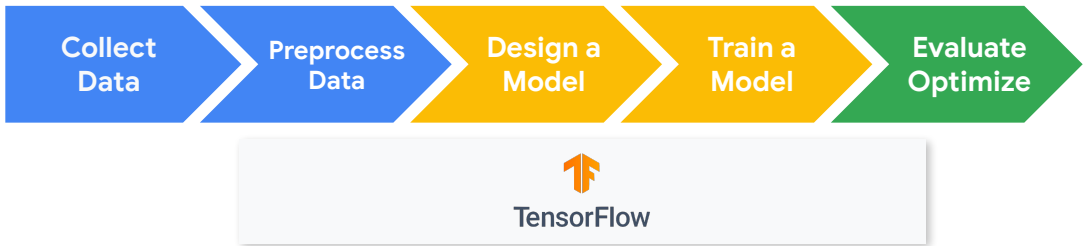
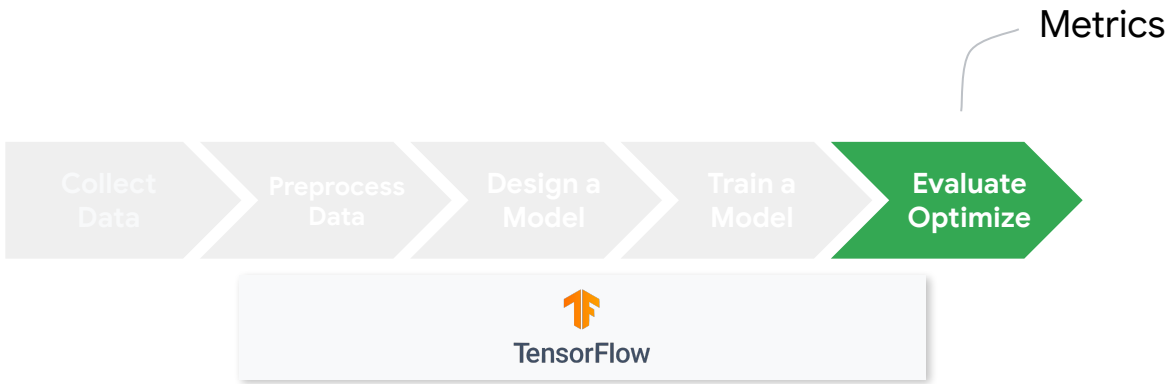
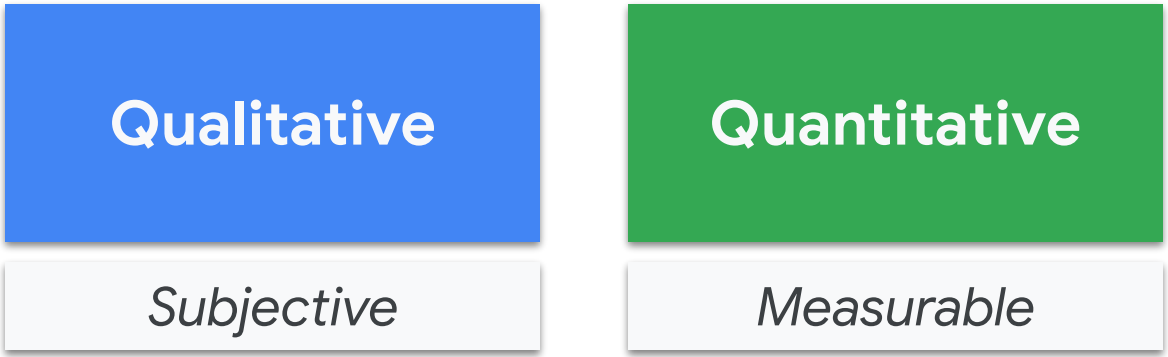
Metrics for *VWW*


Collect Data
Preprocess Data
Design a Model
Train a Model
Evaluate Optimize
TensorFlow


Metrics
Collect Data
Preprocess Data
Design a Model
Train a Model
Evaluate Optimize
TensorFlow

