

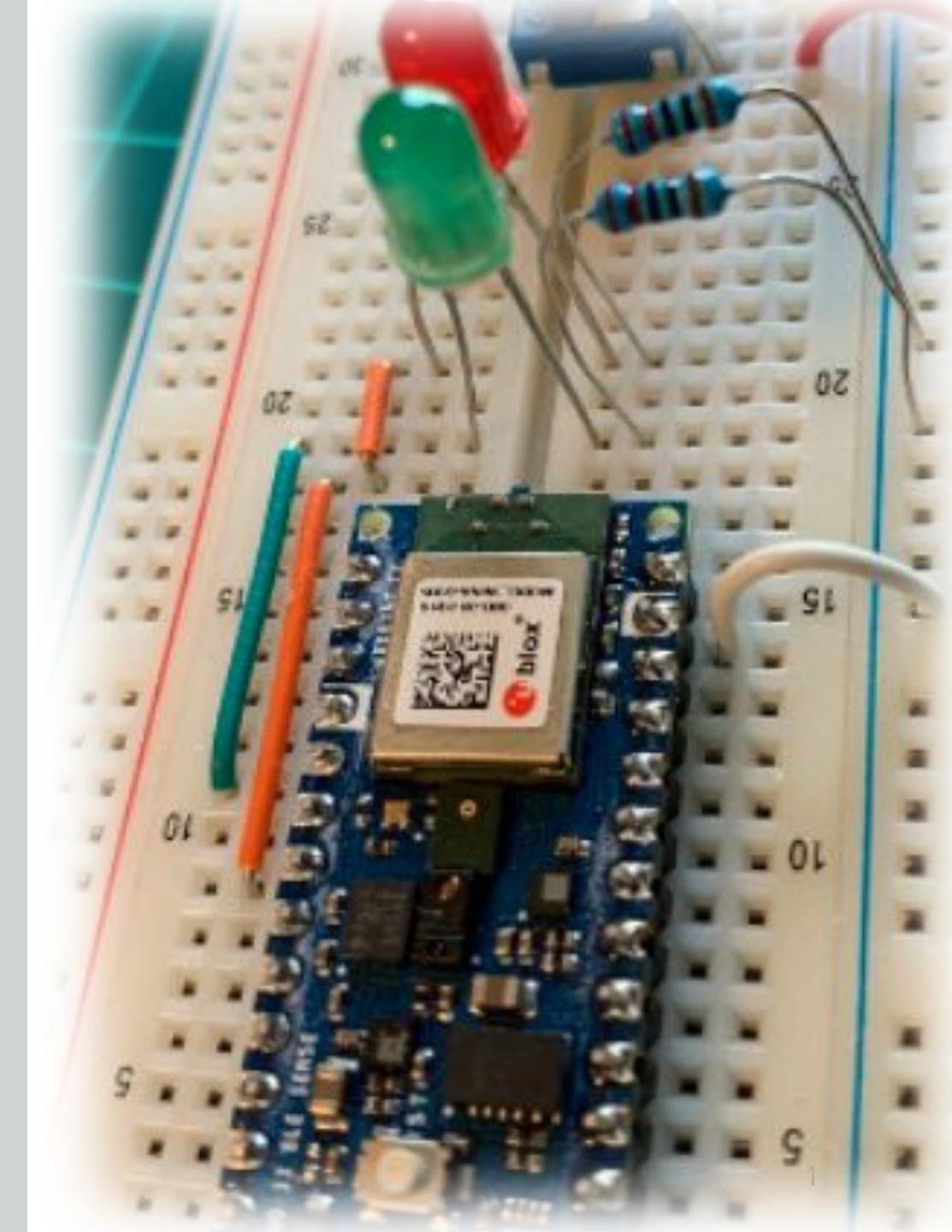
IESTI01 – TinyML

Embedded Machine Learning

12. Introduction to Edge Impulse – CNN with Cifar-10



Prof. Marcelo Rovai
UNIFEI

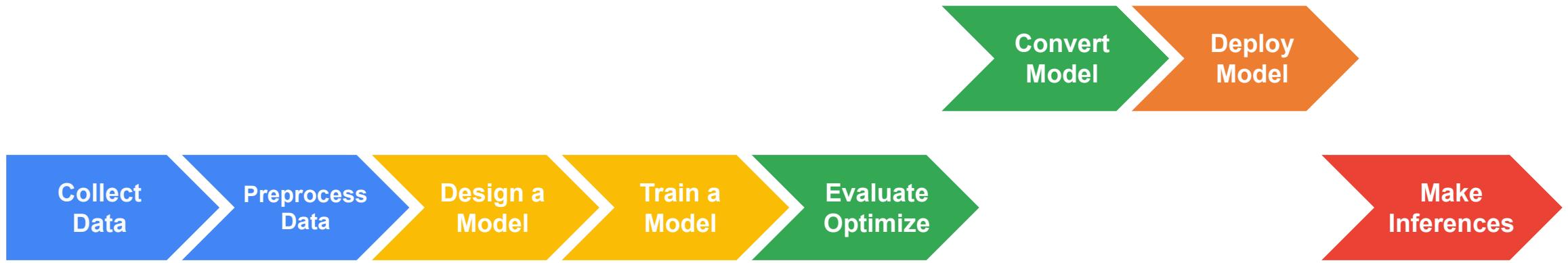


Embedded Machine Learning (TinyML) Workflow Review

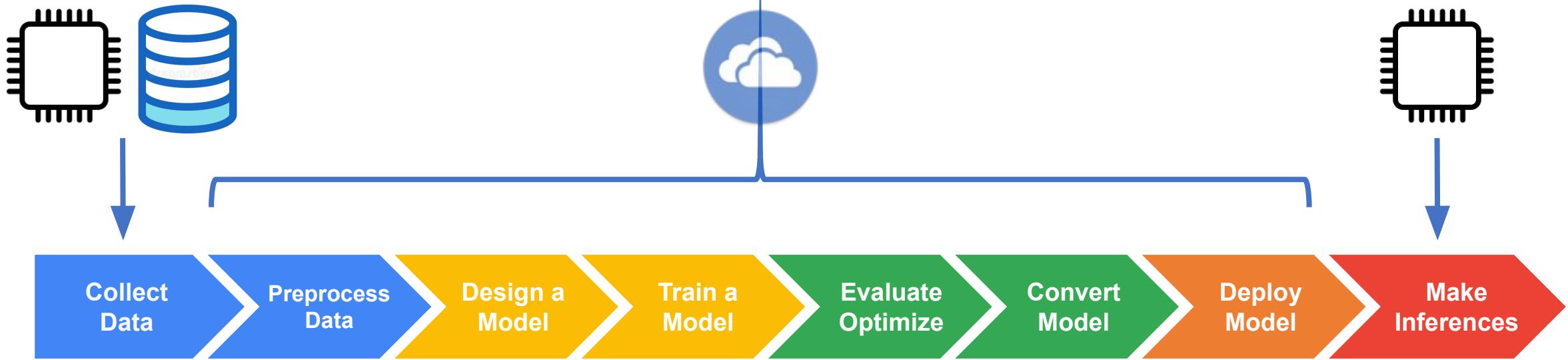
Machine Learning Workflow



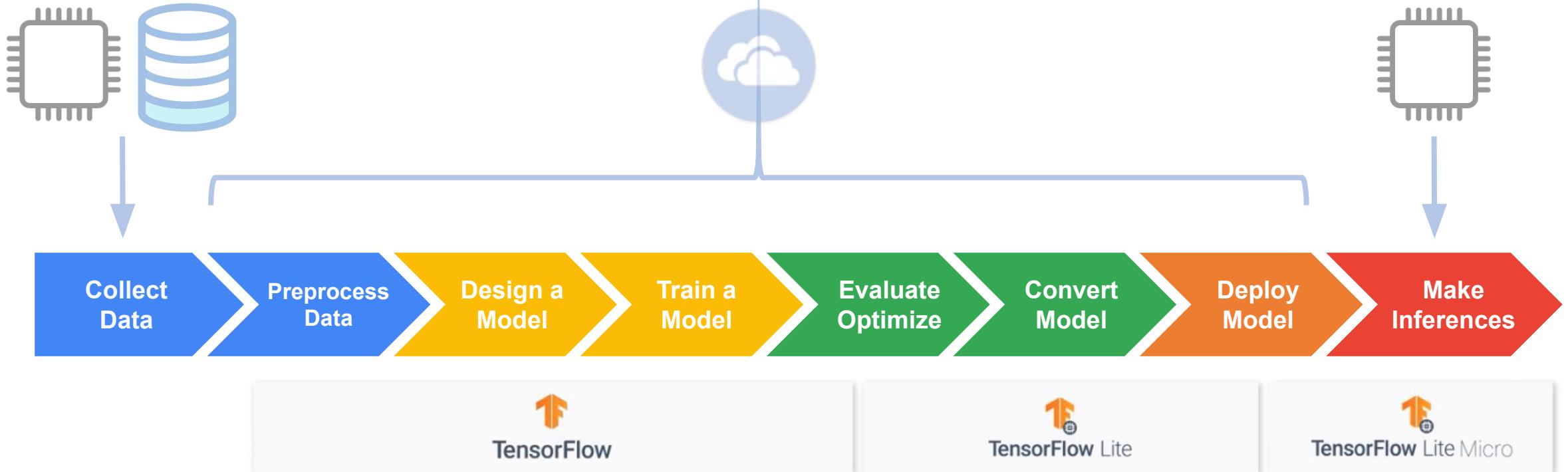
Tiny Machine Learning Workflow (“What”)



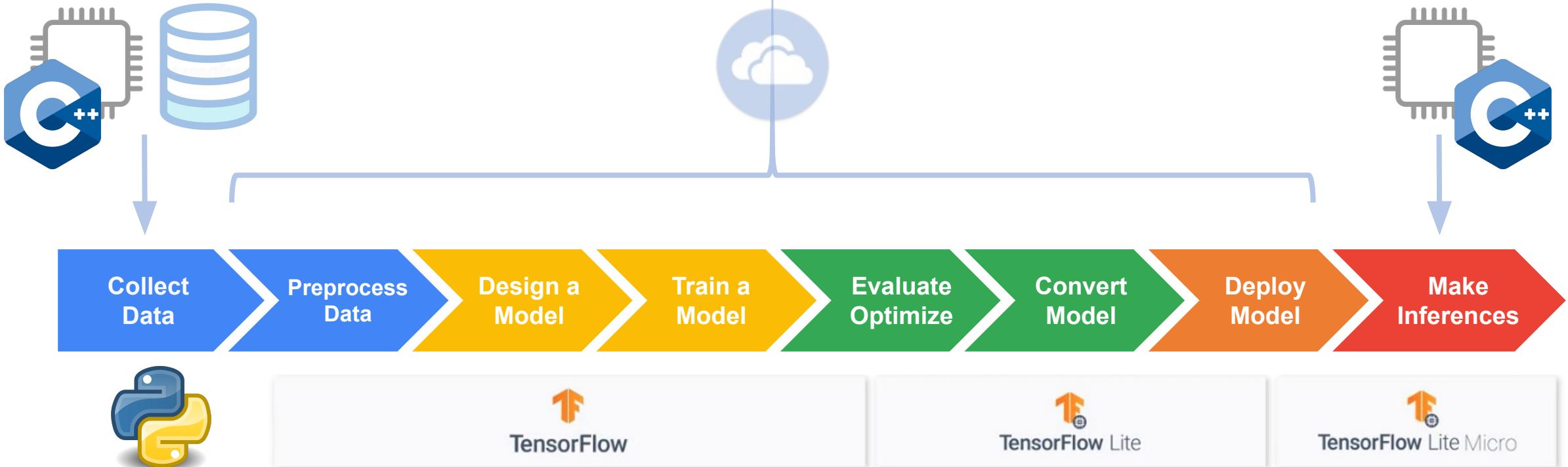
Tiny Machine Learning Workflow (“Where”)



