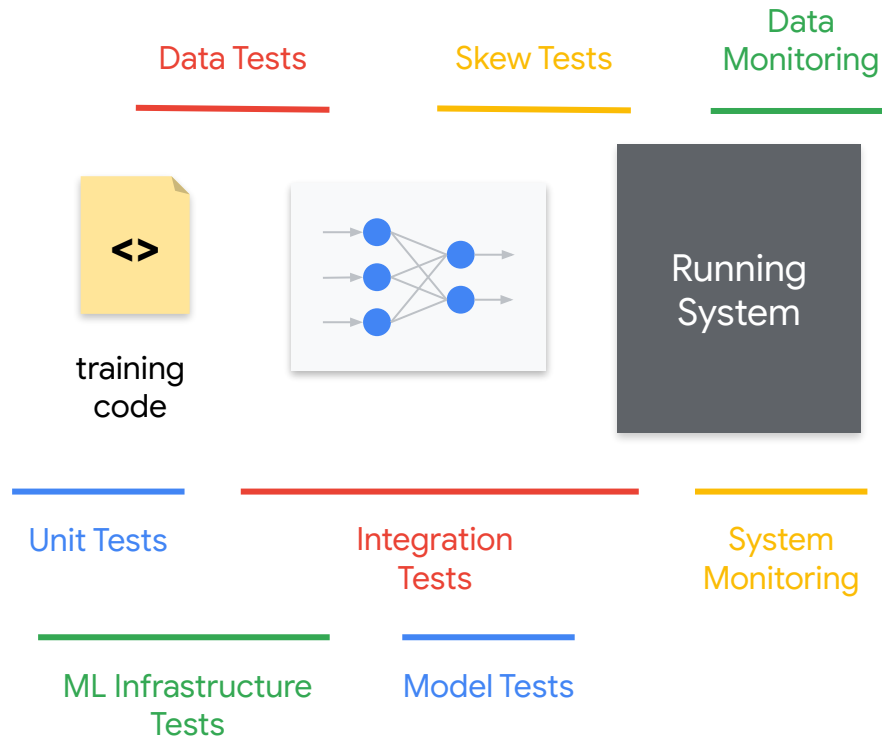
Traditional System Testing



ML-based System Testing



Traditional System Testing



ML-based System Testing

MLOps Tools

Custom
Scripts



Frameworks



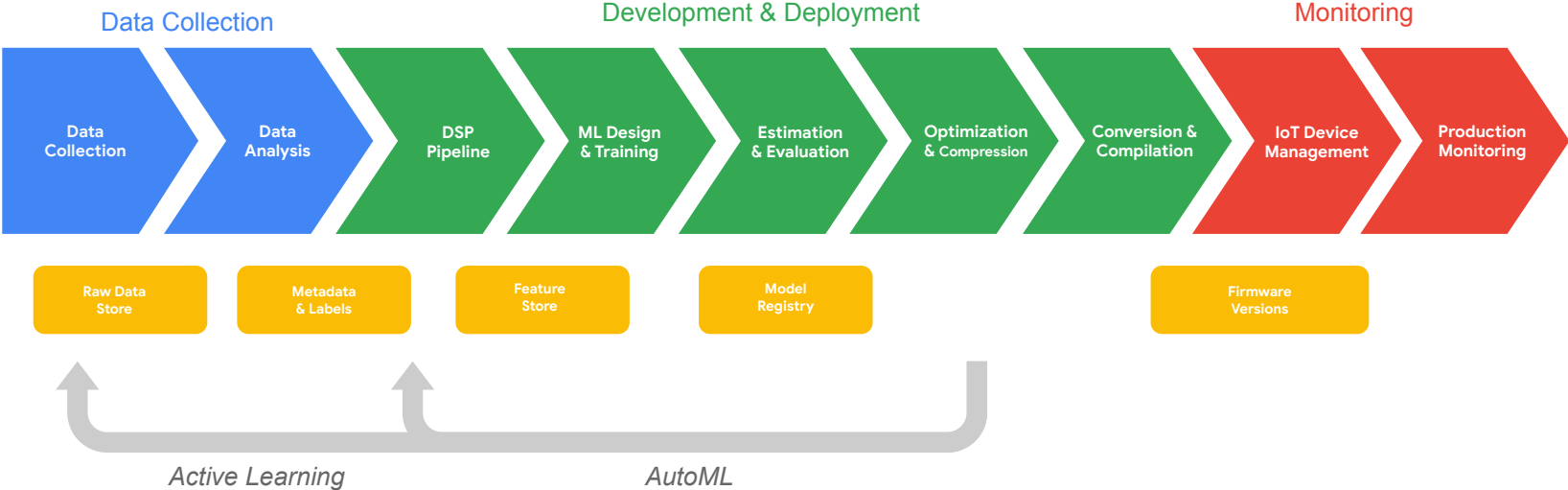
TensorFlow Extended
(TFX)

