



# The Big Impact of TinyML: Embedded Machine Learning at the Extreme Edge



*Brian Plancher*  
*Barnard College, Columbia University*  
[brianplancher.com](http://brianplancher.com)



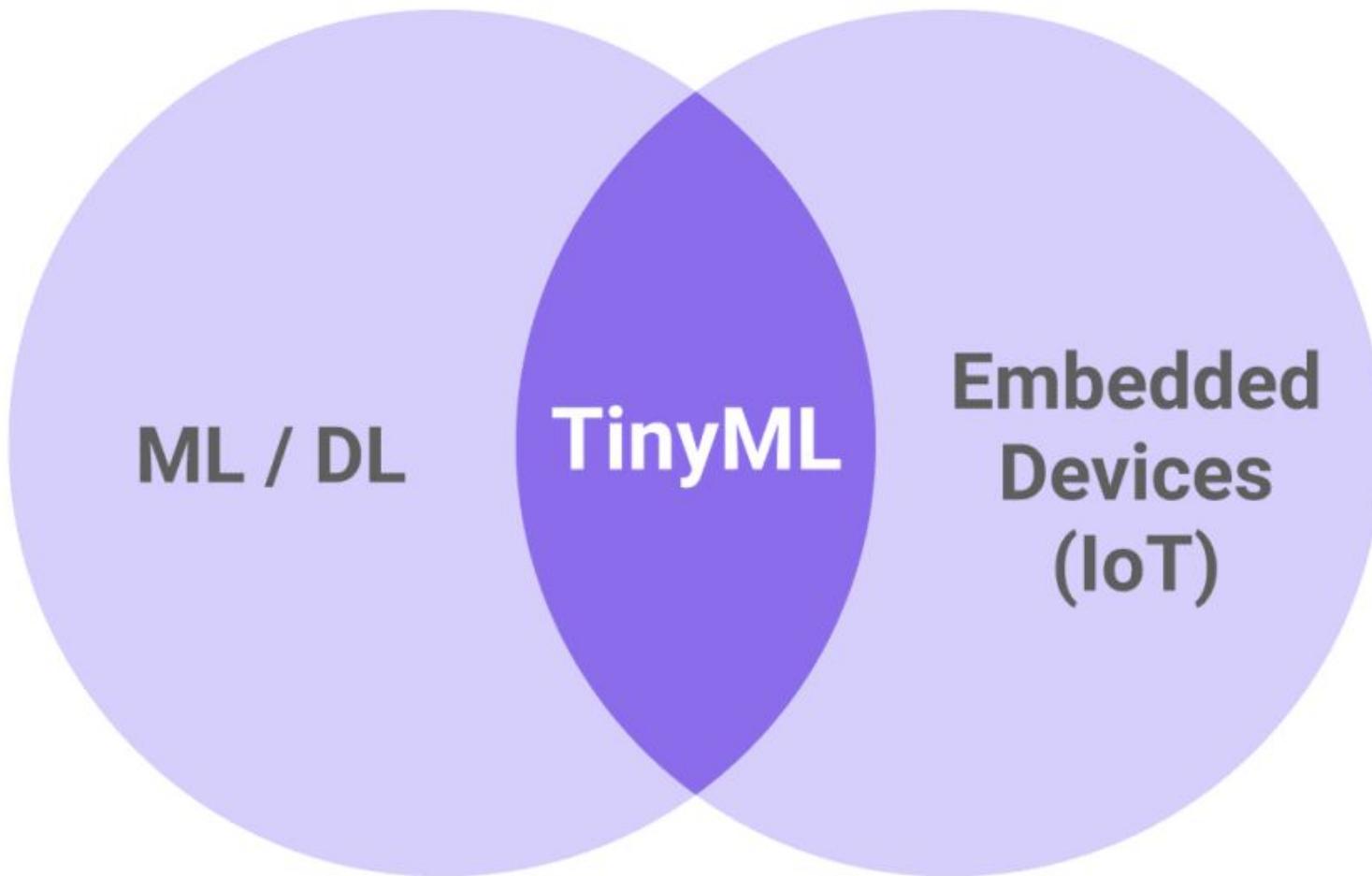
# The Big Impact of TinyML:

## Embedded Machine Learning at the Extreme Edge

Copyright (c) 2023 TinyMLedu.  
All rights reserved.  
CC BY-NC-SA 4.0

1. What is **Embedded (Tiny)** Machine Learning?
2. **Beneficial Applications** of Embedded Machine Learning
3. **Challenges** of Deploying ML at the Extreme Edge
4. **Sustainability & Privacy:** MLSensors at Scale

# What is Embedded (Tiny) Machine Learning?





# Beneficial Applications of Embedded Machine Learning

# Zero Hunger & Good Health and Well-Being (SDG #2 & #3)



Credit: PlantVillage Nuru

Nuru, an ML app more accurate than humans at detecting plant diseases. Increased a farmer's sales by 55% & **yields by 146%**.



Credit: Crop Angel Ltd

Tiny drones can provide targeted pesticide applications that **reduce use to 0.1%** of conventional blanket spraying.



Credit: Sinhyu/Getty Images

Using Edge Impulse, a system was prototyped to identify mosquitoes by wing beats sounds with **88.3% accuracy**.

# Zero Hunger & Good Health and Well-Being (SDG #2 & #3)



Mendeley Data

<https://data.mendeley.com/datasets/zp4nf2dxbh/1>

## Poultry Vocalization Signal Dataset for Early Disease Detection

Published: 12 June 2023 | Version 1 | DOI: 10.17632/zp4nf2dxbh.1

Contributors: Halleluyah Aworinde, Segun Adebayo, Akinwale Akinwunmi, Olufemi Alabi, Adebamiji Ayandiji, Olaide Oke, Abel Oyebamiji, Adetoye Adeyemo, Aderonke Sakpere, Kizito Echetama

### Description

The aim of generating this dataset is to deliver open, accessible and quality machine learning dataset for Poultry farm management. The dataset contains a total of 346 audio signal files which are divided into three folders: healthy, noise and unhealthy. The healthy folder contains 139 audio files, noise folder has 86, while unhealthy folder has 121 files. The audio files are saved in .wav file format.

Download All 869 MB



# Life on Land & Below Water

## (SDG #14 & #15)



Credit: Rainforest Connection

Rainforest Connection uses **recycled smartphones** for **solar-powered** listening devices to warn of **deforestation** efforts



Credit: RESOLVE and Bivash Pandav

RESOLVE's AI camera transmits notifications of elephant detection and can **run for more than 1.5 years** on a single battery.



Credit: Tim Cole

To prevent collisions with whales in busy waterways, Google deployed a TinyML model on hydrophones to alert ships.

