Monitoring After Deployment

Responsible AI: Human-Centered Design

START	DESIGN	DEVELOPMENT	DEPLOYMENT	END

Course 1
Fundamentals of TinyML

- **What** am I building?
- **Who** am I building this for?
- What are the **consequences** for the user if it *fails*?

Course 2
Applications of TinyML

- **What data** will be collected to train the model?
- Is the dataset **biased**?
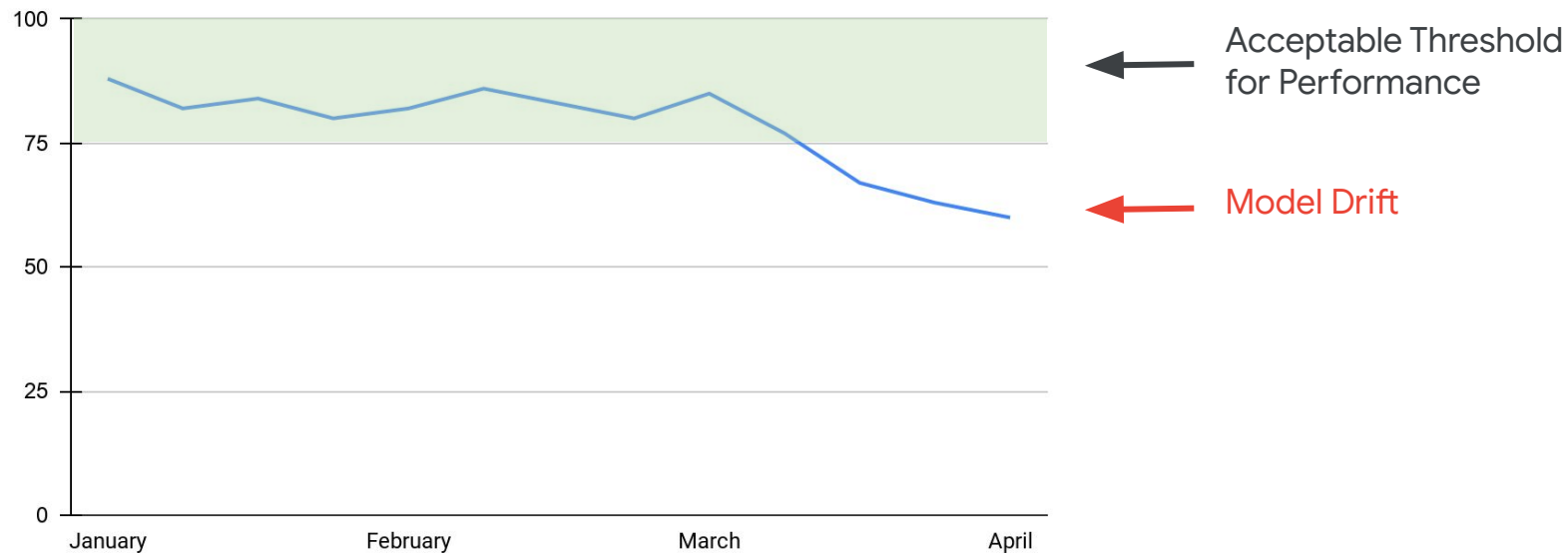- How can we **ensure** the model is **fair**?

