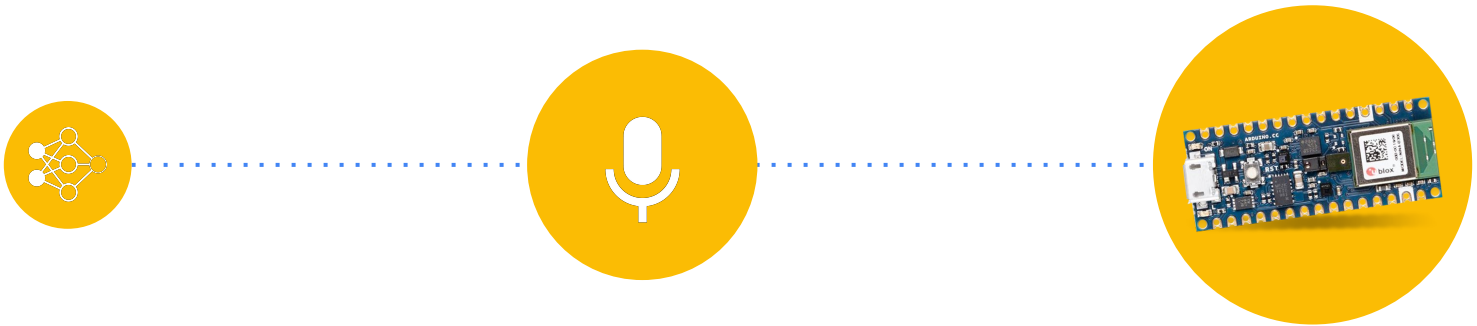
“What Do I Do Now?”

You’ve come a long way.

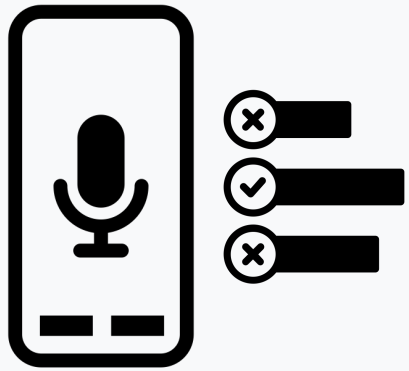
ML Workflow

Collect Data → Preprocess Data → Design a Model → Train a Model → Evaluate Optimize → Convert Model → Deploy Model → Make Inferences



ML Workflow

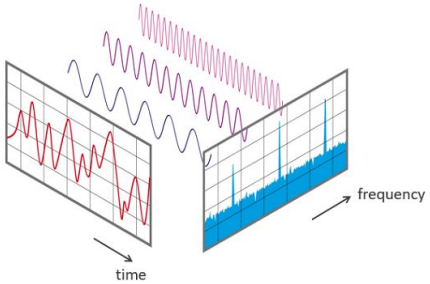
Collect Data → Preprocess Data → Design a Model → Train a Model → Evaluate Optimize → Convert Model → Deploy Model → Make Inferences