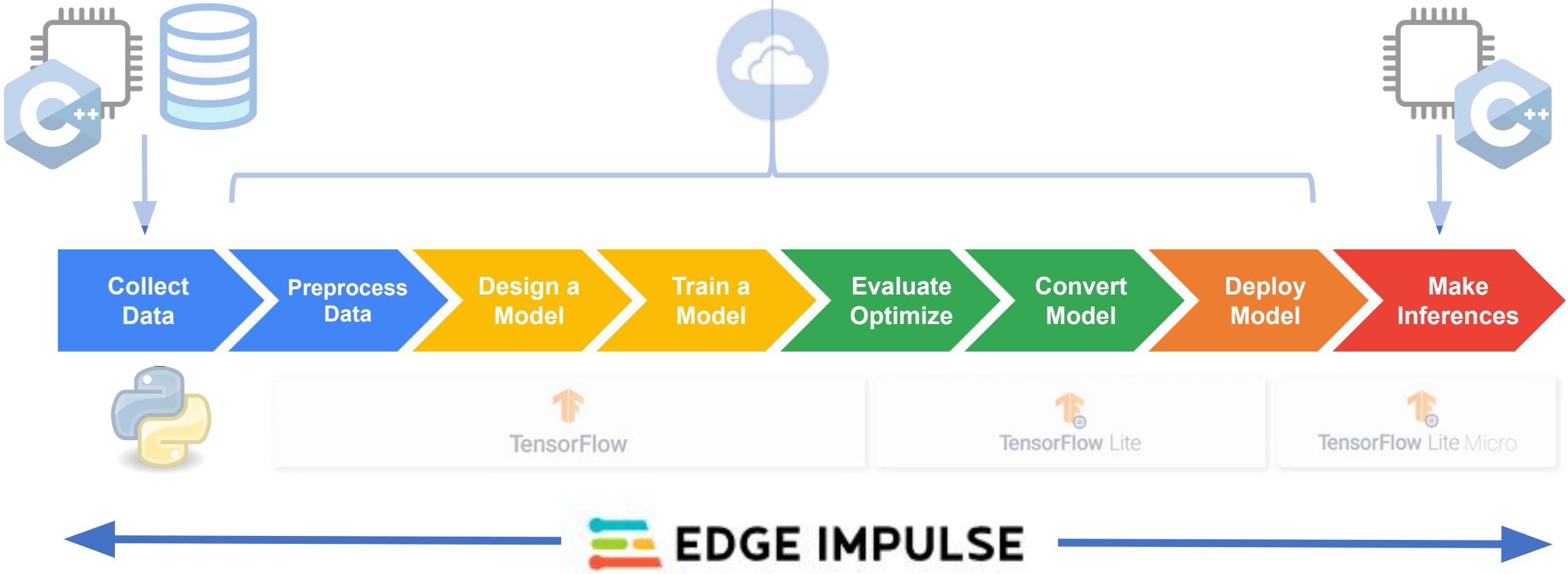
Machine Learning Workflow (“How”)

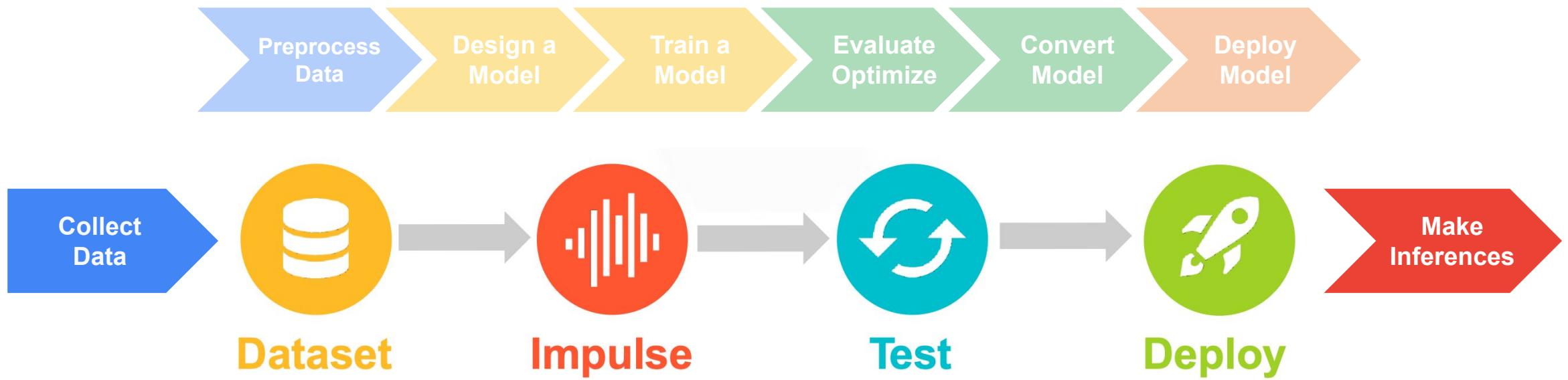


Machine Learning Workflow (“How”)

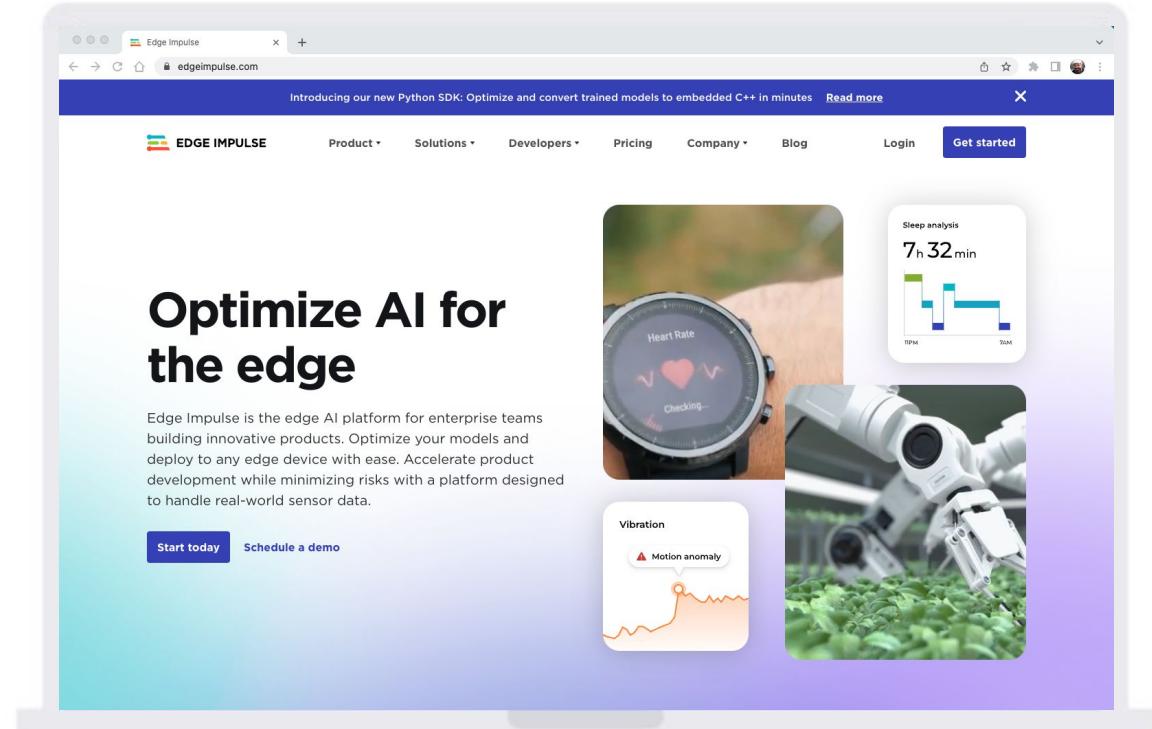
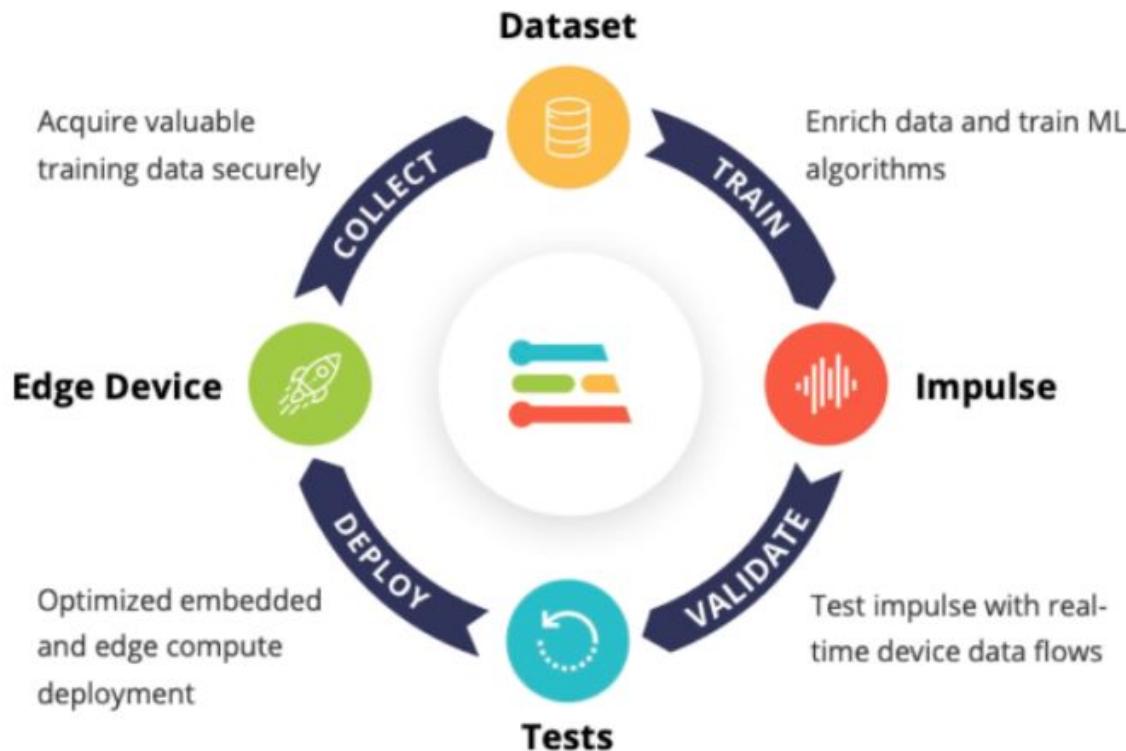


Machine Learning Workflow (“How”)





EI Studio - Embedded ML platform (“AutoML”)



Learn more at <http://edgeimpulse.com>



Cifar10 Edge Impulse Studio



Dataset: <https://github.com/YoongiKim/CIFAR-10-images>

El Studio Public Project: <https://studio.edgeimpulse.com/public/51070/latest>

The screenshot shows the homepage of the Edge Impulse website. At the top, there is a navigation bar with links for 'Product', 'Solutions', 'Developers', 'Pricing', 'Company', and 'Blog'. On the right side of the header, there are 'Login' and 'Get started' buttons. A yellow arrow labeled '1' points to the 'edgeimpulse.com' URL in the browser's address bar. Another yellow arrow labeled '2' points to the 'Get started' button.