End-to-end
Platforms



Edge Impulse Workflow

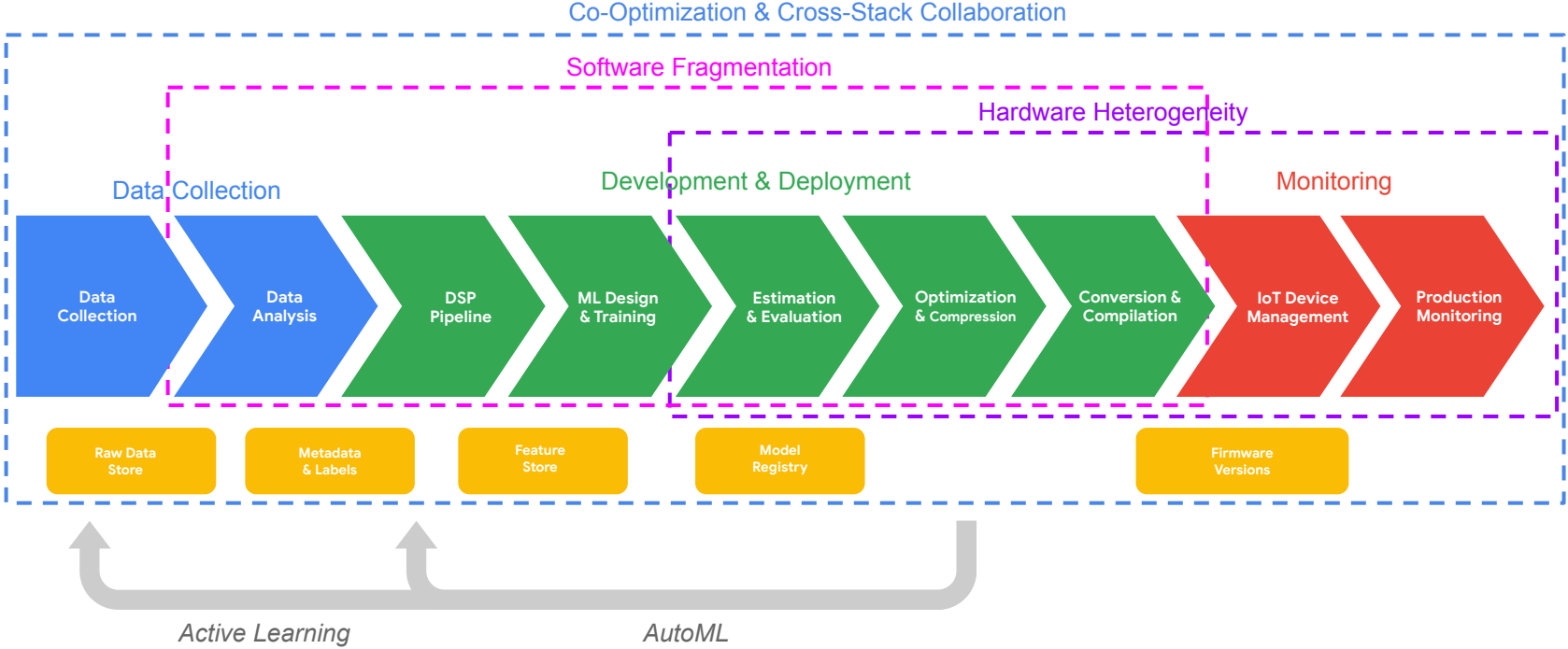
Edge Impulse
Workflow



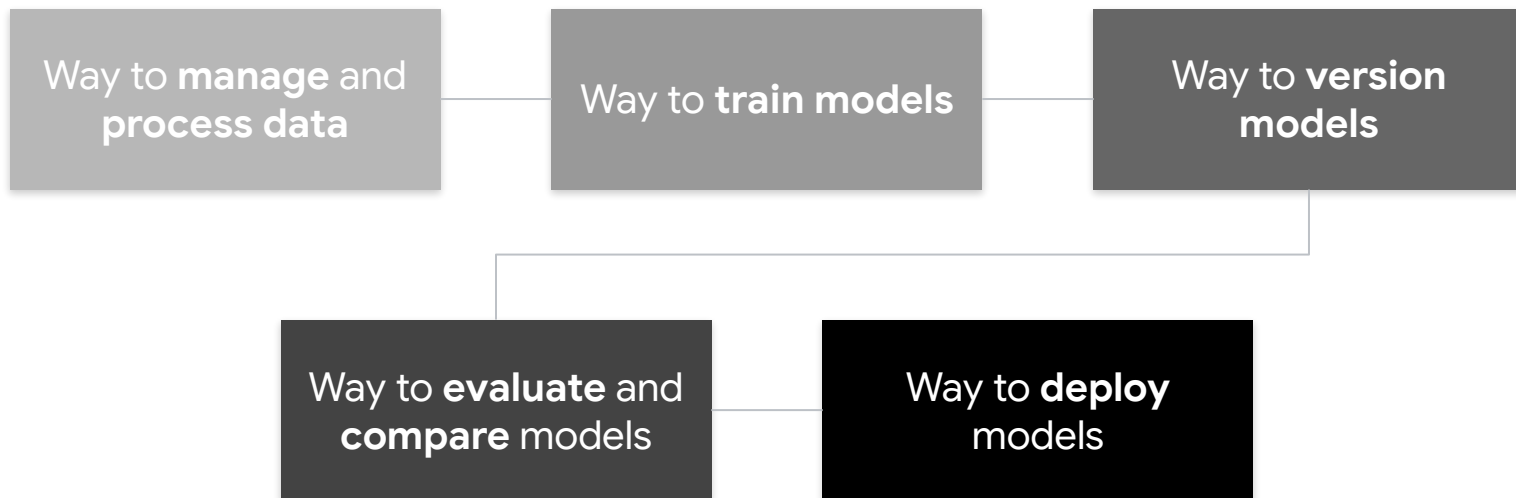