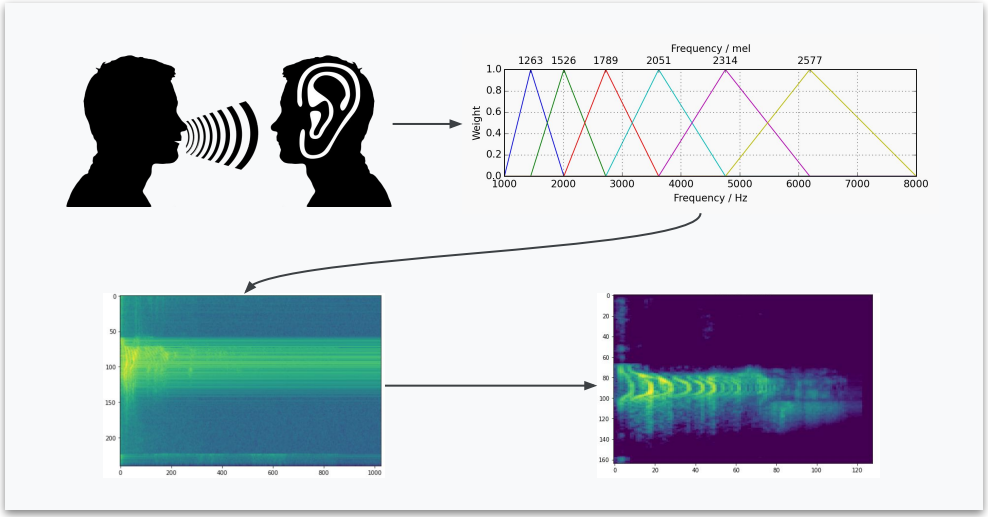


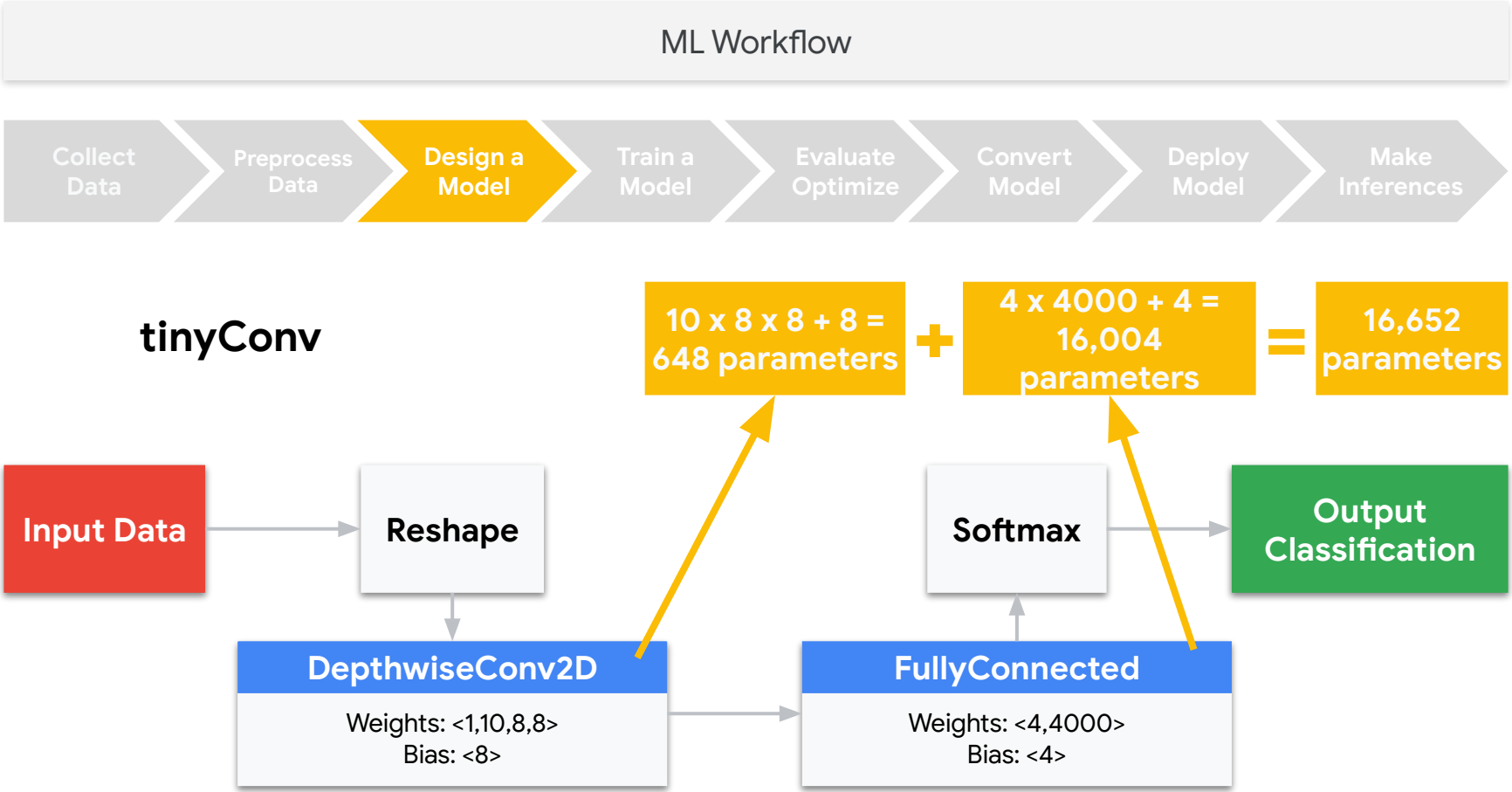
Your custom voice assistant

ML Workflow

Collect Data → **Preprocess Data** → Design a Model → Train a Model → Evaluate Optimize → Convert Model → Deploy Model → Make Inferences