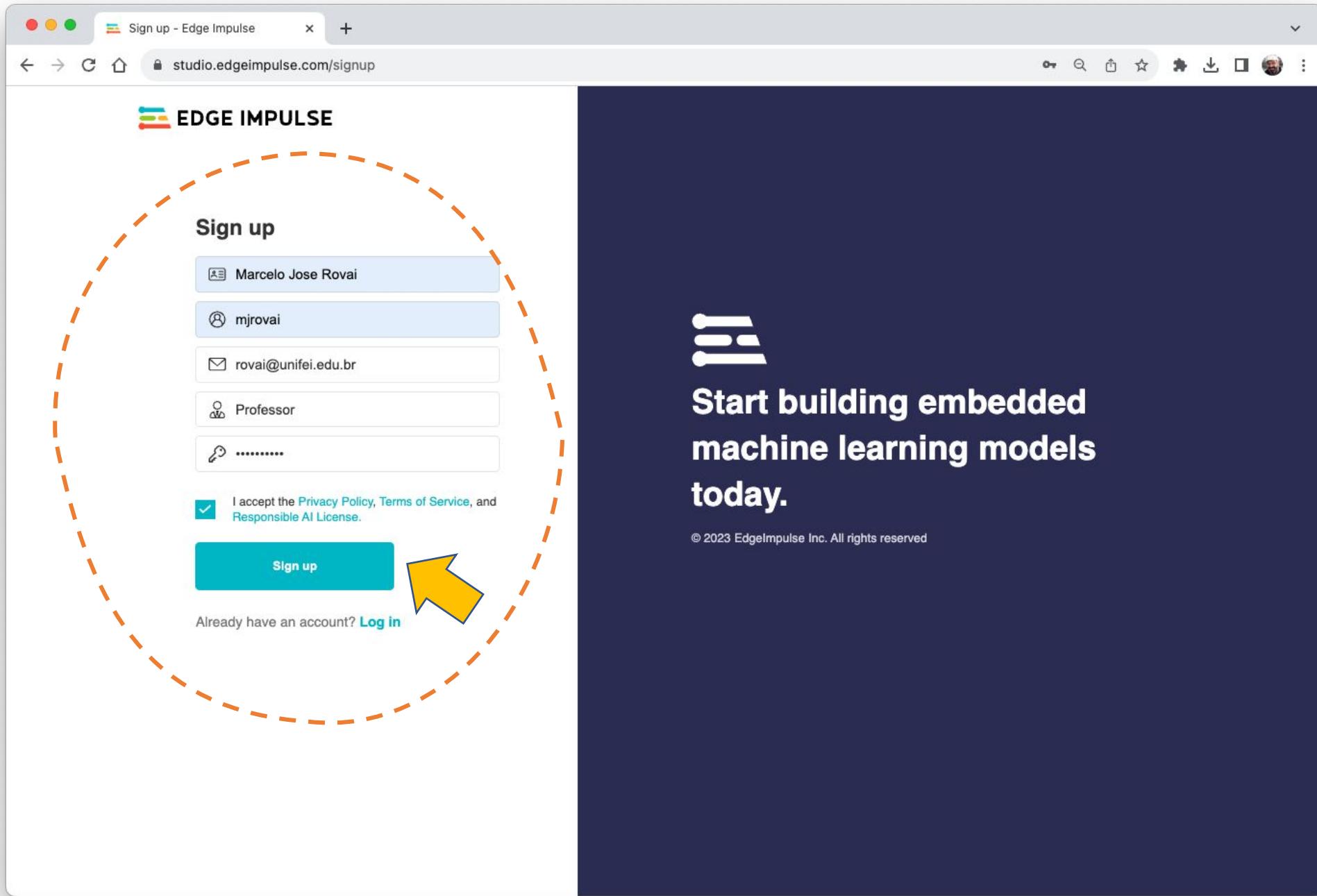
Build. Train. Optimize. AI for the edge.

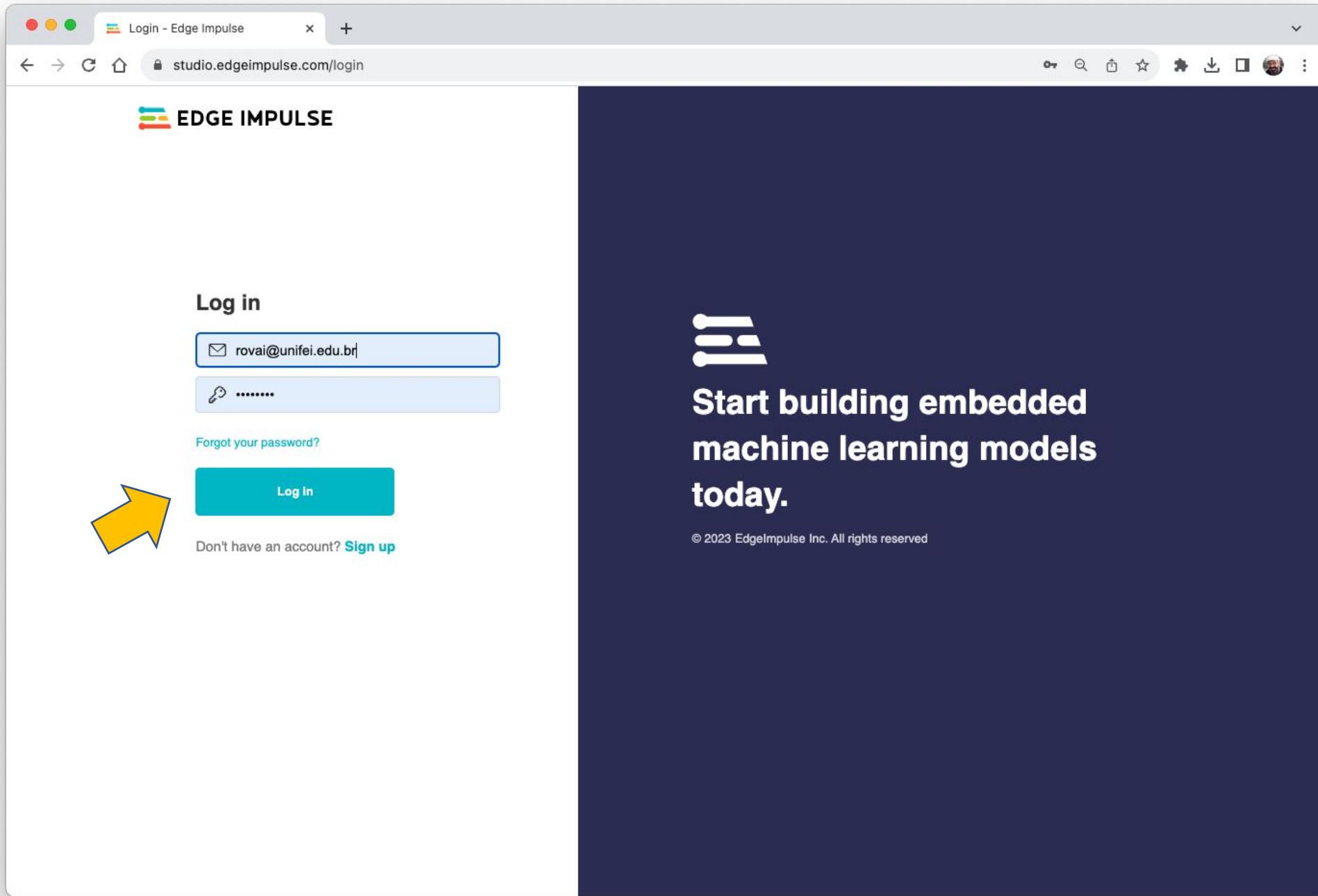
Build datasets, train models, and optimize libraries to run on any edge device, from extremely low-power MCUs to efficient Linux CPU targets and GPUs.

[Get Started](#) [Schedule a demo](#)

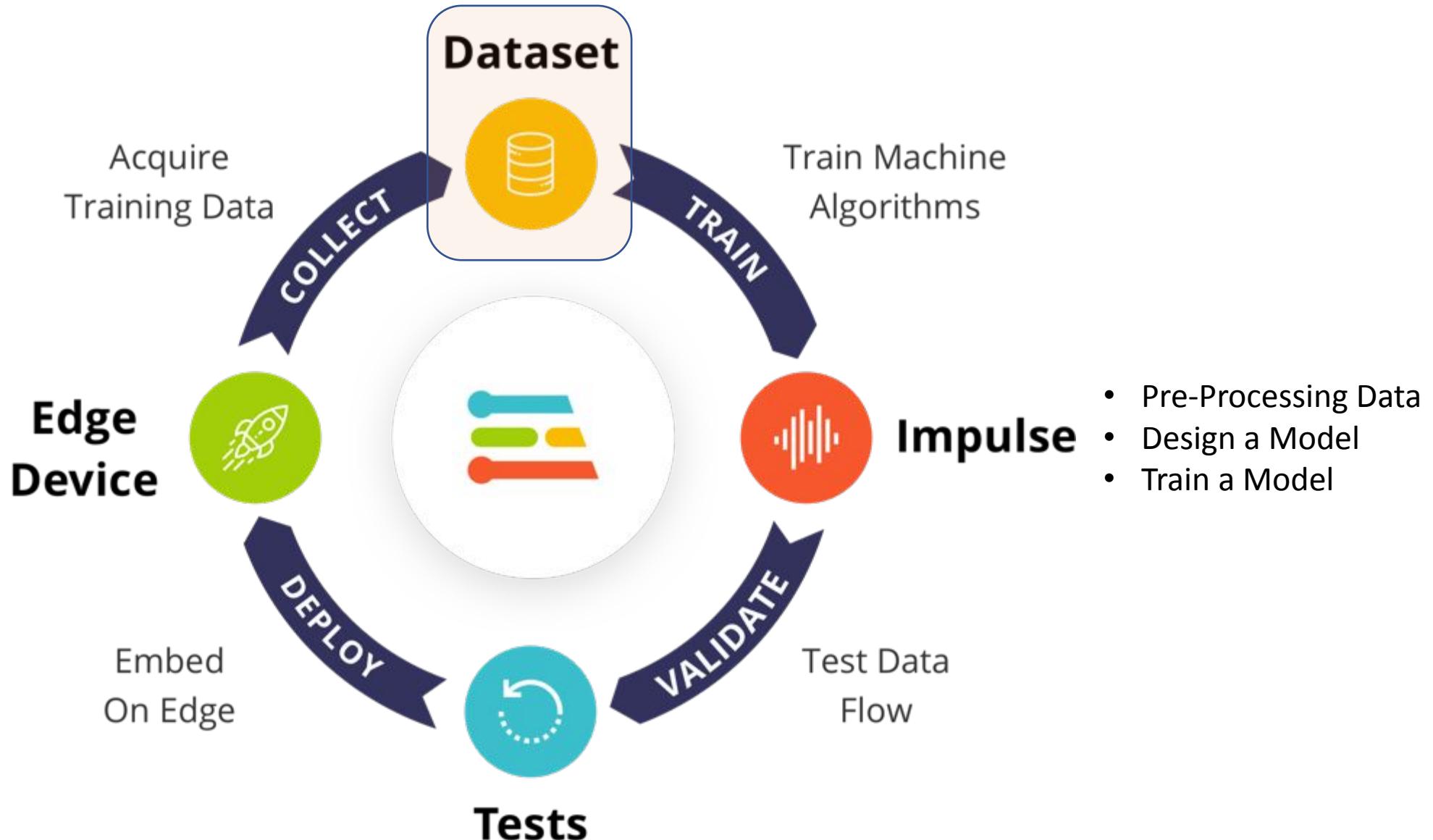
Webinar November 9th: 'Fast Track AI to the Edge with NVIDIA and Edge Impulse' [Register here](#)