Edge Impulse Workflow

Challenges Addressed

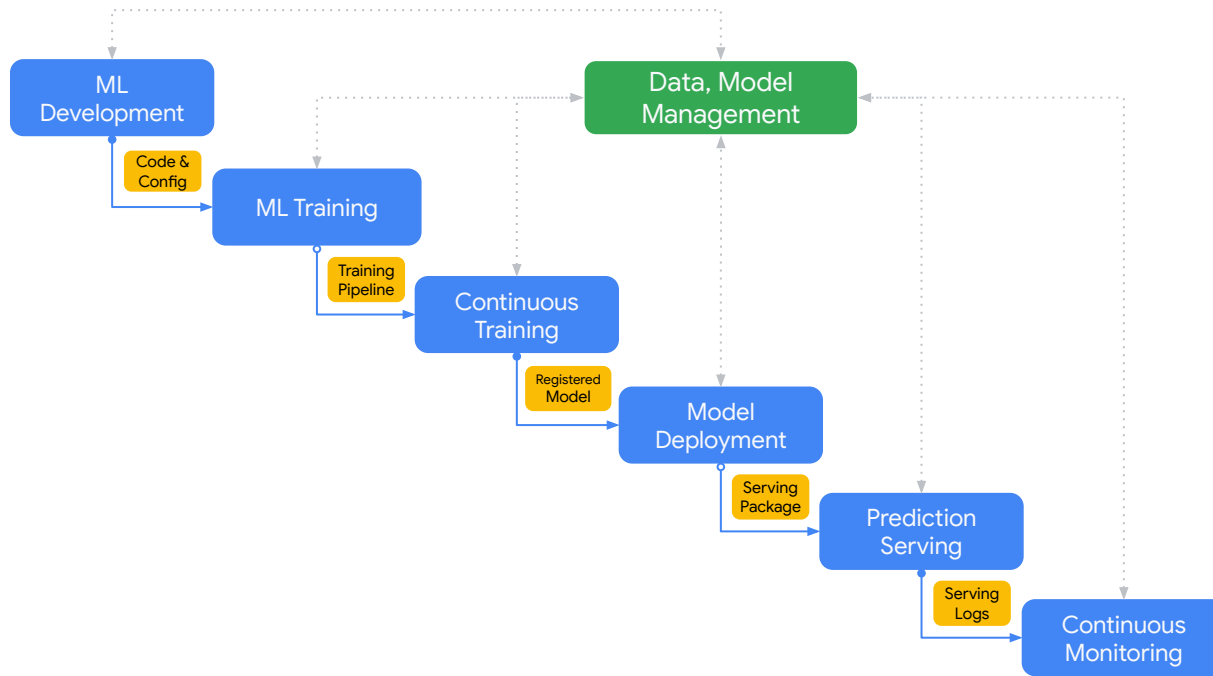
Edge Impulse Workflow



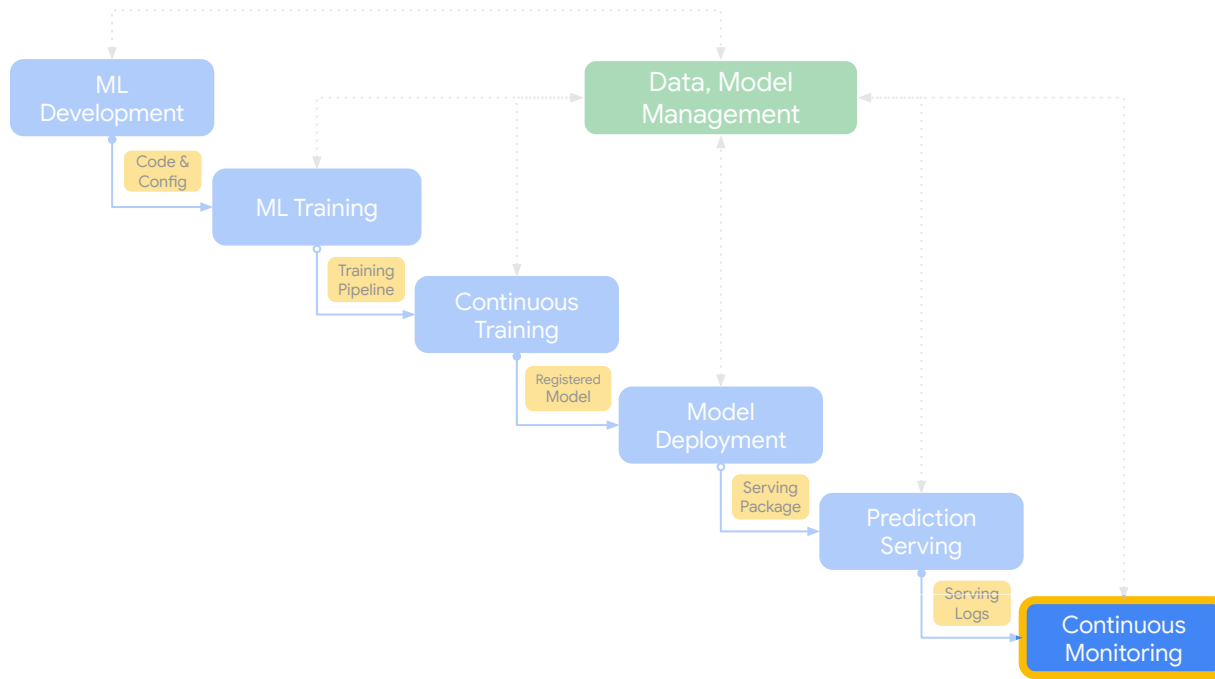