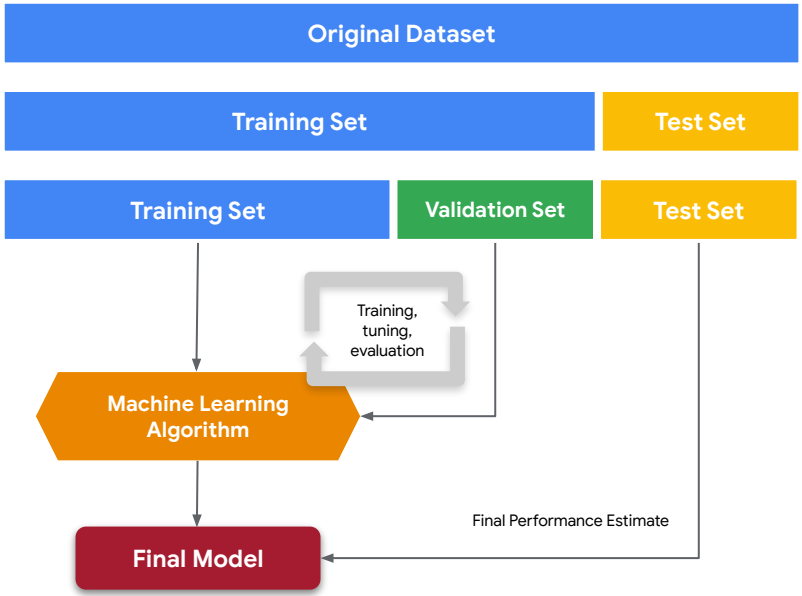


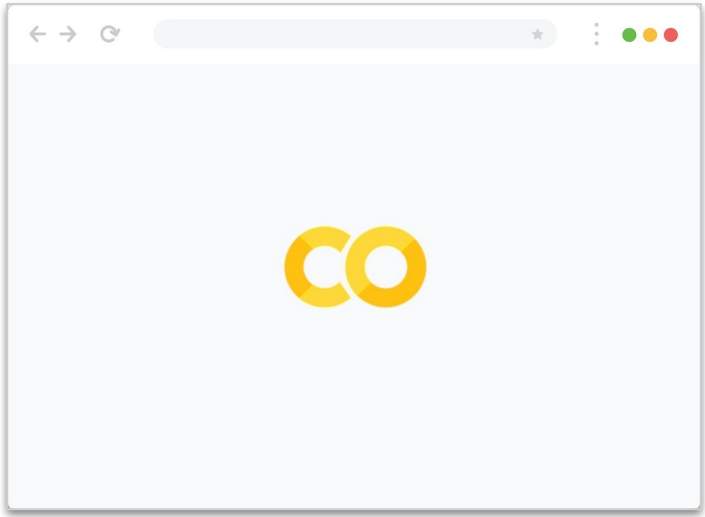



ML Workflow
Collect Data
Preprocess Data
Design a Model
Train a Model
Evaluate Optimize
Convert Model
Deploy Model
Make Inferences
tinyConv
10 x 8 x 8 + 8 = 648 parameters
+
4 x 4000 + 4 = 16,004 parameters
=
16,652 parameters
Input Data
Reshape
DepthwiseConv2D
Weights: <1,10,8,8>
Bias: <8>
FullyConnected
Weights: <4,4000>
Bias: <4>
Softmax
Output Classification

ML Workflow

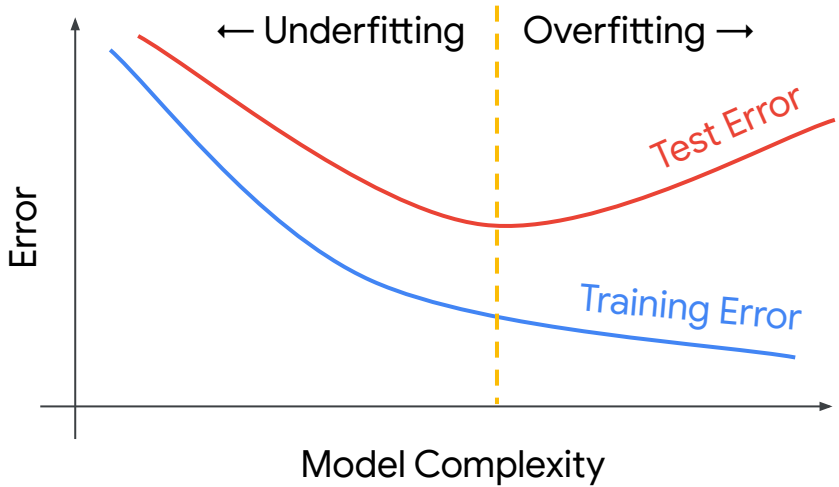
Collect Data → Preprocess Data → Design a Model → **Train a Model** → Evaluate Optimize → Convert Model → Deploy Model → Make Inferences