# Life on Land & Below Water

## (SDG #14 & #15)

Open Access Article

### Design and Development of a Family of Integrated Devices to Monitor Animal Movement in the Wild

by  Laila Daniela Kazmierski <sup>1,\*†</sup> ,  Andrés Oliva Trevisan <sup>2,3,†</sup>,  Erika Kubisch <sup>4</sup>,  Karina Laneri <sup>1</sup>  and  Nicolás Catalano <sup>2,3</sup>

<sup>1</sup> Consejo Nacional de Investigaciones Científicas y Técnicas, Centro Atómico Bariloche (CONICET), Comisión Nacional de Energía Atómica (CNEA), San Carlos de Bariloche R8402AGP, Argentina

<sup>2</sup> Centro Atómico Bariloche, Comisión Nacional de Energía Atómica (CNEA), San Carlos de Bariloche R8402AGP, Argentina

<sup>3</sup> Instituto Balseiro, Universidad Nacional de Cuyo, Comisión Nacional de Energía Atómica (CNEA), San Carlos de Bariloche R8402AGP, Argentina

<sup>4</sup> Instituto de Investigaciones en Biodiversidad y Medioambiente, Consejo Nacional de Investigaciones Científicas y Técnicas (INIBIOMA, CONICET-Universidad Nacional del Comahue), San Carlos de Bariloche R8400AGP, Argentina

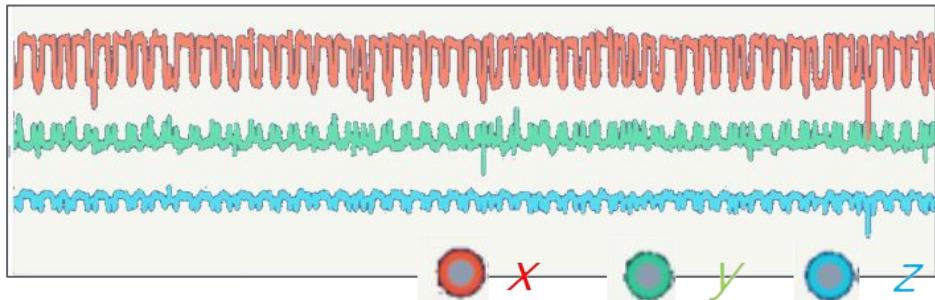
\* Author to whom correspondence should be addressed.

† These authors contributed equally to this work.

<https://www.mdpi.com/1424-8220/23/7/3684>



Example of  
accelerometer  
signal of a female  
digging a nest to  
lay eggs:

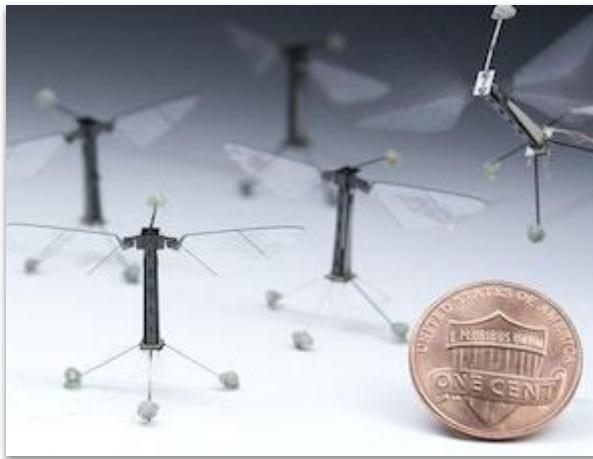


# Climate Action (SDG #13)



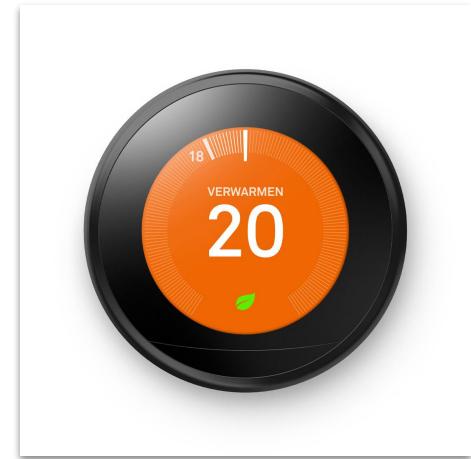
Credit: Ribbit Network

Ribbit Network is **crowdsourcing world's largest greenhouse gas emissions dataset** through distributed intelligent sensors



Credit: Wyss Institute at Harvard University

TinyML can help provide intelligence to **tiny robots like the Robobee** that can be used as artificial pollinators.



Credit: Google Nest

Smart HVAC systems show a **20-40% reduction in building energy usage**.

# And many more!



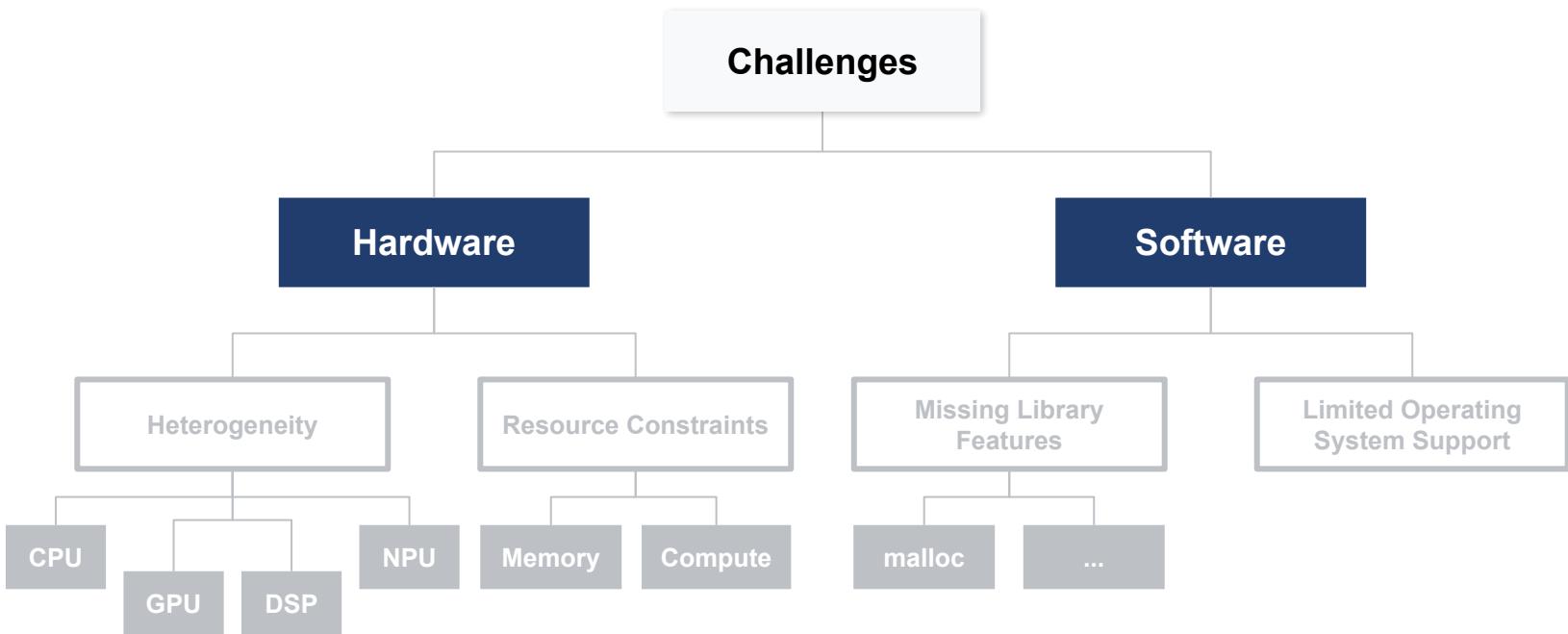
## Tiny Robot Learning: Challenges and Directions for Machine Learning in Resource-Constrained Robots