Basic features of MLOps tool



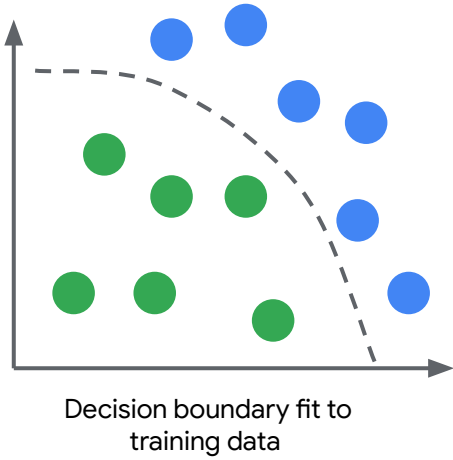
The MLOps Process



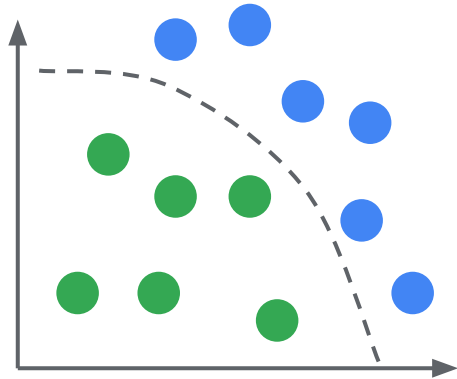
Continuous Monitoring Example