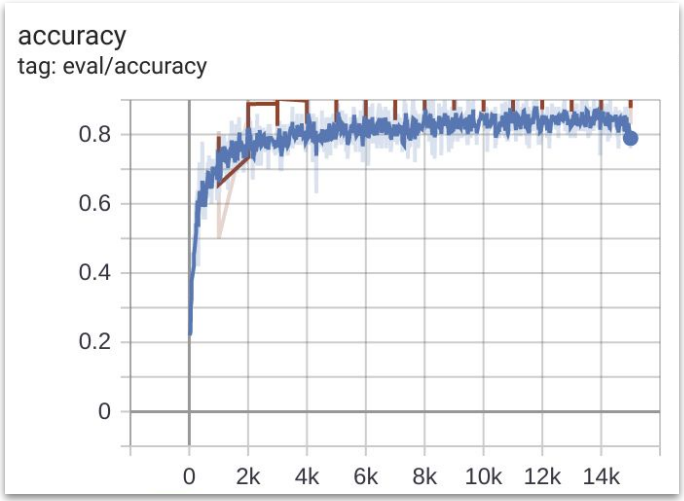




ML Workflow

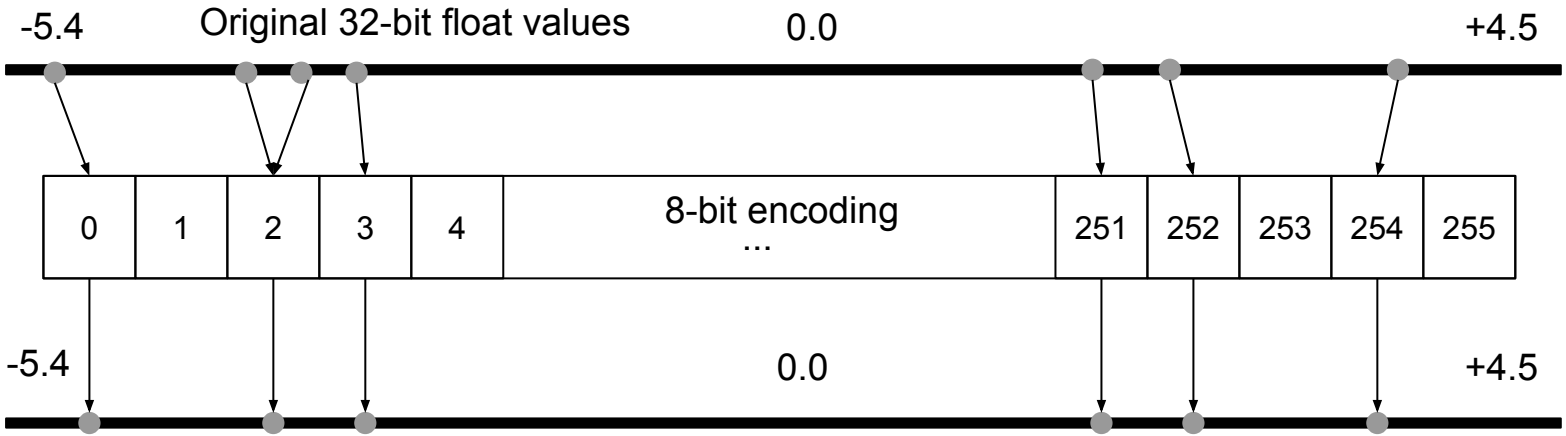
Collect Data → Preprocess Data → Design a Model → Train a Model → **Evaluate Optimize** → Convert Model → Deploy Model → Make Inferences