EDGE IMPULSE



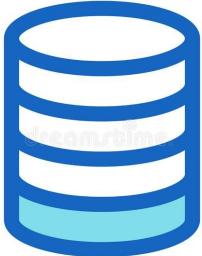


The screenshot shows the Edge Impulse studio interface. On the left, there's a sidebar with a user profile picture of a man with a beard, the name "MJRoBot", and the subtitle "(Marcelo Rovai)". Below this is a section titled "Organizations" with a purple button labeled "EIE". The main area has tabs for "Projects" (which is selected) and "Custom ML blocks". A modal window titled "Create a new project" is open. It contains a text input field with the placeholder "Enter the name for your new project:" and the value "Cifar10_Image_Classification" highlighted with an orange border. Below the input field is a section titled "Choose your project type:" with two options: "Developer" (selected, indicated by a blue radio button) and "Enterprise" (indicated by an empty radio button). Under "Developer", there's a note: "20 min job limit, 4GB or 4 hours of data, limited collaboration." Under "Enterprise", there's a note: "No job or data size limits, higher performance, custom blocks." A dropdown menu "Create under organization:" shows "Edge Impulse Experts". At the bottom of the modal is a green "Create new project" button. To the right of the modal, a yellow arrow points upwards towards the "Create new project" button. The background shows a list of existing projects: "MJRoBot (Marcelo Rovai) / video_tinyml_raw", "MJRoBot (Marcelo Rovai) / Pico_Motion_Detection" (PUBLIC), and "MJRoBot (Marcelo Rovai) / Loi_kouic_low_meetups".

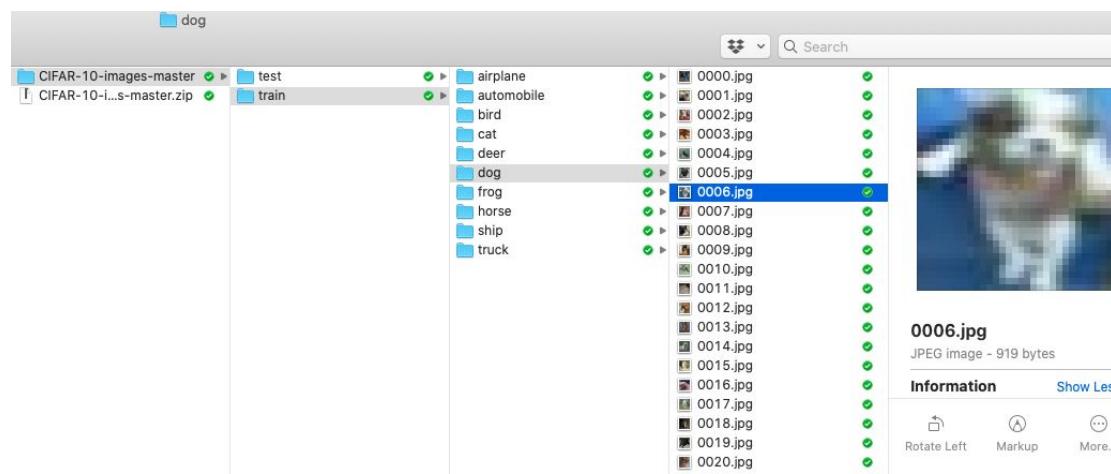


Download Dataset

<https://github.com/YoongiKim/CIFAR-10-images>



The screenshot shows a GitHub repository page for 'CIFAR-10-images'. At the top, there are buttons for 'Code' (highlighted with an orange box), 'Go to file', and 'Add file'. Below the buttons, there's a 'Clone' section with 'HTTPS' selected, showing the URL 'https://github.com/YoongiKim/CIFAR-10-images'. There are also links for 'Open with GitHub Desktop' and 'Download ZIP'. A large yellow arrow points from the monitor icon at the bottom to this 'Download ZIP' button. The main content area displays the repository's README.md file, which contains the text 'CIFAR-10 raw jpeg images' and 'You can just clone this repository to use dataset.'



Collect Data

Upload data - Cifar10_Image_C

stUDIO.edgeimpulse.com/studio/51070/upload

EDGE IMPULSE

Dashboard

Devices

Data acquisition

Impulse design

- Create impulse
- Image
- NN Classifier

EON Tuner

Retrain model

Live classification

Model testing

Versioning

Deployment

GETTING STARTED

Documentation

Forums

UPLOAD DATA (CIFAR10_IMAGE_CLASSIFICATION)

Upload existing data

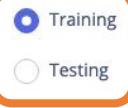
You can upload existing data to your project in the [Data Acquisition Format](#) (CBOR, JSON, CSV), or as WAV, JPG or PNG files.

Select files 

No file chosen

Upload into category

Automatically split between training and testing ?

Training 

Testing

Label

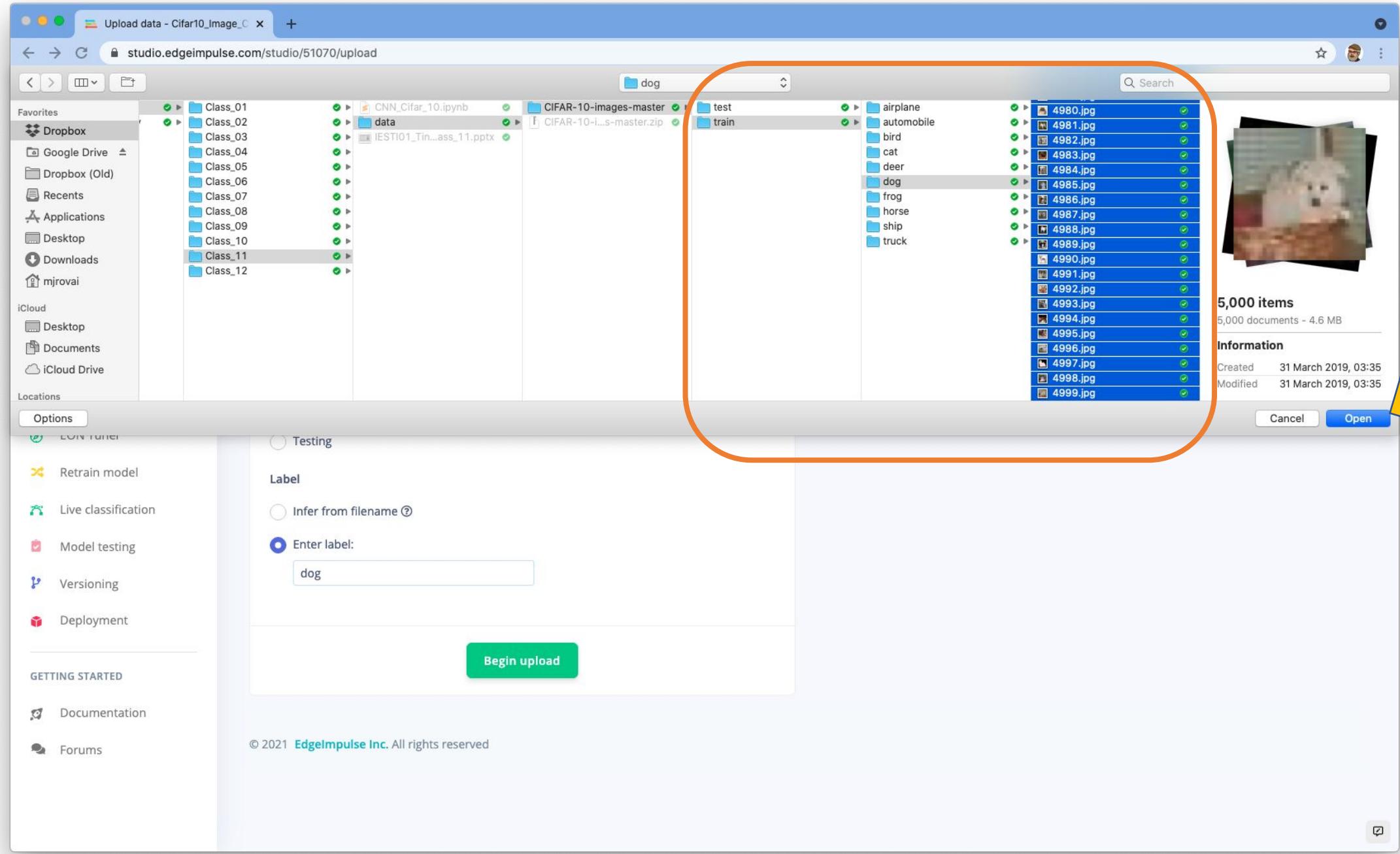
Infer from filename ?

Enter label: 

dog

Begin upload

MJRoBot (Marcelo Rovai)



Collect Data

Upload data - Cifar10_Image_C

stUDIO.edgeimpulse.com/studio/51070/upload

EDGE IMPULSE

MJRoBot (Marcelo Rovai)

Dashboard

Devices

Data acquisition

Impulse design

- Create impulse
- Image
- NN Classifier

EON Tuner

Retrain model

Live classification

Model testing

Versioning

Deployment

GETTING STARTED

Documentation

Forums

UPLOAD DATA (CIFAR10_IMAGE_CLASSIFICATION)

Upload existing data

You can upload existing data to your project in the [Data Acquisition Format](#) (CBOR, JSON, CSV), or as WAV, JPG or PNG files.