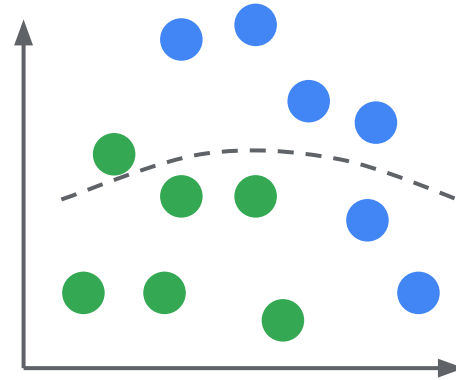
Concept Drift



Concept Drift

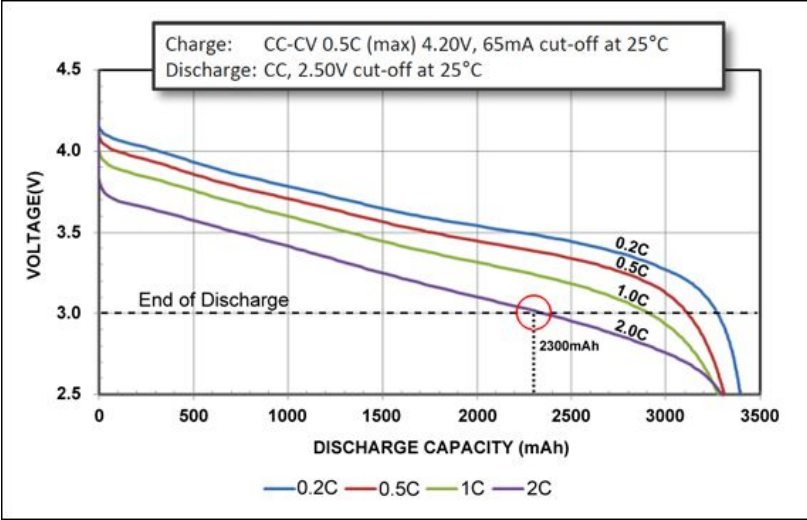
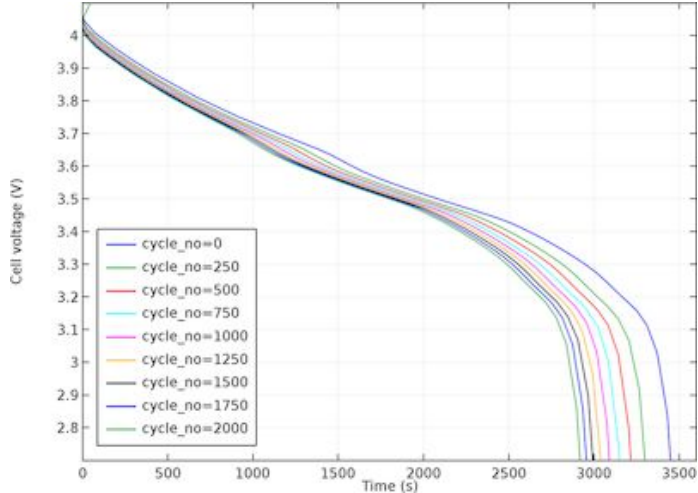