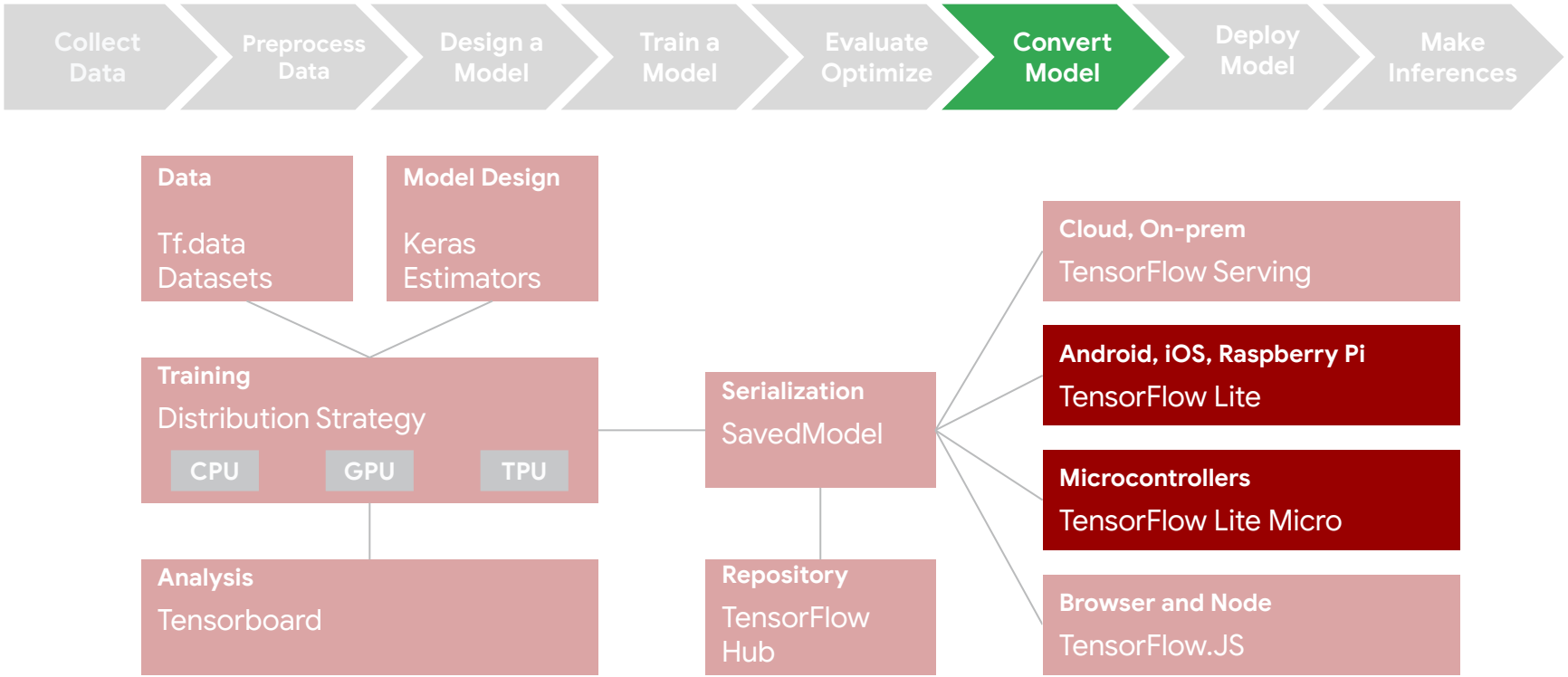


ML Workflow

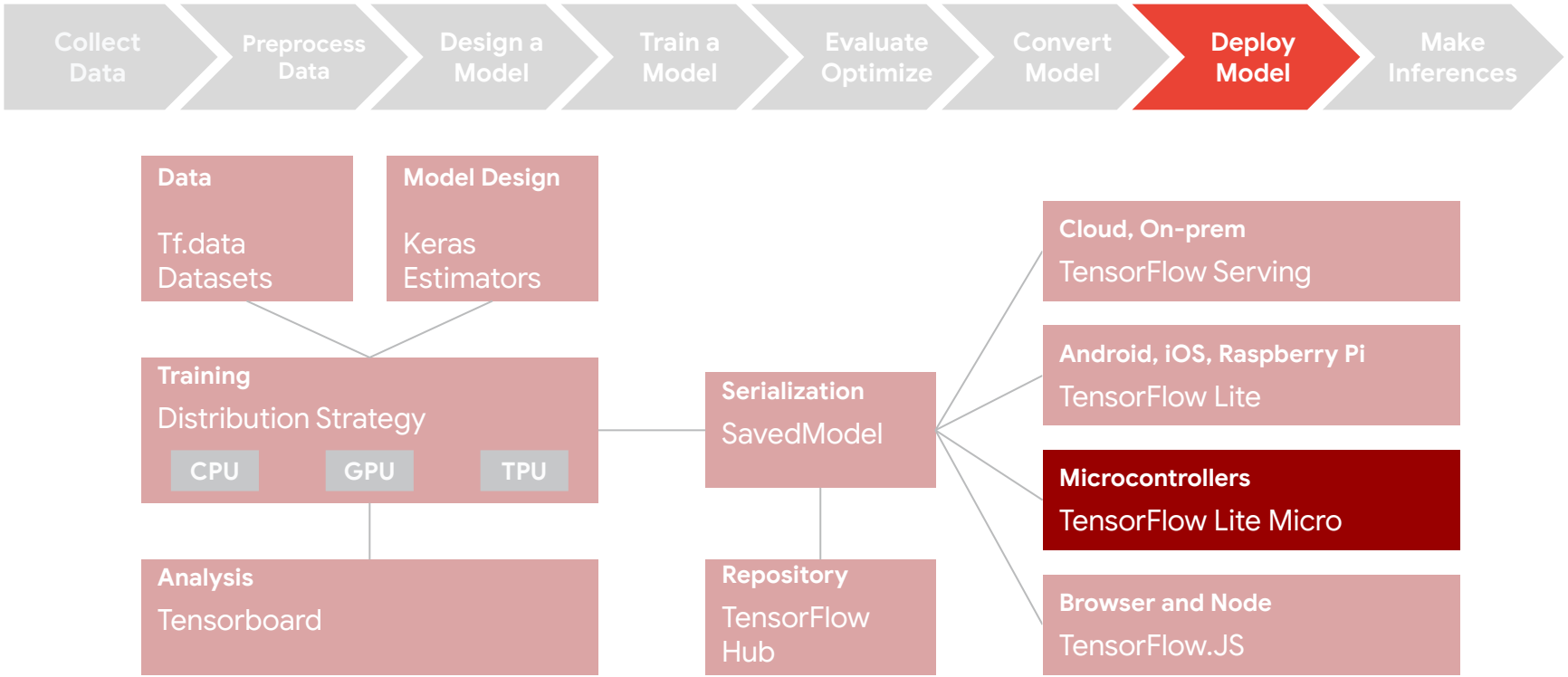
Collect Data → Preprocess Data → Design a Model → Train a Model → **Evaluate Optimize** → Convert Model → Deploy Model → Make Inferences