Metrics **Anatomy**



Common Metrics

Accuracy

Quantitative

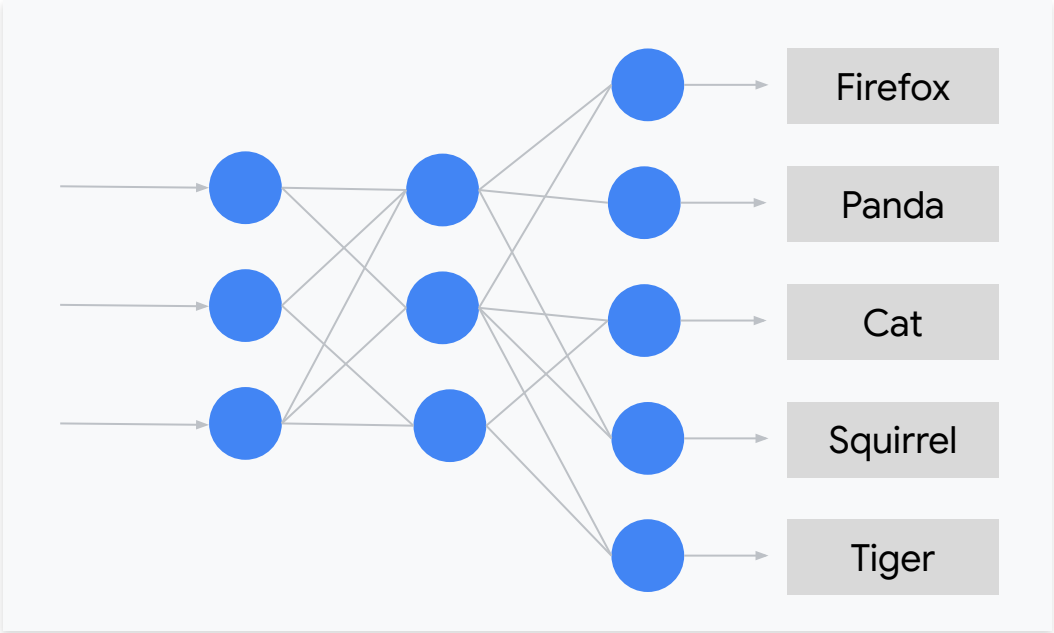
Efficiency

Quantitative

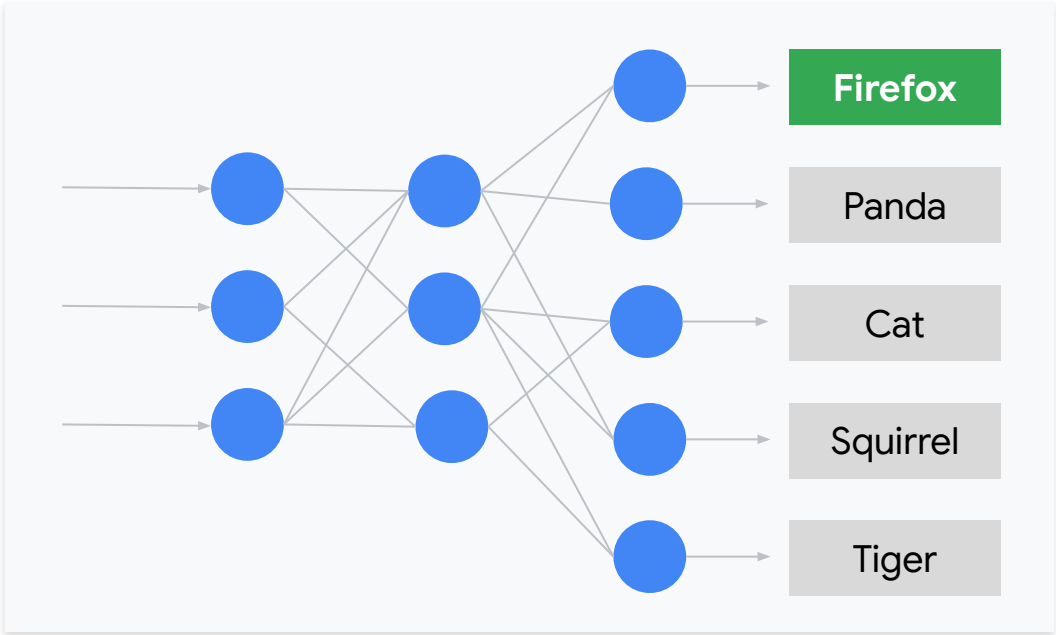