Decision boundary fit to training data

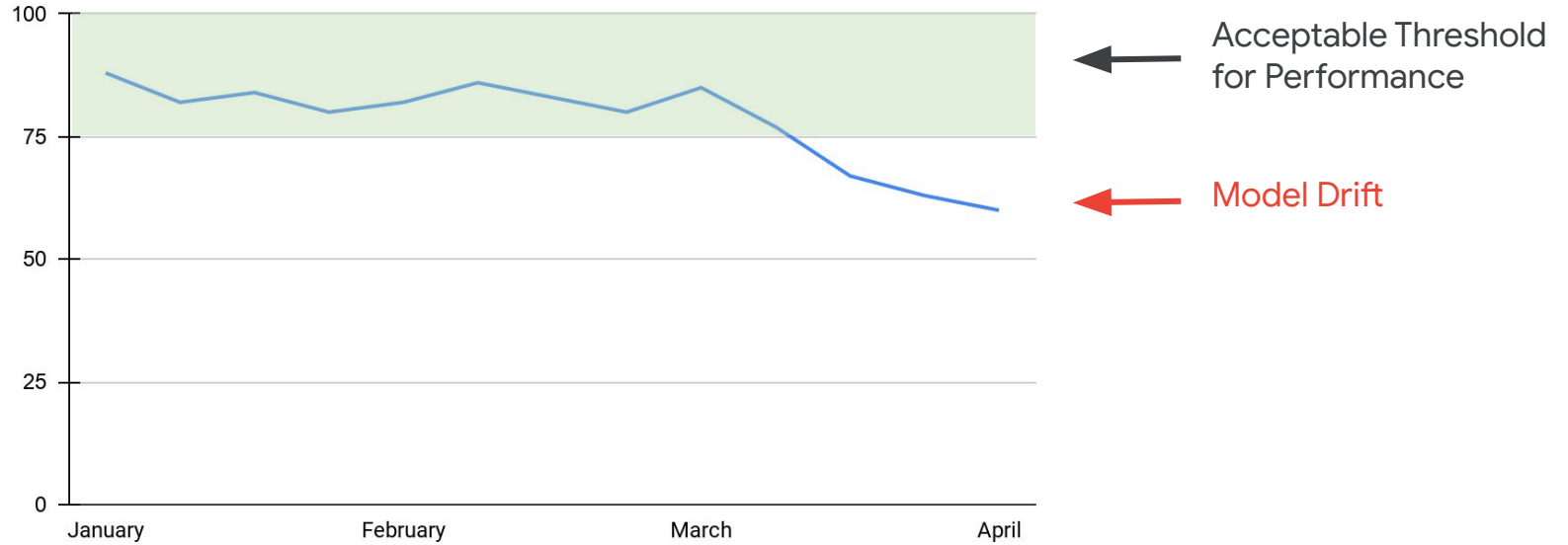


Concept drift, detrimental performance impact

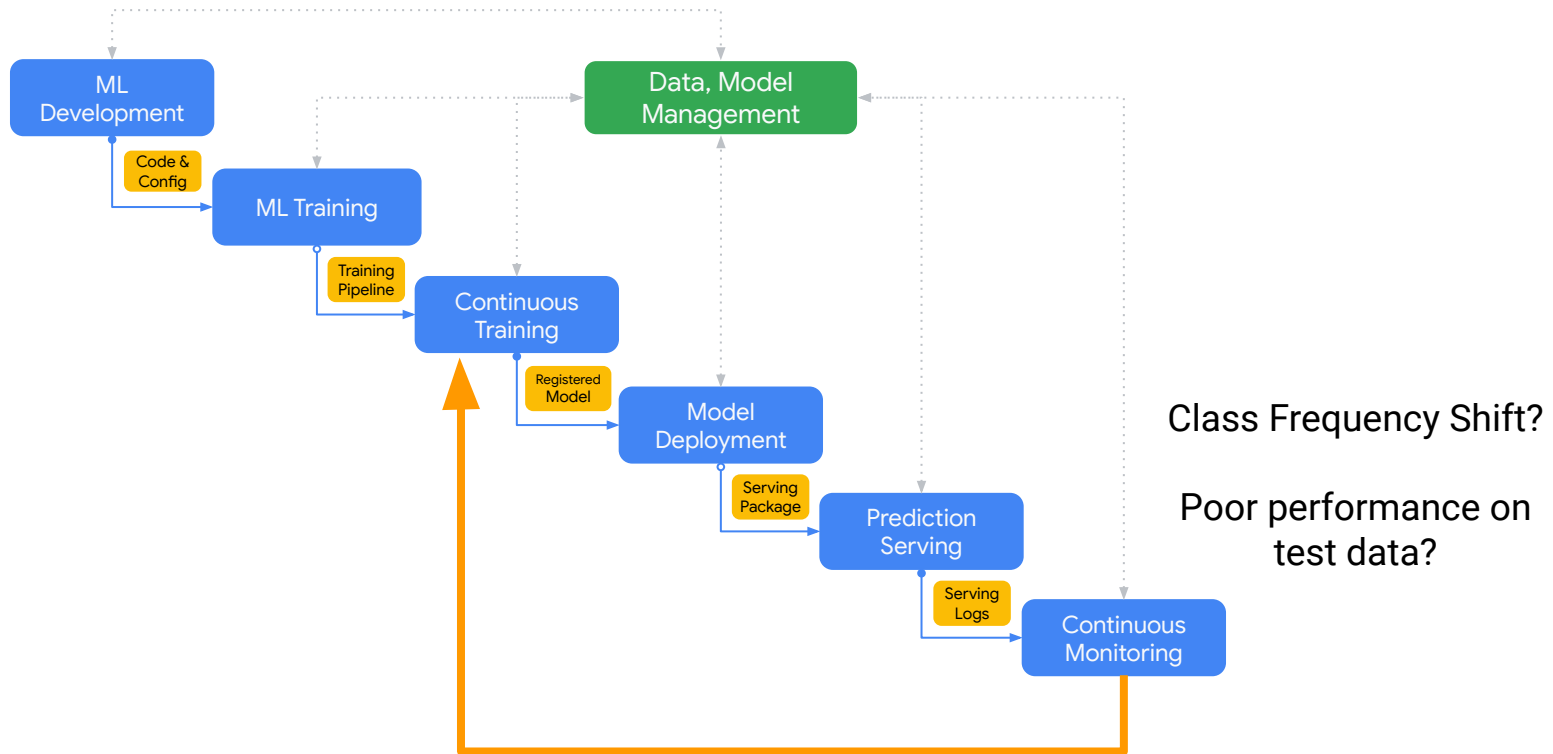
Li-ion Battery Example



Model Performance - Accuracy Rate



Trigger Retraining



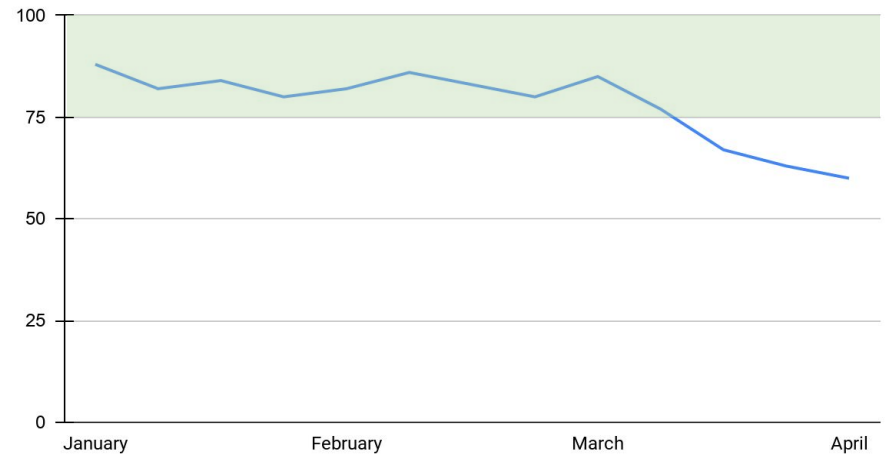
Connectivity Trade-Off



Connectivity Trade-Off



Model Performance - Accuracy Rate



Connectivity Trade-Off



You're **always** doing MLOps,
so make sure you do it well.

MLOps for Scaling TinyML

This course introduces learners to Machine Learning Operations (MLOps) through the lens of TinyML (Tiny Machine Learning). Learners explore best practices to deploy, monitor, and maintain (tiny) Machine Learning models in production at scale.



7 weeks

2–4 hours per week



Self-paced

Progress at your own speed



Free

Optional upgrade available

Choose your session:

After a course session ends, it will be [archived](#) .

Starts Apr 18

Ends Apr 26

Enroll

Starts Apr 26

Enroll