Select files

5000 files

Upload into category

Automatically split between training and testing ?

Training

Testing

Label

Infer from filename ?

Enter label:

dog

The screenshot shows the Edge Impulse web interface for dataset management. A blue arrow on the left points to the 'Data acquisition' section of the sidebar, labeled 'Collect Data'. A yellow arrow at the top points to the browser address bar, which displays the URL studio.edgeimpulse.com/studio/51070/acquisition/training?page=1#. The main interface has a purple header with tabs for 'Dataset' (selected), 'Data explorer', 'Data sources', and 'CSV Wizard'. The 'Dataset' tab displays summary statistics: 'DATA COLLECTED 12,002 items' (with a red pie chart icon) and 'TRAIN / TEST SPLIT 83% / 17%' (with a green pie chart icon). An orange box highlights the 'Dataset' tab and its summary. To the right, a large callout box titled 'Collect data' contains the text 'Connect a device to start building your dataset.' Below the summary, there's a 'RAW DATA' section for a sample named '4999.jpg.2hfb34u0', showing a small image of a dog. The 'Dataset' table lists four samples, all labeled 'dog':

	SAMPLE NAME	LABEL	ADDED
<input type="checkbox"/>	4999.jpg.2hfb34u0	dog	Oct 09 2021, 11:1...
<input type="checkbox"/>	4997.jpg.2hfb34n1	dog	Oct 09 2021, 11:1...
<input type="checkbox"/>	4991.jpg.2hfb34mm	dog	Oct 09 2021, 11:1...
<input type="checkbox"/>	4993.jpg.2hfb34n8	dog	Oct 09 2021, 11:1...

The screenshot shows the Edge Impulse web studio interface. The left sidebar contains a navigation menu with various project management and development tools. The main content area is titled "Data explorer" and provides instructions on how to generate a data explorer for a Cifar10_Image_Classification project. It lists three options: "Using a pretrained visual model" (selected), "Using your trained impulse", and "Using the preprocessing blocks in your impulse". Below this, it discusses dimensionality reduction techniques, specifically "t-SNE" and "PCA". A green button at the bottom right says "Generate data explorer".

Cifar10_Image_Classification

MJRoBot (Marcelo Rovai) / Cifar10_Image_Classification

EDGE IMPULSE

Dataset Data explorer Data sources | CSV Wizard

Data explorer

The data explorer shows a complete view of all data in your project. Use it to quickly label your data, or spot outliers. [Learn more](#).

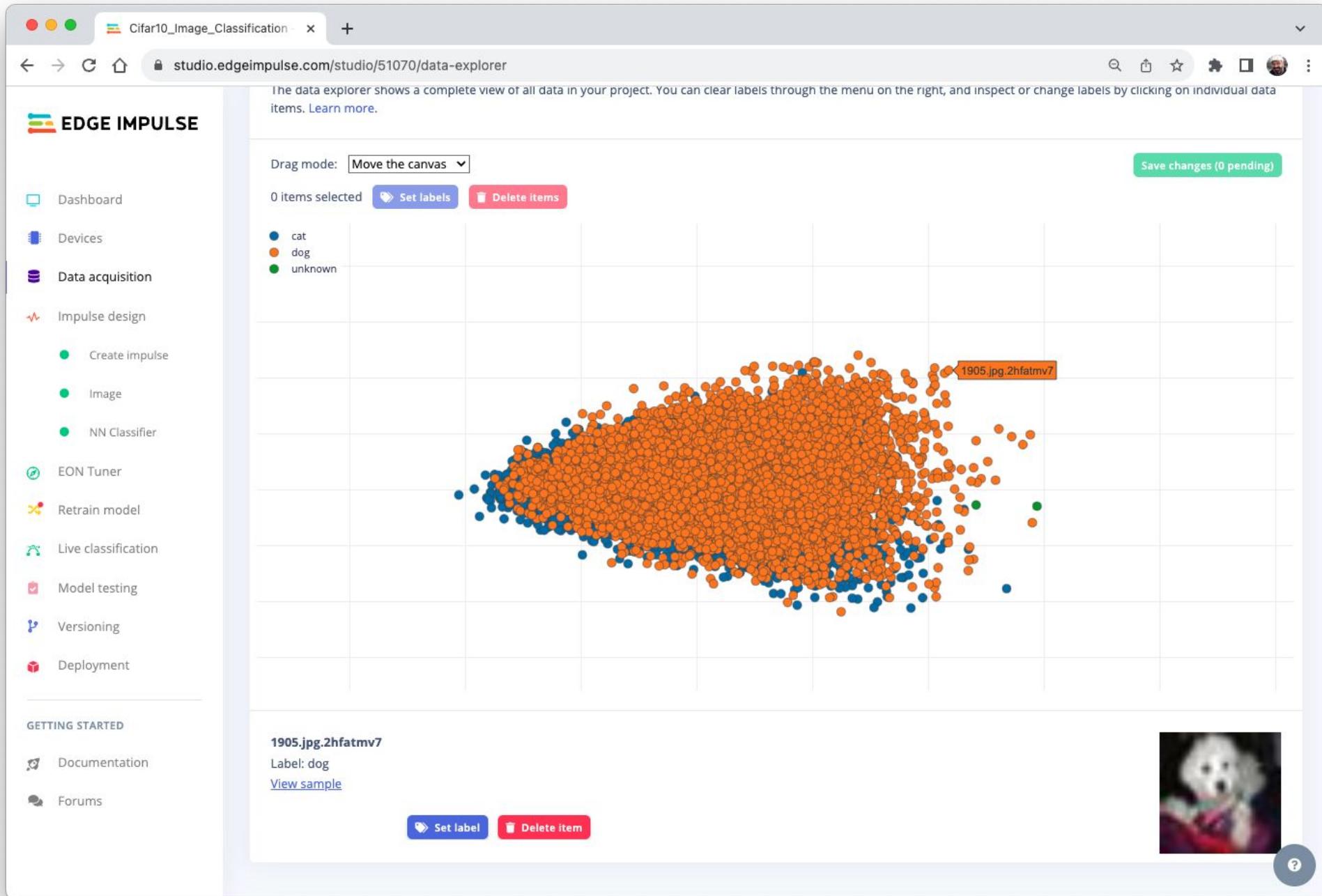
How should we generate the data explorer?

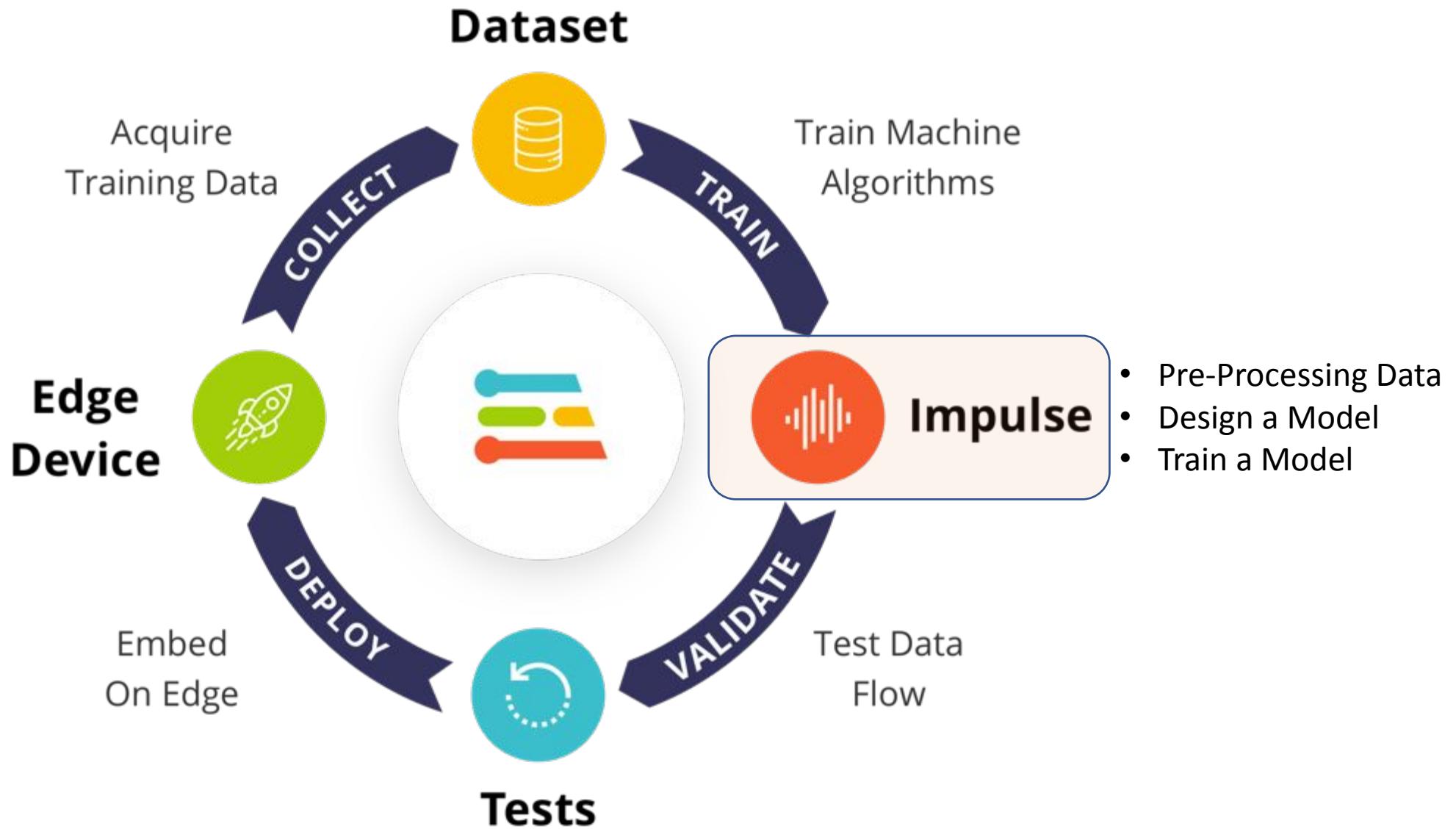
- Using a pretrained visual model
Great for most image classification projects.
- Using your trained impulse
Works great if you have collected some labeled data already and have a trained model.
- Using the preprocessing blocks in your impulse
Use this if you don't have any labels for your data yet, and thus can't train a full model.

Dimensionality reduction technique

- t-SNE
Separates best, but takes a significant amount of time on large datasets.
- PCA
Recommended for your dataset. Separates less well, but works on any dataset size.

Generate data explorer





Create impulse - Cifar10_Image

studio.edgeimpulse.com/studio/51070/create-impulse

EDGE IMPULSE

CREATE IMPULSE (CIFAR10_IMAGE_CLASSIFICATION)

An impulse takes raw data, uses signal processing to extract features, and then uses a learning block to classify new data.