User Experience

Qualitative

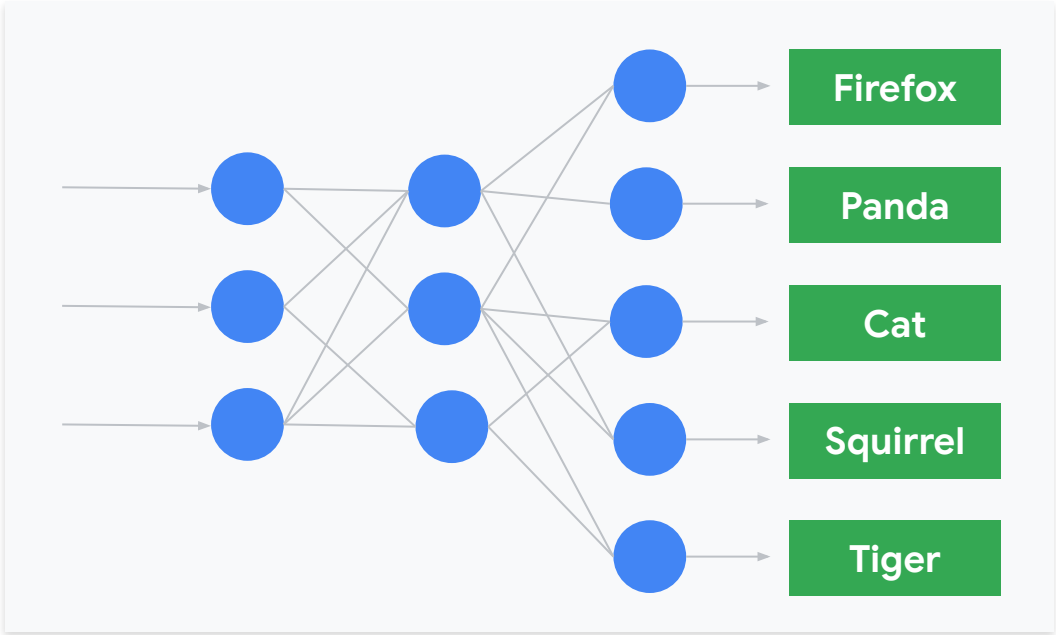
Model **Accuracy**



Top-1 Accuracy



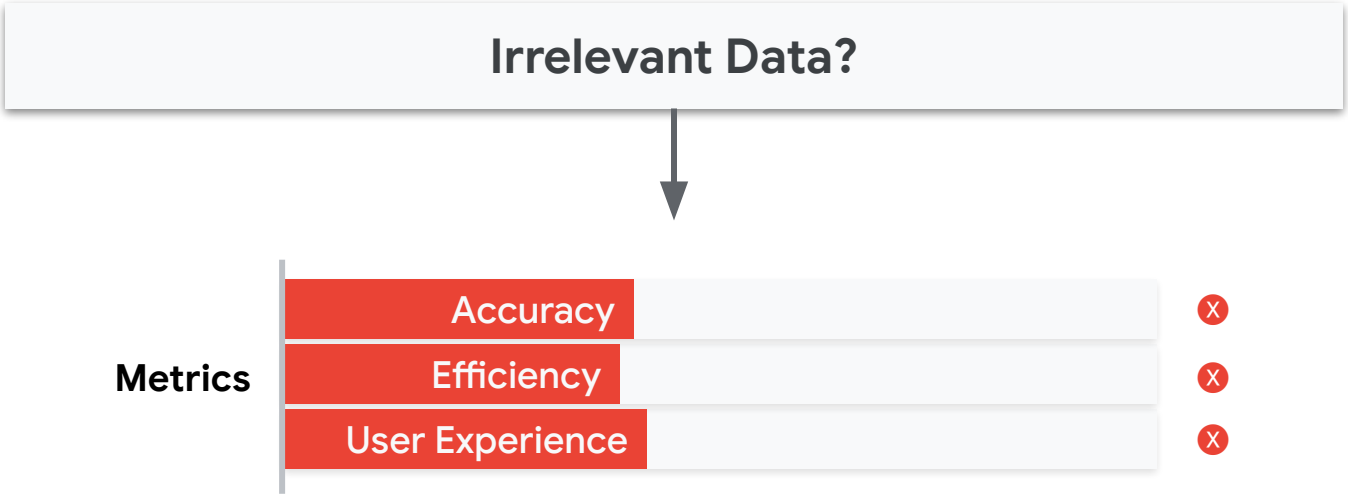
Top-5 Accuracy



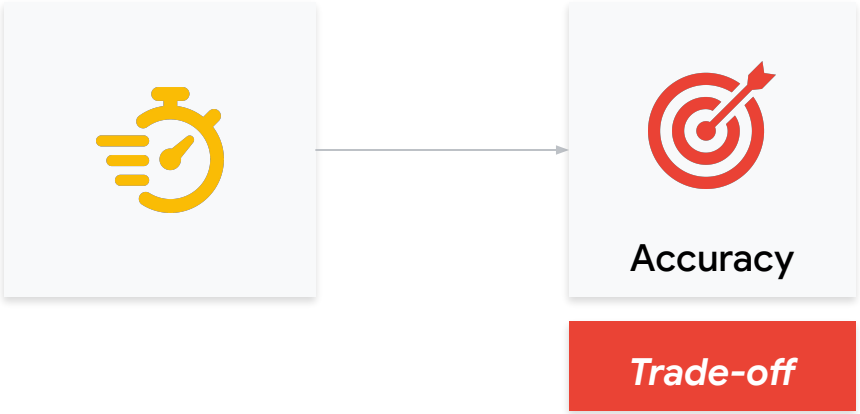
Accuracy

Representative (relevant) data is important because it affects the accuracy of the model