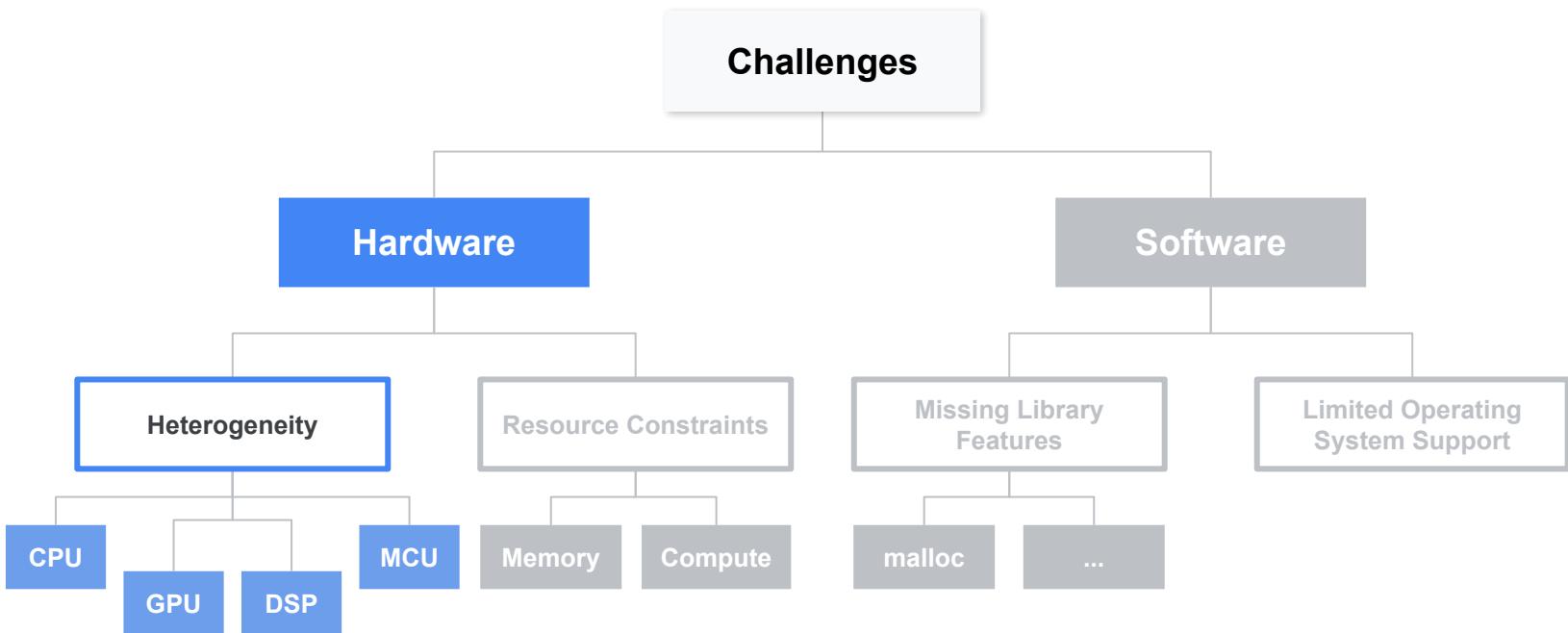
Sabrina M. Neuman<sup>1</sup>, Brian Plancher<sup>1</sup>, Bardienus P. Duisterhof<sup>2</sup>, Srivatsan Krishnan<sup>1</sup>, Colby Banbury<sup>1</sup>,  
Mark Mazumder<sup>1</sup>, Shvetank Prakash<sup>1</sup>, Jason Jabbour<sup>3</sup>, Aleksandra Faust<sup>4</sup>,  
Guido C.H.E. de Croon<sup>5</sup>, and Vijay Janapa Reddi<sup>1</sup>

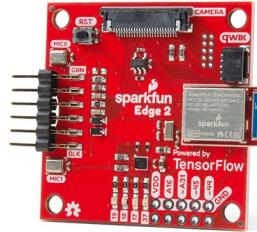
*Harvard University<sup>1</sup>, CMU<sup>2</sup>, University of Virginia<sup>3</sup>, Google Brain<sup>4</sup>, Delft University of Technology<sup>5</sup>*  
{sneuman@seas, brian\_plancher@g, srivatsan@g, cbanbury@g, markmazumder@g, sprakash@g, vj@eecs}.harvard.edu,  
bduister@andrew.cmu.edu, jjj4se@virginia.edu, aleksandra.faust@gmail.com, g.c.h.e.decroon@tudelft.nl

<https://arxiv.org/pdf/2205.05748.pdf>

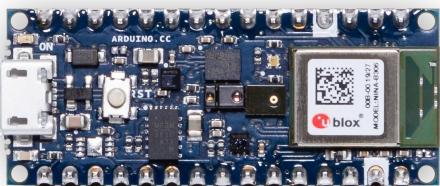
# Challenges of Deploying ML at the Extreme Edge