ML Workflow

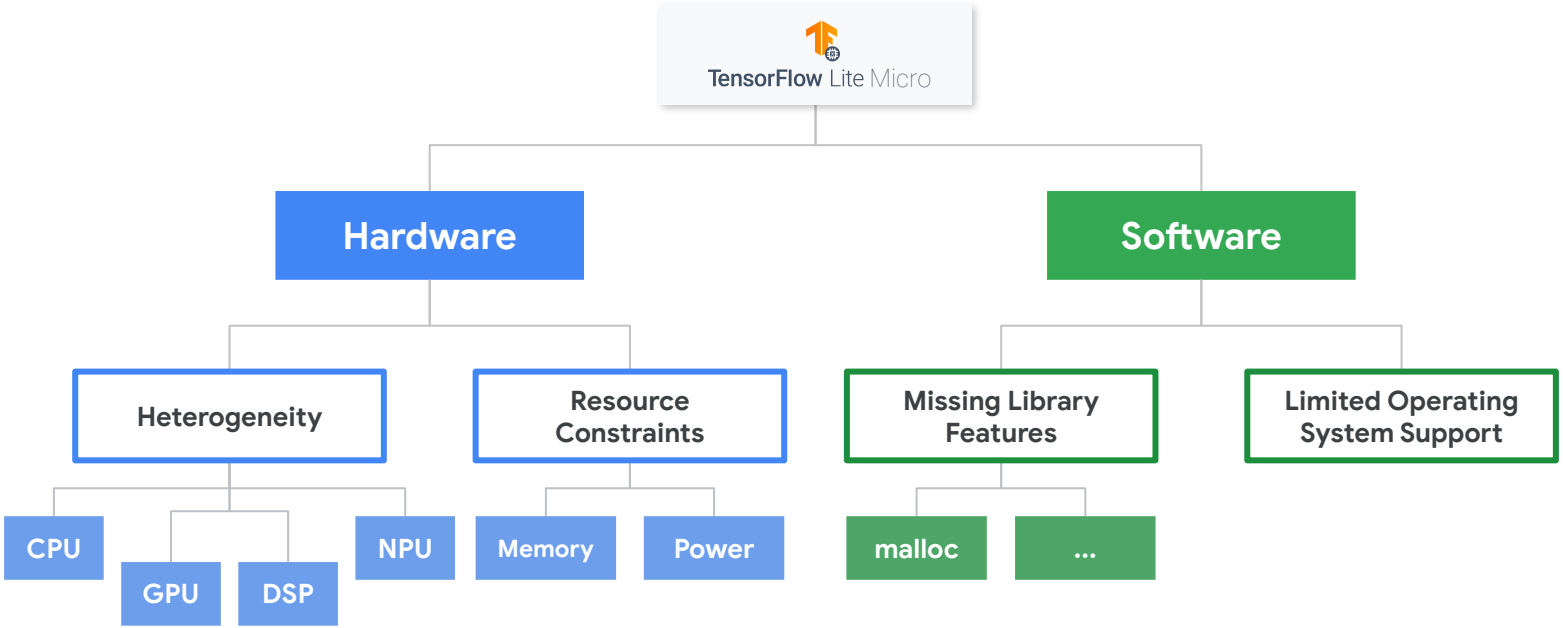

Collect Data
Preprocess Data
Design a Model
Train a Model
Evaluate Optimize
Convert Model
Deploy Model
Make Inferences
Data
Tf.data Datasets
Model Design
Keras Estimators
Training
Distribution Strategy
CPU
GPU
TPU
Analysis
Tensorboard
Serialization
SavedModel
Repository
TensorFlow Hub
Cloud, On-prem
TensorFlow Serving
Android, iOS, Raspberry Pi
TensorFlow Lite
Microcontrollers
TensorFlow Lite Micro
Browser and Node
TensorFlow.JS

ML Workflow


Collect Data
Preprocess Data
Design a Model
Train a Model
Evaluate Optimize
Convert Model
Deploy Model
Make Inferences
Data
Tf.data Datasets
Model Design
Keras Estimators
Training
Distribution Strategy
CPU
GPU
TPU
Analysis
Tensorboard
Serialization
SavedModel
Repository
TensorFlow Hub
Cloud, On-prem
TensorFlow Serving
Android, iOS, Raspberry Pi
TensorFlow Lite
Microcontrollers
TensorFlow Lite Micro
Browser and Node
TensorFlow.JS

ML Workflow


Collect Data