Image data

Axes

Image width: 32 Image height: 32

Resize mode: Fit longest axis

For optimal accuracy with transfer learning blocks, use a 96x96 or 160x160 image size.

Image

Name: Image

Input axes: image

Classification (Keras)

Name: NN Classifier

Input features: Image

Output features: 2 (cat, dog)

Output features

2 (cat, dog)

Save impulse

Image data

Image

Classification (Keras)

Output features

Save impulse

Add a processing block

Add a learning block

© 2021 EdgeImpulse Inc. All rights reserved

MJRoBot (Marcelo Rovai)

Preprocess Data

Image - Cifar10_Image_Classifi x +

studio.edgeimpulse.com/studio/51070/dsp/image/3

EDGE IMPULSE

IMAGE (CIFAR10_IMAGE_CLASSIFICATION)

#1 Click to set a description for this version

Parameters Generate features

Raw data

4999.jpg.2hfb34u0 (dog)

Raw features

0xa9aaa4, 0xafb0aa, 0xabaca6, 0xa9aaa4, 0xaebl8, 0xabaee5, 0xa6a9a0, 0xabaee5, 0xabaee3, 0xaaad...

Parameters

Image

Color depth

RGB

Save parameters

DSP result

Image

Processed features

0.6627, 0.6667, 0.6431, 0.6863, 0.6902, 0.6667, 0.6706, 0.6745, 0.6510, 0.6627, 0.6667, 0.6431, ...

On-device performance

PROCESSING TIME 9 ms.

PEAK RAM USAGE 4 KB

© 2021 EdgeImpulse Inc. All rights reserved

A large yellow arrow points upwards from the "Save parameters" button towards the "Color depth" dropdown menu, which is highlighted with an orange border.

Preprocess Data

Image - Cifar10_Image_Classifi x +

studio.edgeimpulse.com/studio/51070/dsp/image/3/generate-features

EDGE IMPULSE

IMAGE (CIFAR10_IMAGE_CLASSIFICATION)
#1 Click to set a description for this version

Parameters Generate features

Training set

Data in training set 10,000 items

Classes 2 (cat, dog)

Generate features

Feature explorer (10,000 samples) ②

X Axis Y Axis Z Axis

Visualization layer 1 Visualization layer 2 Visualization layer 3

cat dog

8.5
8
7.5
7
6.5
6
5.5
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0

Visualization layer 3
Visualization layer 2
Visualization layer 1

On-device performance ②

PROCESSING TIME 9 ms.

PEAK RAM USAGE 4 KB

© 2021 EdgeImpulse Inc. All rights reserved

MJRoBot (Marcelo Rovai)

colab.research.google.com

CNN_Cifar_10.ipynb

File Edit View Insert Runtime Tools Help

+ Code + Text Copy to Drive Connect GPU

10 plt.show()

model accuracy

accuracy

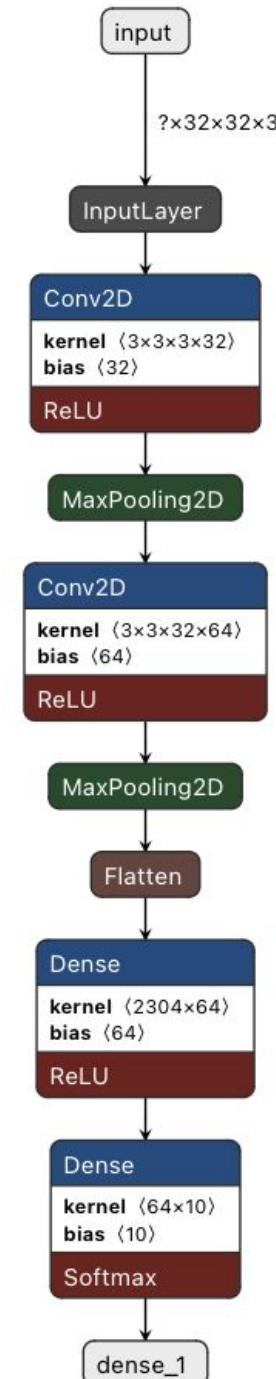
epoch

{x}

Evaluate Model

```
[ ] 1 print("Train data: ", model.evaluate(train_images, train_labels)[1])
2 print("Eval data: ", model.evaluate(val_images, val_labels)[1])
3 print("Test data: ", model.evaluate(test_images, test_labels)[1])
```

1407/1407 [=====] - 6s 4ms/step - loss: 0.2313 - accuracy: 0.9199
Train data: 0.9198889136314392
157/157 [=====] - 1s 4ms/step - loss: 1.2736 - accuracy: 0.6920
Eval data: 0.6919999718666077
313/313 [=====] - 1s 4ms/step - loss: 1.3474 - accuracy: 0.6807
Test data: 0.6807000041007996



The screenshot shows the Edge Impulse Studio interface for a Cifar10_Image_Classification project. The left sidebar includes links for Dashboard, Devices, Data acquisition, Impulse design (Create impulse, Image, NN Classifier), EON Tuner, Retrain model, Live classification, Model testing, Versioning, Deployment, Documentation, and Forums. A yellow arrow on the left points to the 'Design a Model' section.

Design a Model

Training settings

- Number of training cycles: 10
- Learning rate: 0.0005

Advanced training settings

- Validation set size: 20%
- Batch size: 32
- Auto-balance dataset:
- Profile int8 model:

Neural network architecture

- Input layer (3,072 features)
- 2D conv / pool layer (32 filters, 3 kernel size, 1 layer)
- 2D conv / pool layer (64 filters, 3 kernel size, 1 layer)
- Flatten layer
- Dense layer (64 neurons)
- Add an extra layer
- Output layer (2 classes)

Start training

Training output (CPU) (0)

Model

Last training performance (validation set)

- ACCURACY: 74.4%
- LOSS: 0.56

Confusion matrix (validation set)

	CAT	DOG
CAT	80.6%	19.4%
DOG	32.1%	67.9%
F1 SCORE	0.76	0.72

Feature explorer (full training set)

Legend: cat - correct (green), dog - correct (blue), cat - incorrect (red), dog - incorrect (orange)

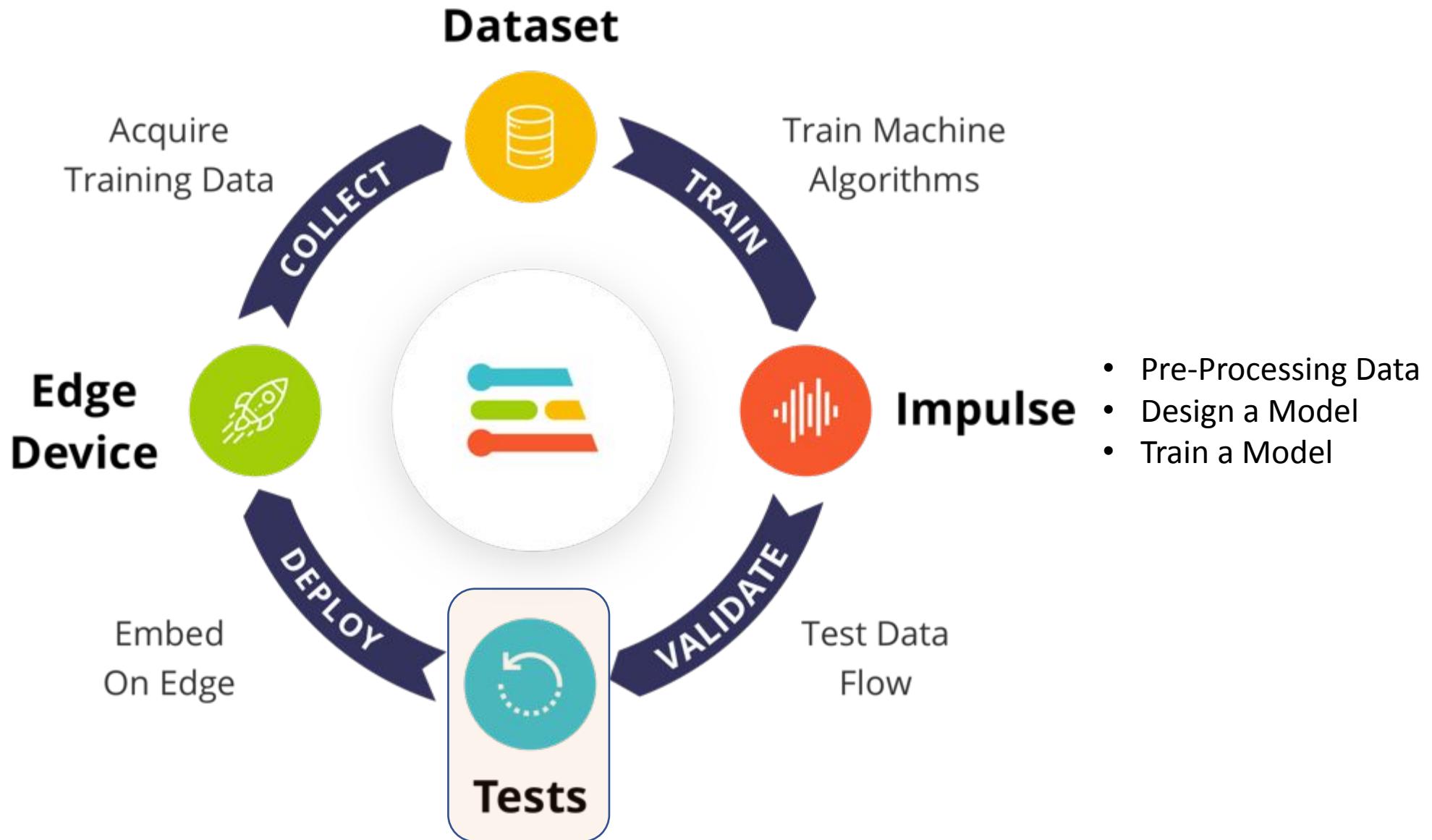
On-device performance

- INFERENCE TIME: 93 ms.
- PEAK RAM USAGE: 44.7K
- FLASH USAGE: 308.2K

Train a Model

Target: Arduino Nicla Vision (Cortex-M7 480MHz)

Model version: Quantized (int8)



Model testing - Cifar10_Image... Upload data - Cifar10 Classific... | +

studio.edgeimpulse.com/studio/51070/validation

EDGE IMPULSE

- Dashboard
- Devices
- Data acquisition
- Impulse design
 - Create impulse
 - Image
 - NN Classifier
- EON Tuner
- Retrain model
- Live classification
 - Model testing
 - Versioning
- Deployment

Evaluate Optimize

Test data

Set the 'expected outcome' for each sample to the desired outcome to automatically score the impulse.