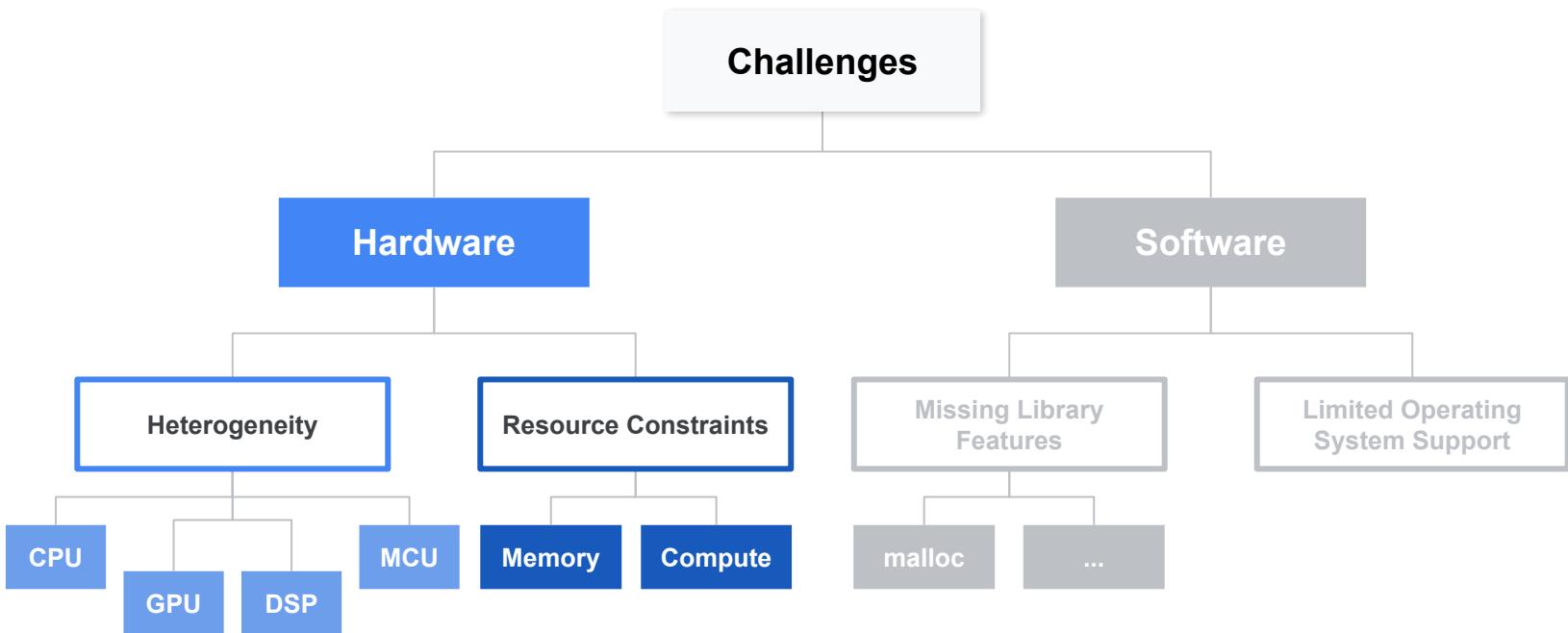






# 250 Billion MCUs today

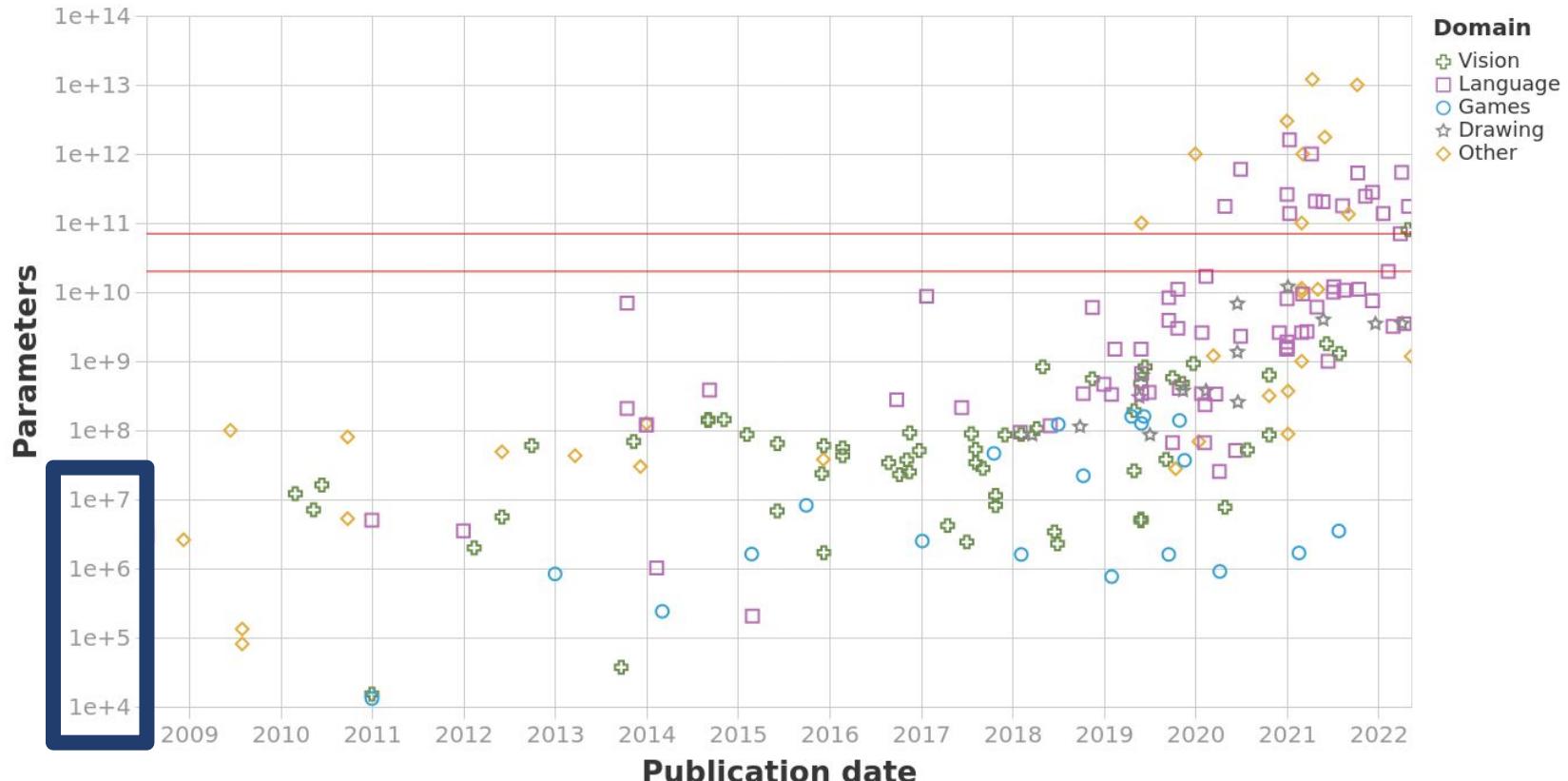


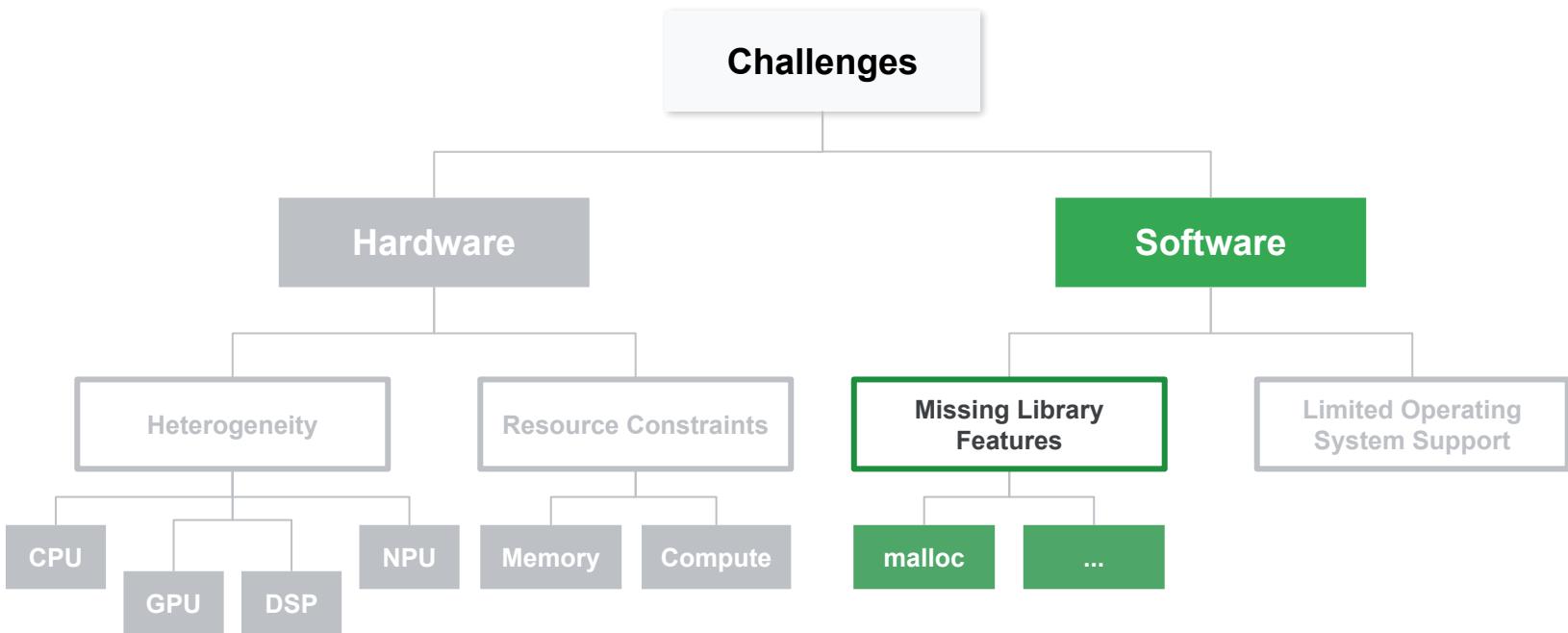


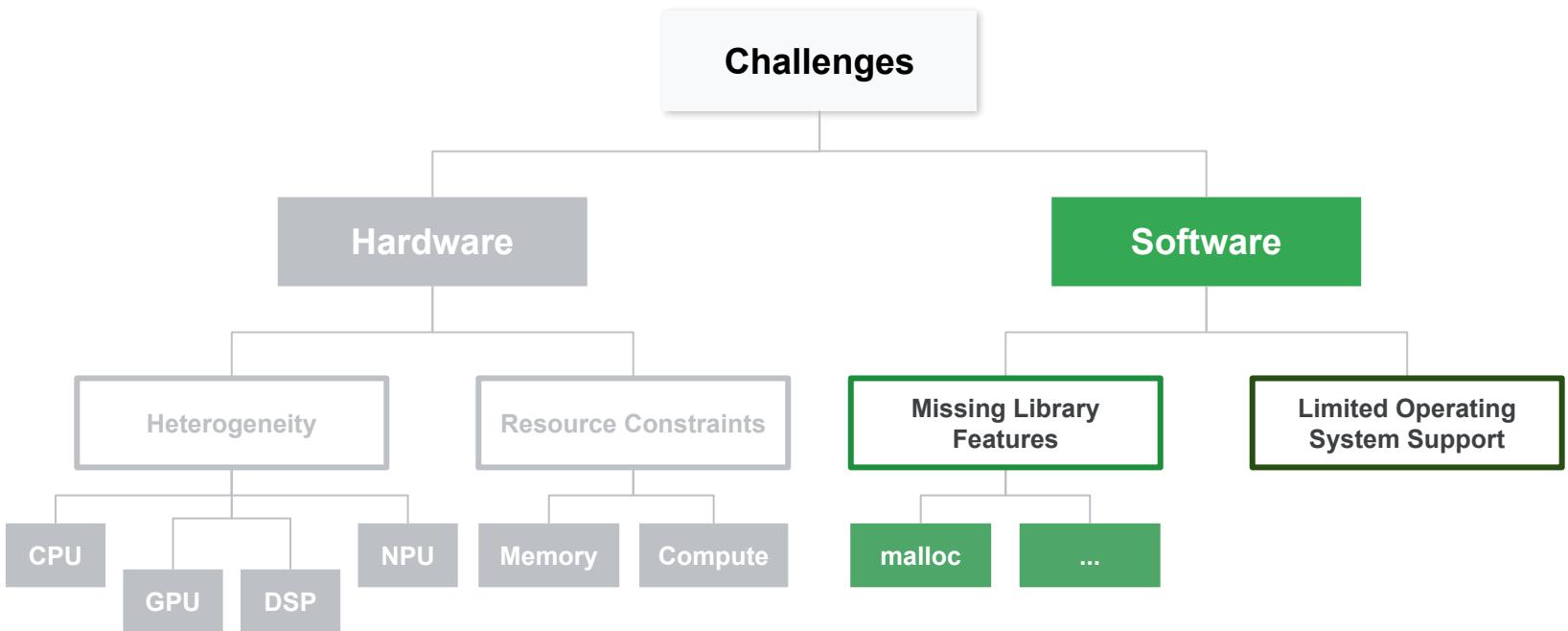
Board	MCU / ASIC	Clock	Memory	Sensors	Radio
	Himax WE-I Plus EVB HX6537-A 32-bit EM9D DSP	400 MHz	2MB flash 2MB RAM	Accelerometer, Mic, Camera	None
	Arduino Nano 33 BLE Sense 32-bit nRF52840	64 MHz	1MB flash 256kB RAM	Mic, IMU, Temp, Humidity, Gesture, Pressure, Proximity, Brightness, Color	BLE
	SparkFun Edge 2 32-bit ArtemisV1	48 MHz	1MB flash 384kB RAM	Accelerometer, Mic, Camera	BLE
	Espressif EYE 32-bit ESP32-D0WD	240 MHz	4MB flash 520kB RAM	Mic, Camera	WiFi, BLE

## Parameters of milestone Machine Learning systems over time

n = 203









# TensorFlow Lite Micro

TensorFlow Learn API Resources Community Why TensorFlow

Search Language

## For Mobile & Edge

Overview Guide Android iOS Other Edge Models Examples API

Filter

Quickstart  
Linux-based devices

Images & video  
Object detection [ ]  
Image classification [ ]  
Video classification [ ]

Audio  
Sound classification [ ]

Microcontrollers  
Overview  
Get started  
Hotword detection [ ]  
Understand the C++ library  
Build and convert models

TensorFlow > Learn > For Mobile & Edge > Other Edge

Was this helpful?

## TensorFlow Lite for Microcontrollers

TensorFlow Lite for Microcontrollers is designed to run machine learning models on microcontrollers and other devices with only few kilobytes of memory. The core runtime just fits in 16 KB on an Arm Cortex M3 and can run many basic models. It doesn't require operating system support, any standard C or C++ libraries, or dynamic memory allocation.

**Note:** The [TensorFlow Lite for Microcontrollers Experiments](#) features work by developers combining Arduino and TensorFlow to create awesome experiences and tools. Check out the site for inspiration to create your own TinyML projects.

### Why microcontrollers are important

Microcontrollers are typically small, low-powered computing devices that are embedded within hardware that requires basic computation. By bringing machine learning to tiny microcontrollers, we can boost the intelligence of billions of devices that we use in our lives, including household appliances and Internet of Things devices, without relying on expensive hardware or reliable internet connections, which is often subject to bandwidth and power constraints and results in high latency. This can also help preserve privacy, since no data leaves the device. Imagine smart appliances that can adapt to your daily routine, intelligent industrial sensors that understand the difference between problems and normal operation, and magical toys that can help kids learn in fun and delightful ways.