Preprocess Data
Design a Model
Train a Model
Evaluate Optimize
Convert Model
Deploy Model
Make Inferences


TensorFlow Lite Micro
Hardware
Software
Heterogeneity
Resource Constraints
Missing Library Features
Limited Operating System Support
CPU
GPU
DSP
NPU
Memory
Power
malloc
...

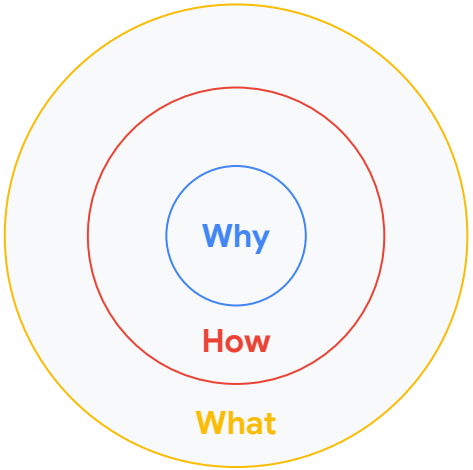
ML Workflow


Collect Data
Preprocess Data
Design a Model
Train a Model
Evaluate Optimize
Convert Model
Deploy Model
Make Inferences

ML Workflow

Applying what you've learned to a new **TinyML** application:

Start by thinking through carefully with "the golden circle" approach



ML Workflow

Why

Your purpose, cause, beliefs.

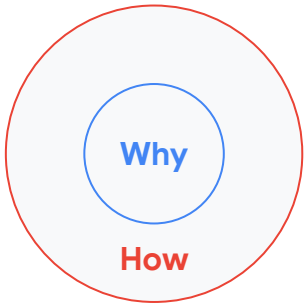
ML Workflow

Why

Your purpose, cause, beliefs.

How

Value proposition, what you add (unique from your competition).



ML Workflow

Why

Your purpose, cause, beliefs.

What

What you do. Products/services.

How

Value proposition, what you add (unique from your competition).