SAMPLE NAME	EXPECTED OUTCOME	LENGTH	ACCURACY	RESULT	⋮
testing.2hfe5uat	testing	-		1 cat	⋮
0999.jpg.2hfb7a...	cat	-	100%	1 cat	⋮
0997.jpg.2hfb7a...	cat	-	100%	1 cat	⋮
0998.jpg.2hfb7a...	cat	-	100%	1 cat	⋮
0996.jpg.2hfb7a...	cat	-	0%	1 uncertain	⋮
0993.jpg.2hfb7a...	cat	-	100%	1 cat	⋮
0995.jpg.2hfb7a...	cat	-	0%	1 dog	⋮
0991.jpg.2hfb7a...	cat	-	100%	1 cat	⋮
0994.jpg.2hfb7a...	cat	-	100%	1 cat	⋮
0992.jpg.2hfb7a...	cat	-	100%	1 cat	⋮
0990.jpg.2hfb79...	cat	-	100%	1 cat	⋮
0988.jpg.2hfb79...	cat	-	100%	1 cat	⋮
0985.jpg.2hfb79...	cat	-	0%	1 uncertain	⋮

Model testing output

Classifying data for NN classifier...
 Copying features from processing blocks...
 Copying features from DSP block...
 Copying features from DSP block OK
 Copying features from processing blocks OK

Classifying data for float32 model...
 Scheduling job in cluster...
 Job started
 Classifying data for NN Classifier OK

Job completed

Model testing results

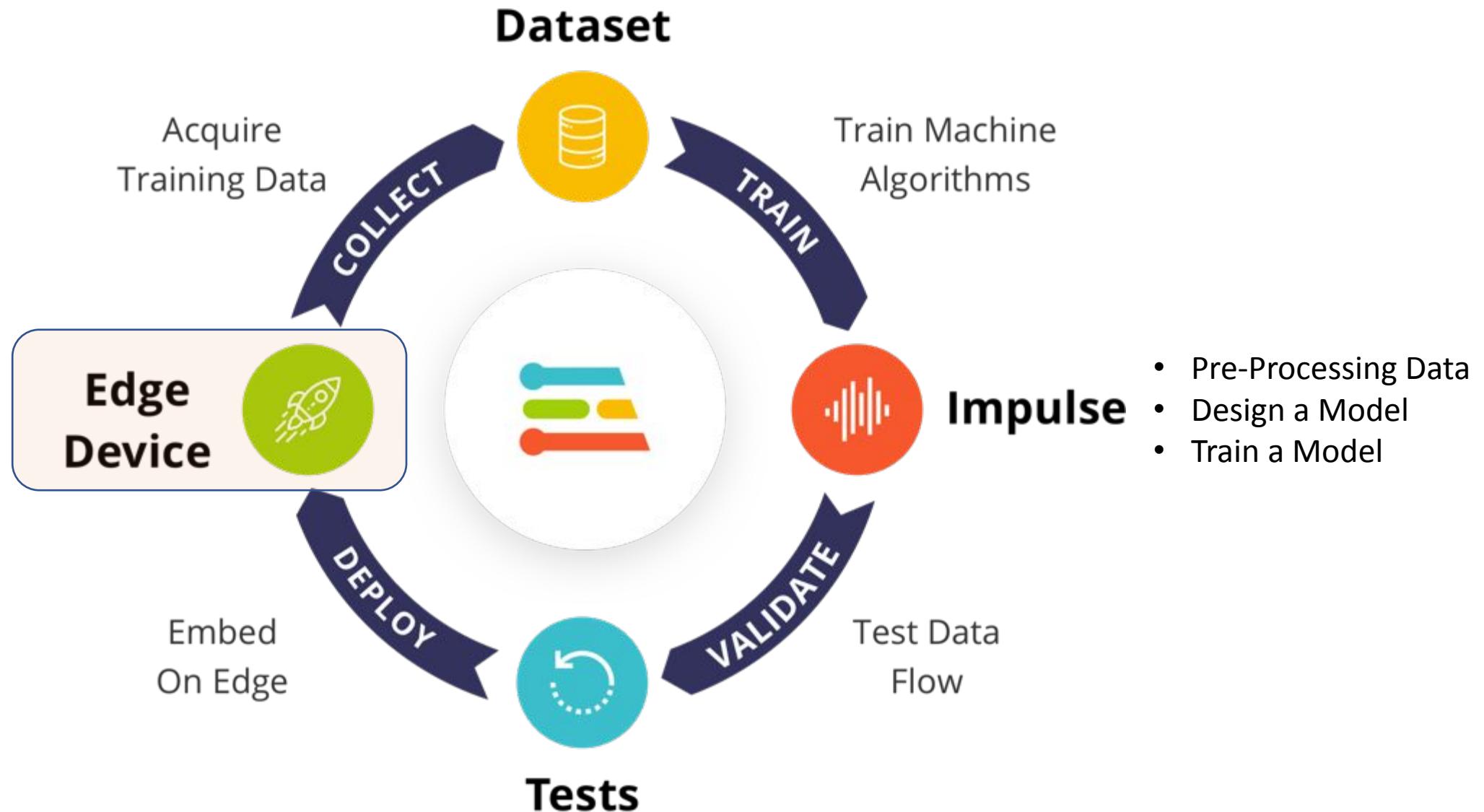
ACCURACY **64.95%** %

	CAT	DOG	UNCERTAIN
CAT	67.7%	16%	16.3%
DOG	21.2%	62.2%	16.6%
F1 SCORE	0.72	0.70	

Feature explorer ⓘ

Legend:

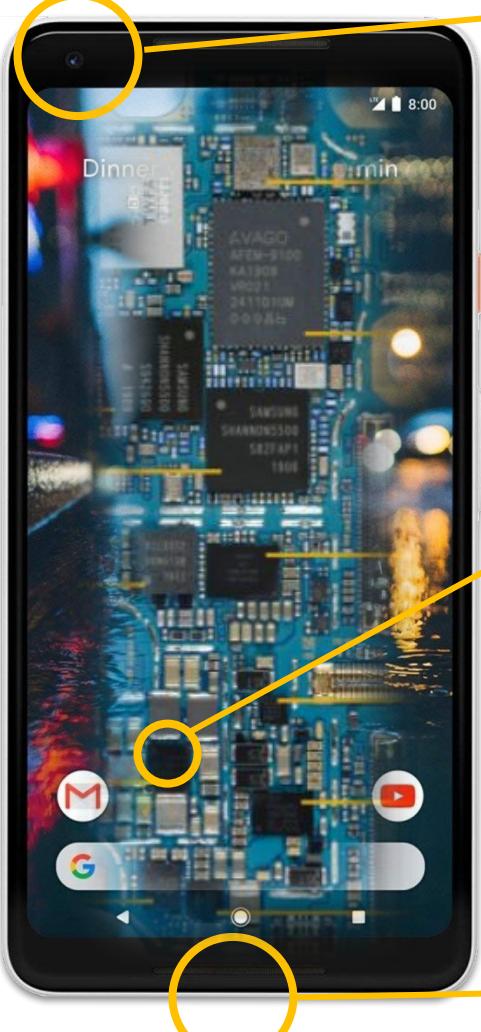
- cat - correct
- dog - correct
- cat - incorrect
- dog - incorrect
- testing



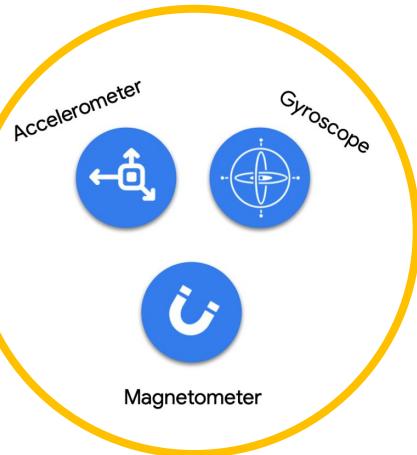
Edge Device



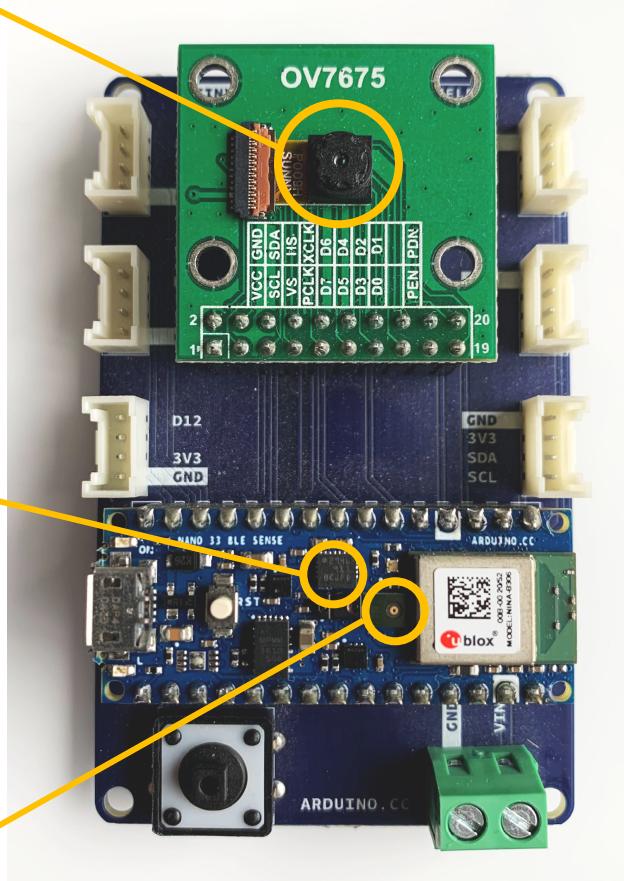
& Sensors



Camera



Microphone



Deploy Model

The screenshot shows the Edge Impulse studio interface for deploying a Cifar10_Image_Classification model. The left sidebar includes links for Dashboard, Devices, Data acquisition, Impulse design (Create impulse, Image, NN Classifier), EON Tuner, Retrain model, Live classification, Model testing, Versioning, Deployment, Documentation, and Forums.

The main area is titled "Configure your deployment". It explains that the model can be deployed to any device, running without an internet connection, minimizing latency, and with minimal power consumption. A "Search deployment options" input field is present.

DEFAULT DEPLOYMENT: C++ library
A portable C++ library with no external dependencies, which can be compiled with any modern C++ compiler.

MODEL OPTIMIZATIONS: Model optimizations can increase on-device performance but may reduce accuracy.