### Supported platforms

TensorFlow Lite for Microcontrollers is written in C++ 11 and requires a 32-bit platform. It has been tested extensively with many processors based on the [Arm Cortex-M Series](#) architecture, and has been ported to other architectures [including ESP32](#). The framework is available as an Arduino library. It can also connect directly to development

On this page  
[Why microcontrollers are important](#)  
[Supported platforms](#)  
[Explore the examples](#)  
[Workflow](#)  
[Limitations](#)  
[Next steps](#)

# EDGE IMPULSE



## Create library

Turn your impulse into optimized source code that you can run on any device.



C++ library



Arduino library



Cube.MX CMSIS-  
PACK



WebAssembly



TensorRT library

## Build firmware

Get a ready-to-go binary for your development board that includes your impulse.



ST IoT Discovery Kit



Arduino Nano 33 BLE  
Sense



Espressif ESP-EYE  
(ESP32)



Raspberry Pi  
RP2040



SiLabs Thunderboard  
Sense 2



SiLabs xG24 Dev Kit



Himax WE-I Plus



Nordic nRF52840 DK  
+ IKS02A1



Nordic nRF5340 DK  
+ IKS02A1

## Run your impulse directly

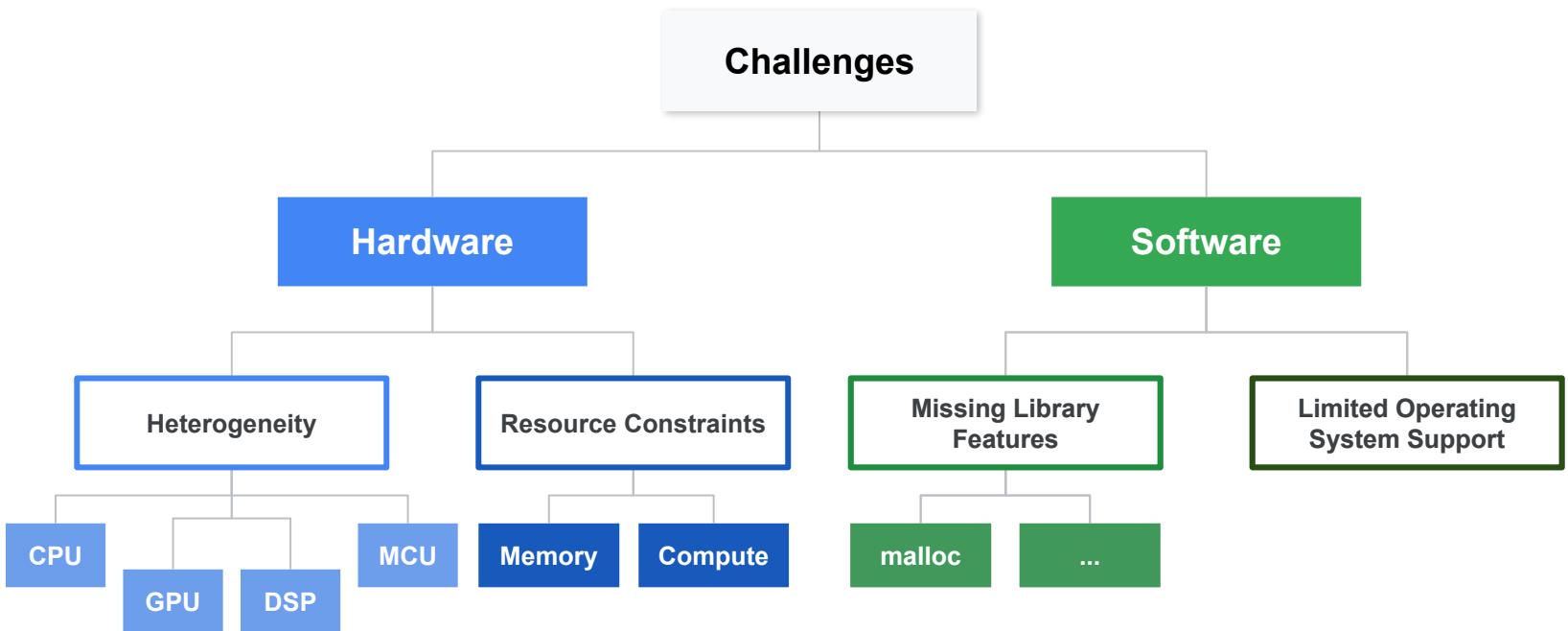
Run this impulse directly on your mobile phone or computer, no app required.



Computer



Mobile phone



# We need to adapt our approach to Edge ML to support a **Responsible AI** future!

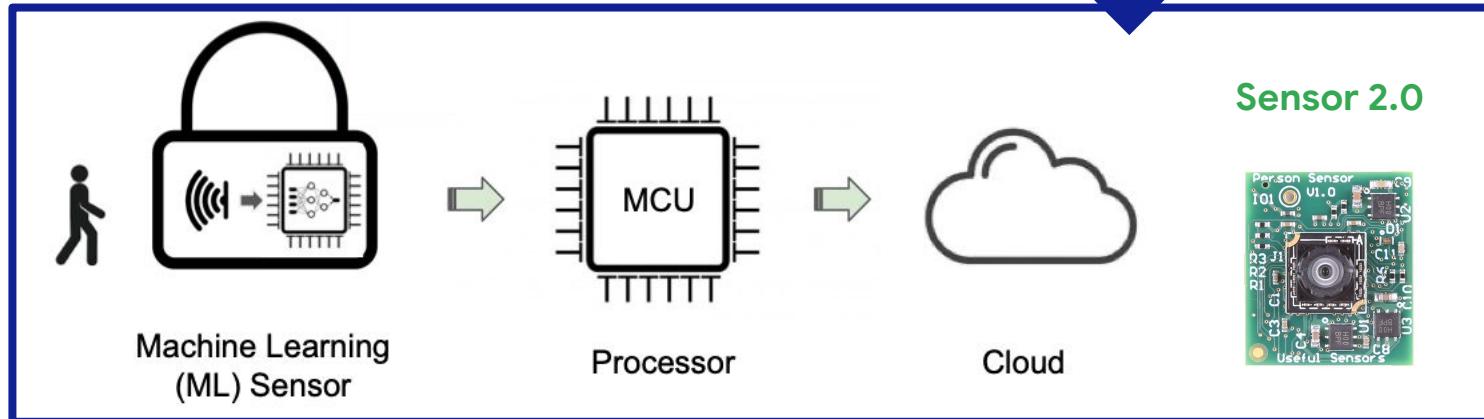
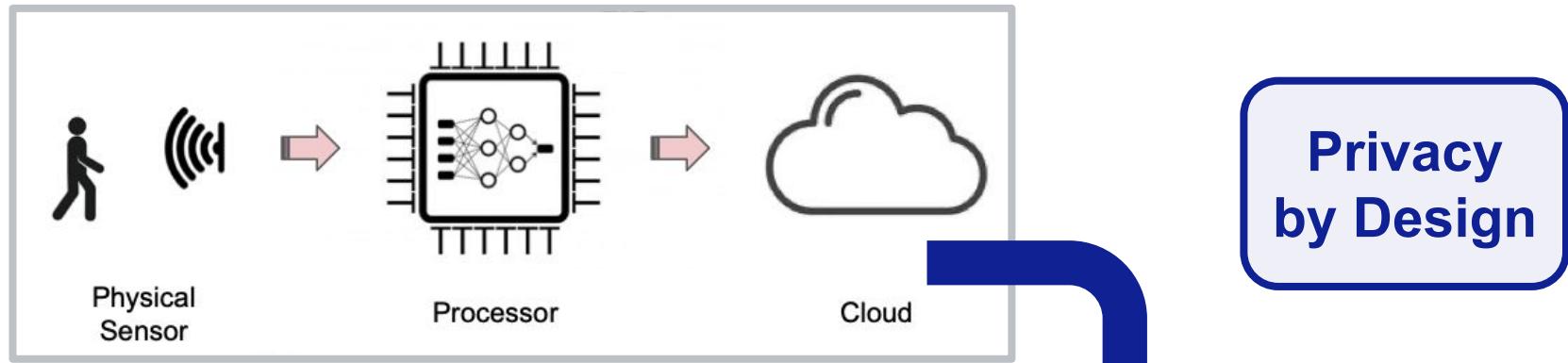
The collage consists of five screenshots from different news websites:

- Top Left:** A screenshot of a browser window showing multiple tabs. One tab is titled "How to Stop Your Smart TV From Spying on You" from [Academy](#). The page content discusses how a voice command starts your TV's recognition and viewing data, and how interconnectivity has privacy implications.
- Top Middle:** A screenshot of a browser window showing multiple tabs. One tab is titled "AI is harming our planet: addressing AI's staggering energy cost" from [DZone](#). The page content discusses the environmental impact of AI, specifically its energy consumption.
- Bottom Left:** A screenshot of a browser window showing multiple tabs. One tab is titled "Google Calls Hidden Microphone in Its Nest Home Security Devices an 'Error'" from [Popular Mechanics](#). The page content discusses Google's admission of a design oversight in their Nest smart home devices.
- Bottom Middle:** A screenshot of a browser window showing multiple tabs. One tab is titled "How to stop your smart home spying on you" from [The Observer](#). The page content provides tips for preventing smart home devices from spying on users.
- Bottom Right:** A screenshot of a browser window showing multiple tabs. One tab is titled "Deep Learning's Carbon Emissions Problem" from [Forbes](#). The page content discusses the significant carbon footprint of deep learning models and training processes.

A screenshot of a news article from [WIRED](#) titled "AI Can Do Great Things—if It Doesn't Burn the Planet". The article, written by Will Knight on January 21, 2020, at 7:00 AM, discusses the computing power required for AI landmarks, such as recognizing images and defeating humans at Go, which increased 300,000-fold from 2012 to 2018. The article includes a large image of server racks in a data center.

# Sustainability & Privacy: MLSensors at Scale

# What is a Machine Learning Sensor?



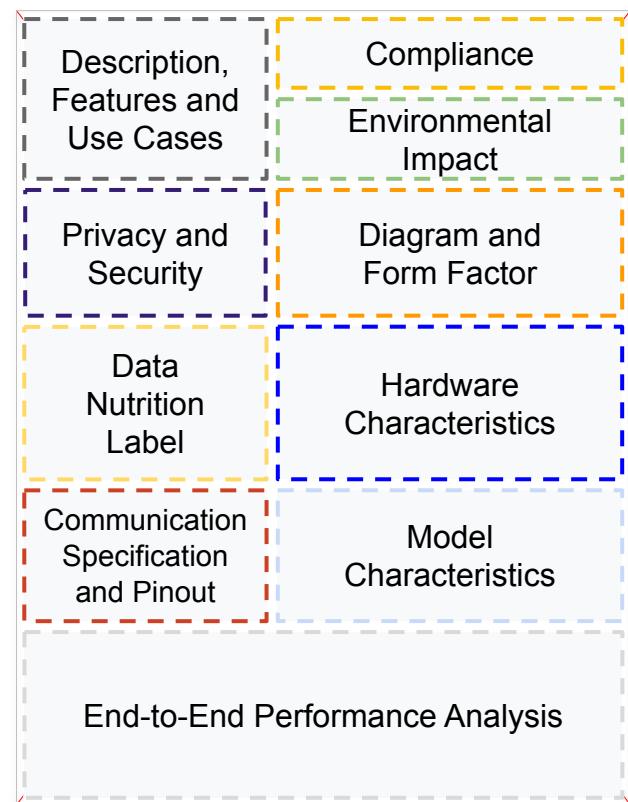
# We suggest **transparency** as a core value to overcome these challenges.

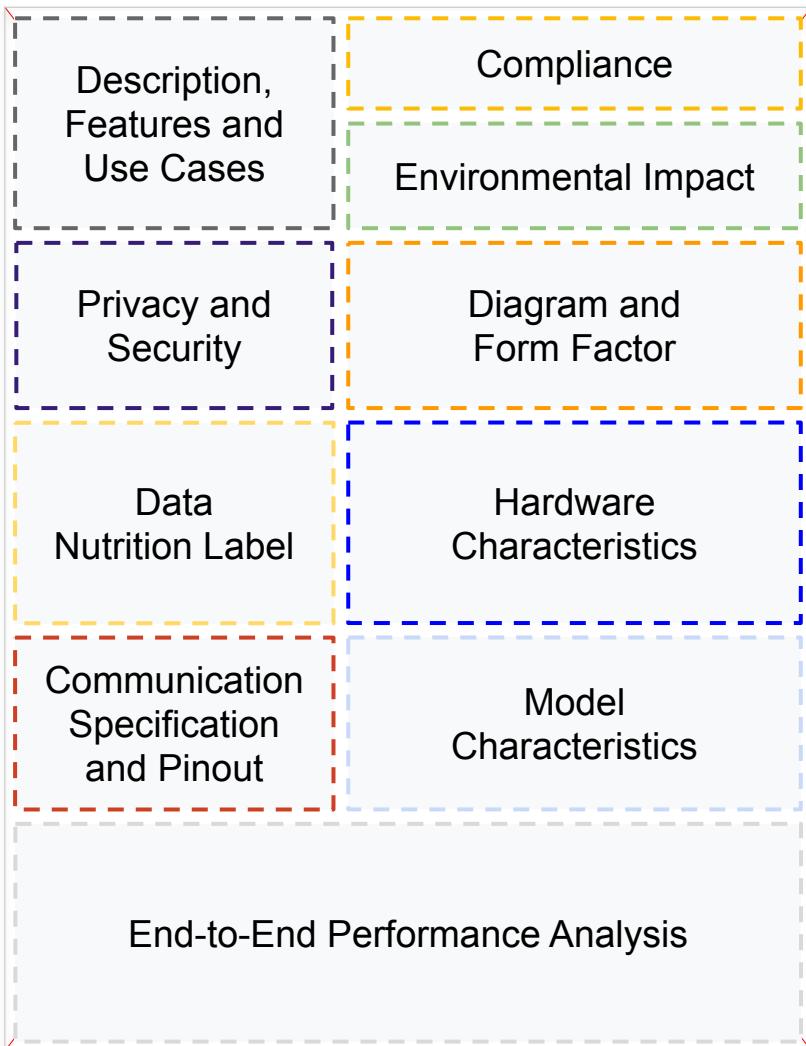
## Datasheets for Machine Learning Sensors