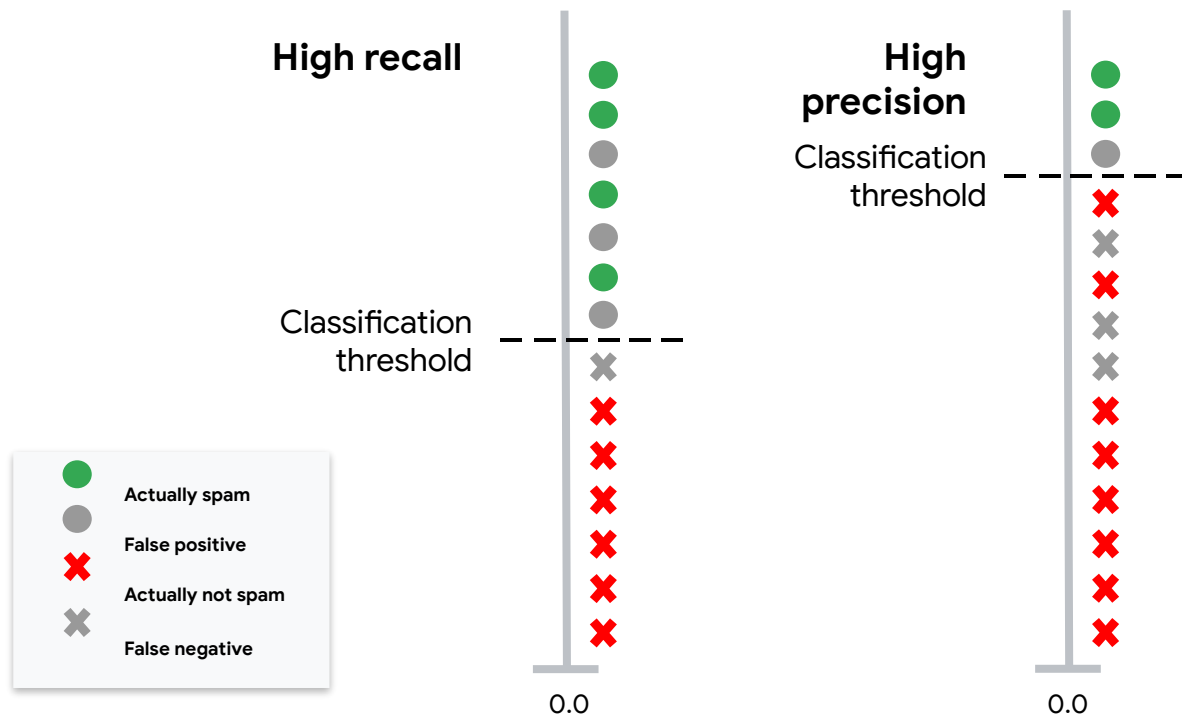**Course 3**
Deploying TinyML

- How will **model drift** be monitored?
- How should security breaches be addressed?
- How should the user's privacy be protected?

Detecting Drift—Performance

Model Performance - Accuracy Rate

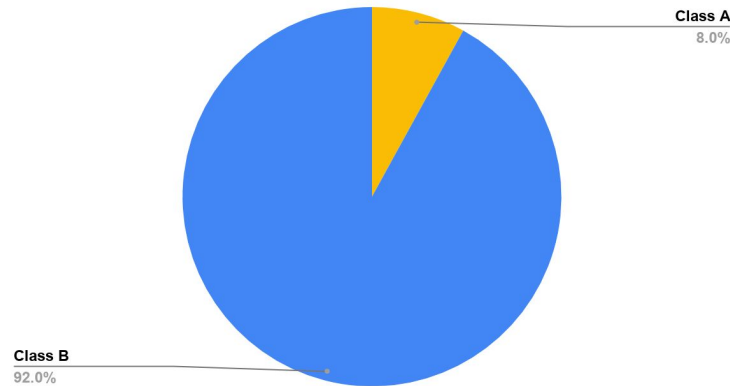Acceptable Threshold for Performance

Model Drift

Precision and Recall

High recall

High precision

Classification threshold

Classification threshold

Actually spam

False positive

Actually not spam

False negative

0.0

0.0

Detecting Drift—Predictions

Baseline Model Predictions

Class B
50.3%

Class A
49.7%

Target Model Predictions

Class A
8.0%

Class B
92.0%

Detecting Drift—Data Distributions

BASELINE
60% male, income 50k, etc.

TARGET
80% male, income 60k, etc.

Retraining

Concept Drift

the affected *old data*
needs to be relabeled

Data Drift

enough new data
needs to be labeled

Retraining for **TinyML**

- Retraining may **not always** be a **feasible** option
  - Low power communication protocol
  - Device isn't wifi-enabled
- Retraining opens up **security risks**