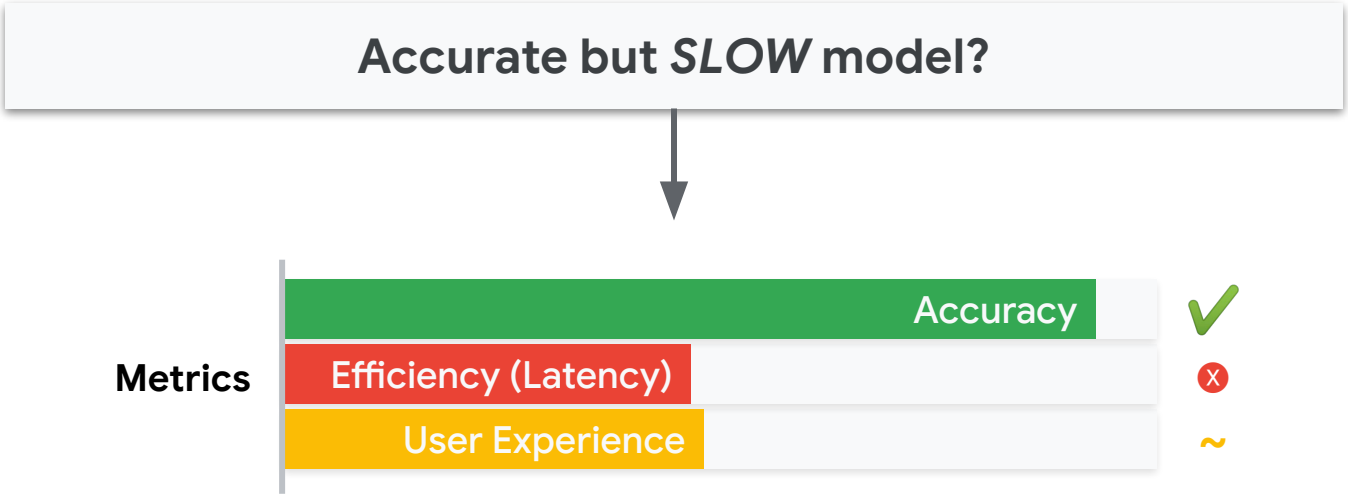
Accuracy


Irrelevant Data?
Metrics
Accuracy
Efficiency
User Experience
X
X
X

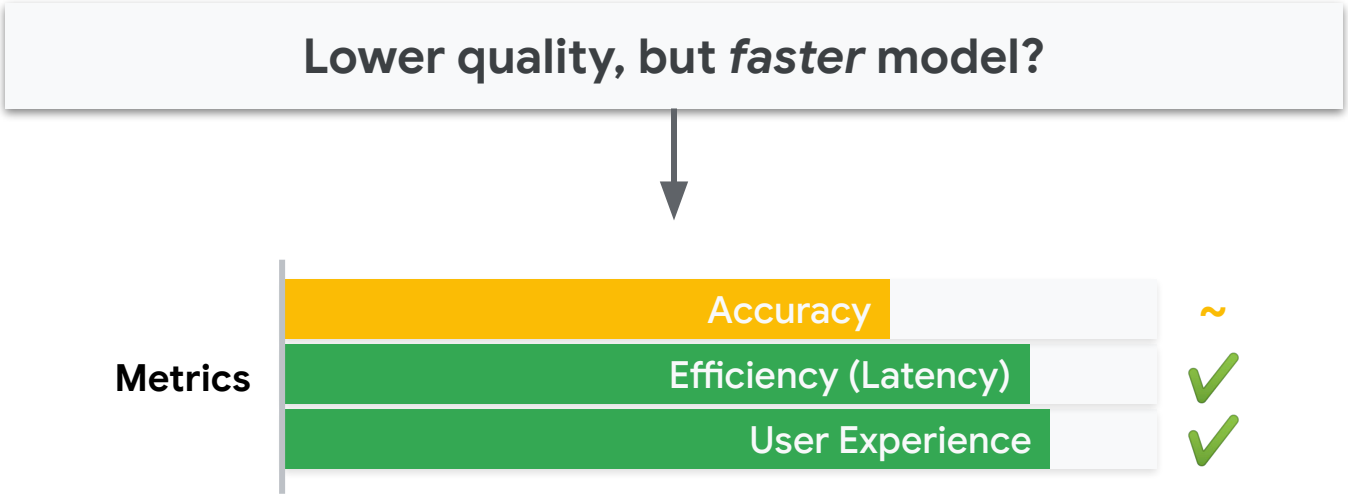
Latency



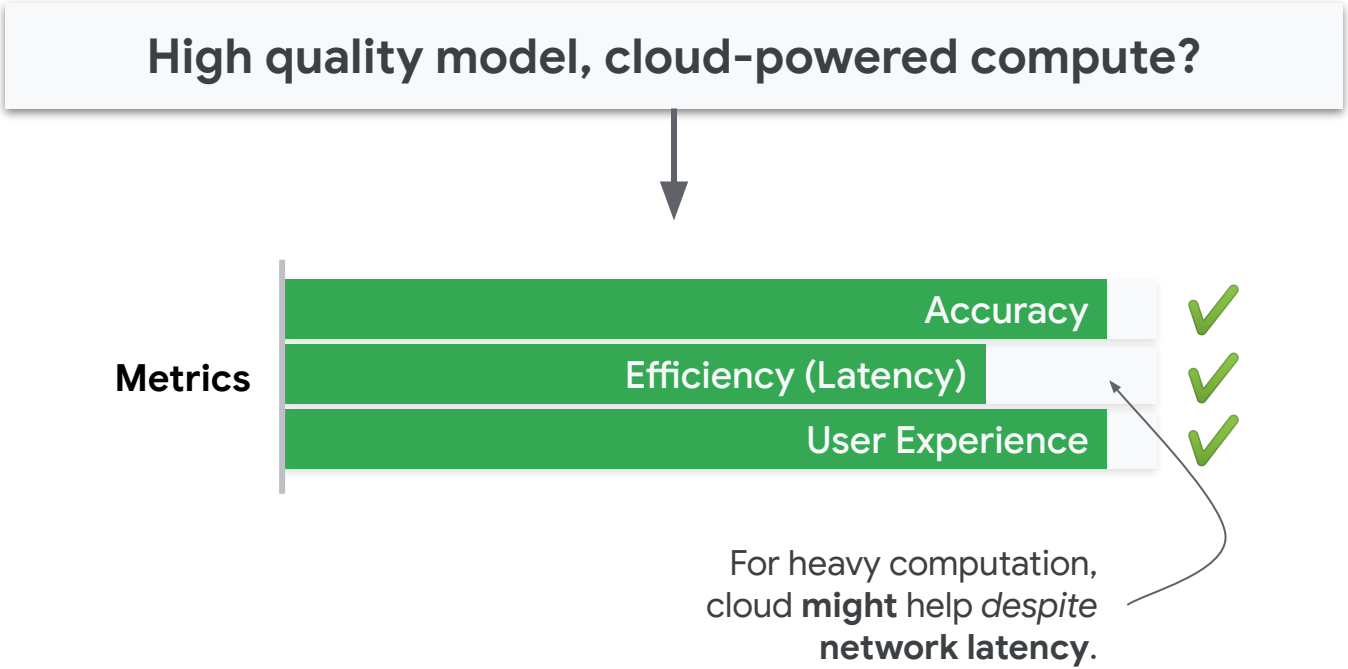
Latency


Accurate but *SLOW* model?
Metrics
Accuracy
Efficiency (Latency)
User Experience
✓
x
~

Latency