Enable EON™ Compiler: Same accuracy, up to 50% less memory. [Learn more](#)

Quantized (int8) (Selected):

	IMAGE	NN CLASSIFIER	TOTAL
LATENCY	1 ms.	93 ms.	94 ms.
RAM	4.0K	44.7K	44.7K
FLASH	-	308.2K	-
ACCURACY			-

Unoptimized (float32) (Select):

	IMAGE	NN CLASSIFIER	TOTAL
LATENCY	1 ms.	206 ms.	207 ms.
RAM	4.0K	163.4K	163.4K
FLASH	-	1.1M	-
ACCURACY			64.95%

To compare model accuracy, run model testing for all available optimizations. [Run model testing](#)

Estimate for Arduino Nida Vision (Cortex-M7 480MHz) - Change target

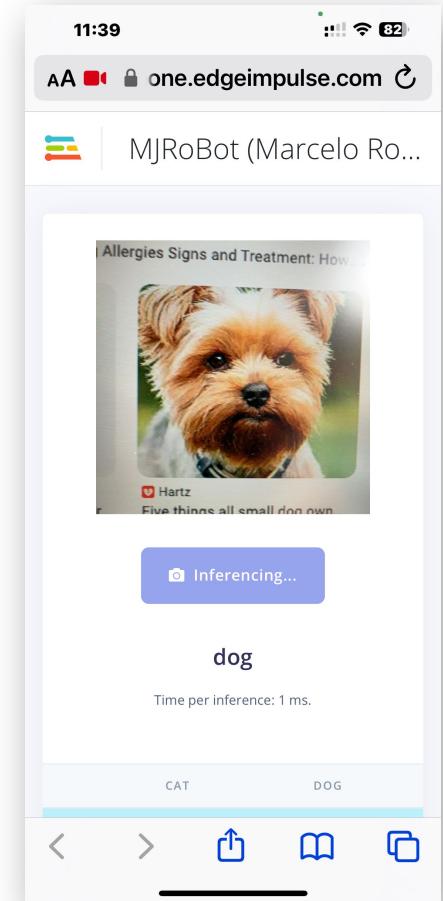
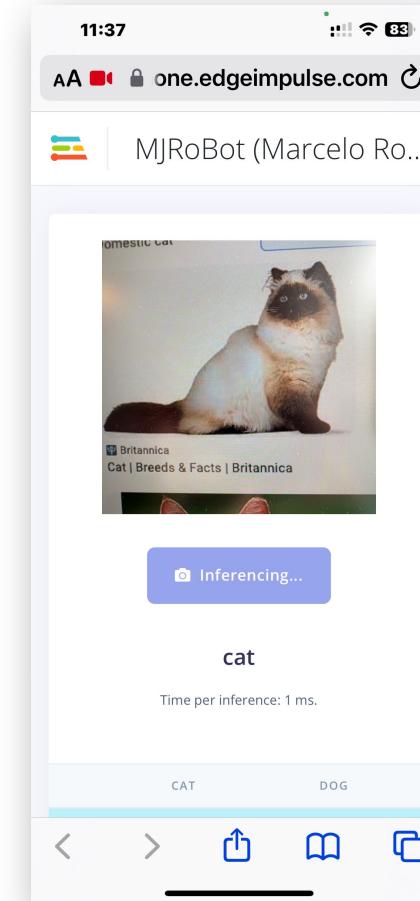
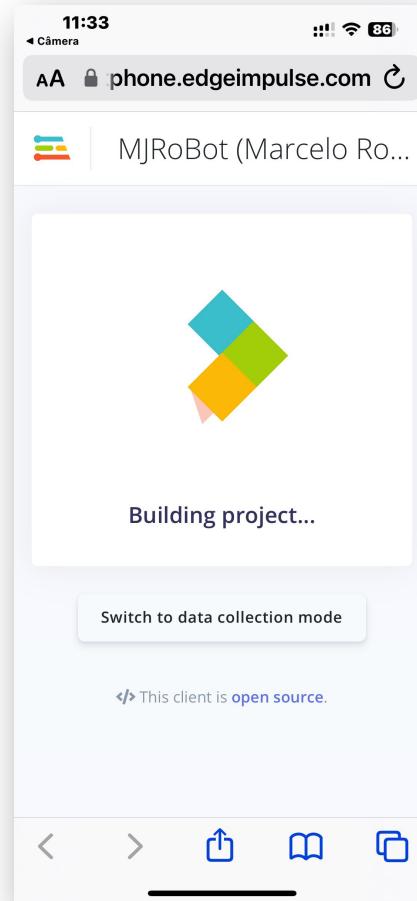
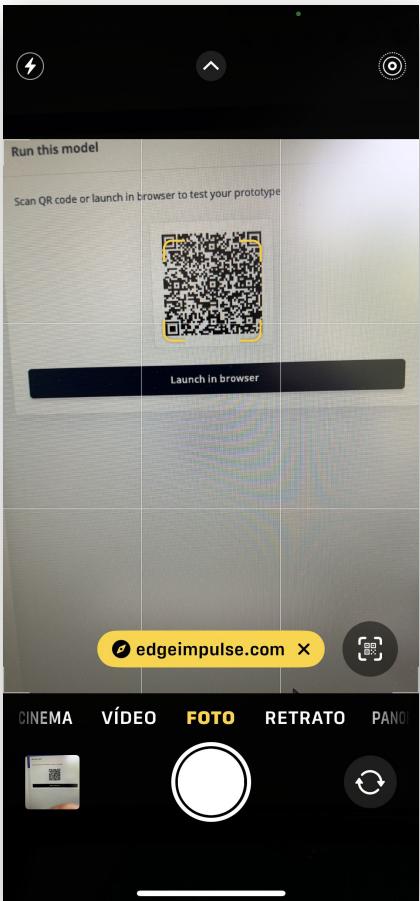
Build

Run this model: Scan QR code or launch in browser to test your prototype

A QR code is displayed with the instruction: "Scan QR code or launch in browser to test your prototype". Below it is a "Launch in browser" button.

A photograph shows a hand holding a smartphone. The screen displays the Edge Impulse website with the same QR code and deployment options visible. In the background, a computer monitor also shows the Edge Impulse interface.

Make Inferences



EI Studio “under the hood” Code Time!

ei-cifar10_image_classification-nn-classifier-explained.ipynb



EDGE IMPULSE

NN CLASSIFIER (CIFAR10_IMAGE_CLASSIFICATION)

#2 ▾ Model - Same as done in class (IESTI01) ★ Primary version

Neural Network settings

Training settings

Neural network architecture

```

1 import tensorflow as tf
2 from tensorflow.keras.models import Sequential
3 from tensorflow.keras.layers import Dense, InputLayer, Dropout, Conv2D, Flatten,
4     Reshape, MaxPooling1D, MaxPooling2D, BatchNormalization, TimeDistributed
5
6 # model architecture
7 model = Sequential()
8 model.add(Conv2D(32, kernel_size=3, activation='relu', kernel_constraint=tf.keras
    .constraints.MaxNorm(1), padding='same'))
9 model.add(MaxPooling2D(pool_size=2, strides=2, padding='same'))
10 model.add(Conv2D(64, kernel_size=3, activation='relu', kernel_constraint=tf.keras
    .constraints.MaxNorm(1), padding='same'))
11 model.add(MaxPooling2D(pool_size=2, strides=2, padding='same'))
12 model.add(Flatten())
13 model.add(Dense(64, activation='relu',
    activity_regularizer=tf.keras.regularizers.l1(0.0001)))
14 model.add(Dense(classes, activation='softmax', name='y_pred'))
15
16 # this controls the learning rate
17 opt = Adam(lr=0.0005, beta_1=0.9, beta_2=0.999)
18 # this controls the batch size, or you can manipulate the tf.data.Dataset objects
19 # yourself
20 BATCH_SIZE = 32
21 train_dataset = train_dataset.batch(BATCH_SIZE, drop_remainder=False)
22 validation_dataset = validation_dataset.batch(BATCH_SIZE, drop_remainder=False)
23 callbacks.append(BatchLoggerCallback(BATCH_SIZE, train_sample_count))
24
25 # train the neural network
26 model.compile(loss='categorical_crossentropy', optimizer=opt, metrics=['accuracy'])

```

Start training

Training output

Model

Model version: ⓘ Quantized (int8)

Last training performance (validation set)

ACCURACY 74.4% LOSS 0.56

Confusion matrix (validation set)

	CAT	DOG
CAT	80.6%	19.4%
DOG	32.1%	67.9%
F1 SCORE	0.76	0.72

Feature explorer (full training set) ⓘ

- cat - correct
- dog - correct
- cat - incorrect
- dog - incorrect

On-device performance ⓘ

INFERENCING TIME 1,048 ms. PEAK RAM USAGE 44.7K FLASH USAGE 308.2K

<https://studio.edgeimpulse.com/studio/51070/learning/keras/6#>

Connect device for data capture
Mobile Phone

The screenshot shows the Edge Impulse Studio interface. On the left, there's a sidebar with various options like Dashboard, Devices, Data acquisition, etc. The 'Devices' option is highlighted with an orange border and has an orange arrow pointing to it from below. The main area shows a 'Collect new data' modal with three options: 'Scan QR code to connect to your phone' (which is also highlighted with an orange border), 'Connect to your computer', and 'Connect your device or development board'. The URL in the browser bar is studio.edgeimpulse.com/studio/51070/devices#collect-new-data.

The image displays the Edge Impulse studio interface across three devices: a desktop browser and two mobile phones.

Desktop Browser (Left):

- The URL is `studio.edgeimpulse.com/studio/51070/devices`.
- The sidebar on the left has a highlighted "Devices" option.
- A modal window titled "Collect new data" is open, showing a green checkmark icon and the message "Device 'phone_lorah242' is now connected".
- Below the modal, there's a "Get started!" button.

Mobile Phone (Top Right):

- The screen shows a list of connected devices.
- The device "phone_lorah242" is listed with a status of "Connected" and last seen at "Today, 11:53:13".
- The main area says "Data collection" with a large green checkmark icon.
- Text below says: "Connected as phone_lorah242. You can collect data from this device from the Data acquisition page in the Edge Impulse studio."
- Two buttons are highlighted with orange boxes:
 - "Collecting images?" (with a camera icon)
 - "Collecting audio?" (with a microphone icon)
- At the bottom are standard navigation icons.

Mobile Phone (Bottom Right):

- The screen shows a camera view with a red "AA" battery icon.
- The URL is `one.edgeimpulse.com`.
- The main area says "Data collection".
- A blue bar at the top indicates the category is "Training".
- A large image of an orange and white cat is displayed.
- A "Capture" button is visible at the bottom.
- At the bottom are standard navigation icons.

Reading Material

Main references

- [Harvard School of Engineering and Applied Sciences - CS249r: Tiny Machine Learning](#)
- [Professional Certificate in Tiny Machine Learning \(TinyML\) – edX/Harvard](#)
- [Introduction to Embedded Machine Learning - Coursera/Edge Impulse](#)
- [Computer Vision with Embedded Machine Learning - Coursera/Edge Impulse](#)
- Fundamentals textbook: “[Deep Learning with Python](#)” by François Chollet
- Applications & Deploy textbook: “[TinyML](#)” by Pete Warden, Daniel Situnayake
- Deploy textbook “[TinyML Cookbook](#)” by Gian Marco Iodice

I want to thank **Shawn Hymel** and **Edge Impulse**, **Pete Warden** and **Laurence Moroney** from Google, Professor **Vijay Janapa Reddi** and **Brian Plancher** from Harvard, and the rest of the **TinyMLEdu** team for preparing the excellent material on TinyML that is the basis of this course at UNIFEI.

The IESTI01 course is part of the [TinyML4D](#), an initiative to make TinyML education available to everyone globally.

Thanks



UNIFEI