Lower quality, but *faster* model?
Metrics
Accuracy
Efficiency (Latency)
User Experience
~
✔
✔

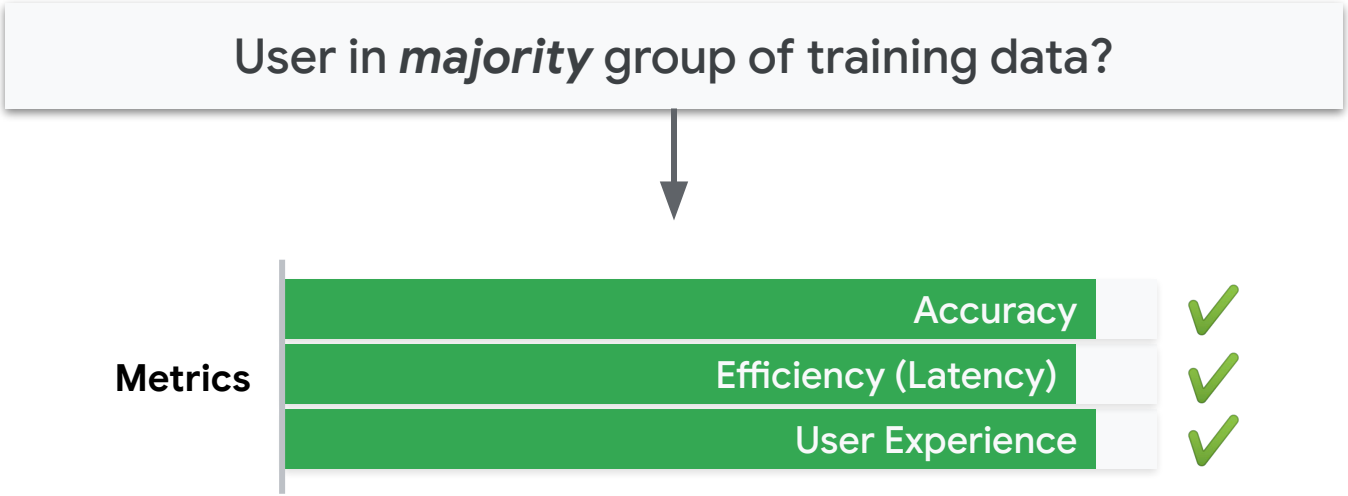
Latency


High quality model, cloud-powered compute?
Metrics
Accuracy
Efficiency (Latency)
User Experience
For heavy computation, cloud **might** help *despite* **network latency**.

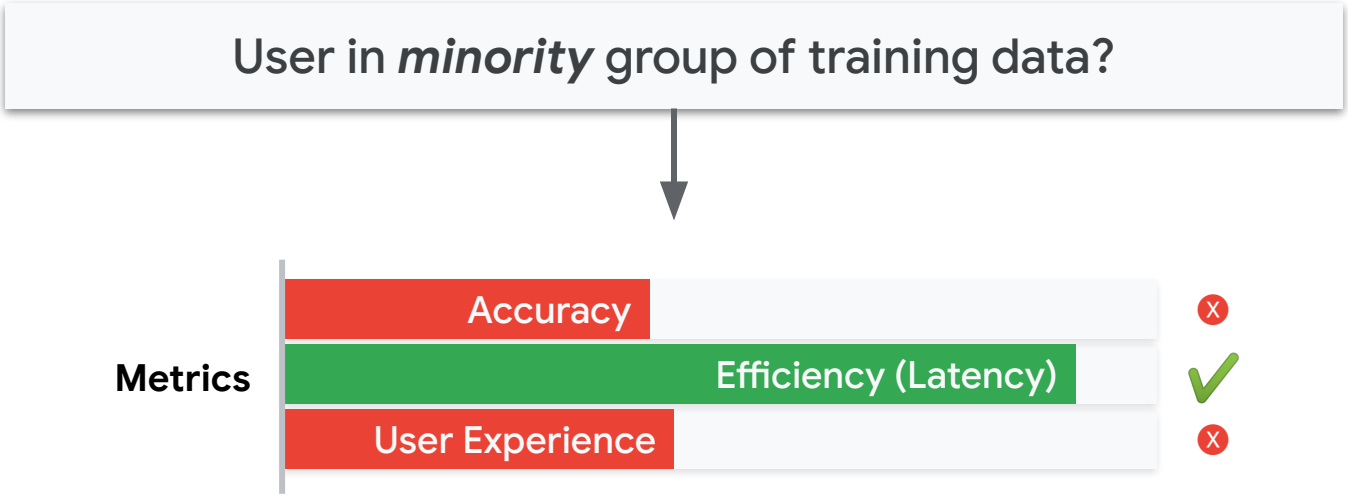
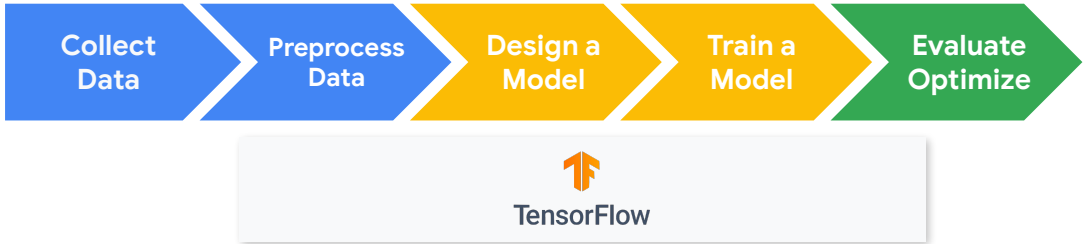
Fairness

***Diverse*, representative data** is important because it enables fair use (equal performance) across populations

Fairness


User in *majority* group of training data?
Metrics
Accuracy
Efficiency (Latency)
User Experience

Fairness


User in *minority* group of training data?
Metrics
Accuracy
Efficiency (Latency)
User Experience

Achieving Ideal Metrics: Revisit pipeline