Matthew Stewart<sup>1\*</sup> Pete Warden<sup>2,5</sup> Yasmine Omri<sup>1</sup> Shvetank Prakash<sup>1</sup> Joao Santos<sup>1</sup>  
Shawn Hymel<sup>4</sup> Benjamin Brown<sup>1</sup> Jim MacArthur<sup>1</sup> Nat Jeffries<sup>5</sup> Brian Plancher<sup>3</sup>  
Vijay Janapa Reddi<sup>1</sup>

<sup>1</sup>Harvard University <sup>2</sup>Stanford University <sup>3</sup>Barnard College, Columbia University  
<sup>4</sup>Edge Impulse <sup>5</sup>Useful Sensors

[arxiv.org/abs/2306.08848](https://arxiv.org/abs/2306.08848)

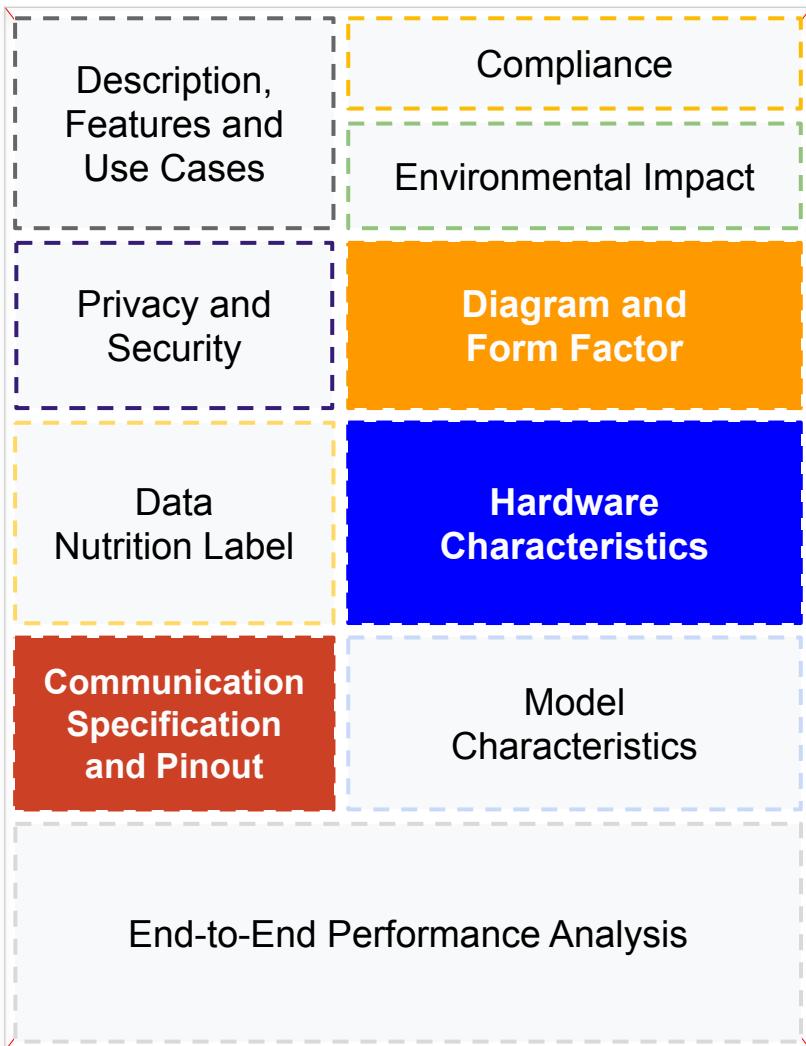




# ML Sensor Datasheets

## Have 3 Goals:

1. Raise the level of **abstraction**
2. Transparent at the **hardware, data, model, and end-to-end layers**
3. Support **Responsible** Use

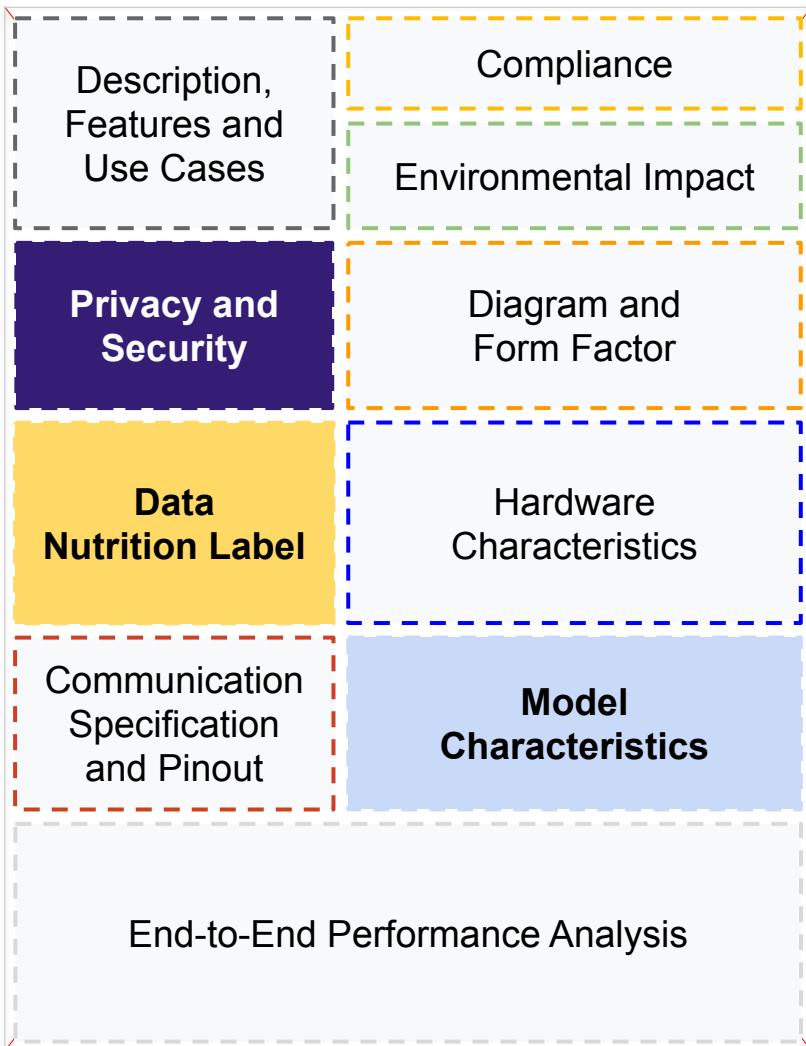


# ML Sensor Datasheets

## Have 3 Goals:

1. Raise the level of abstraction
2. Transparent at the **hardware**, data, model, and end-to-end layers
3. Support Responsible Use

Standard Sensor  
Datasheet



# ML Sensor Datasheets

## Have 3 Goals:

1. Raise the level of abstraction
2. Transparent at the hardware, **data, model**, and end-to-end layers
3. Support **Responsible** Use

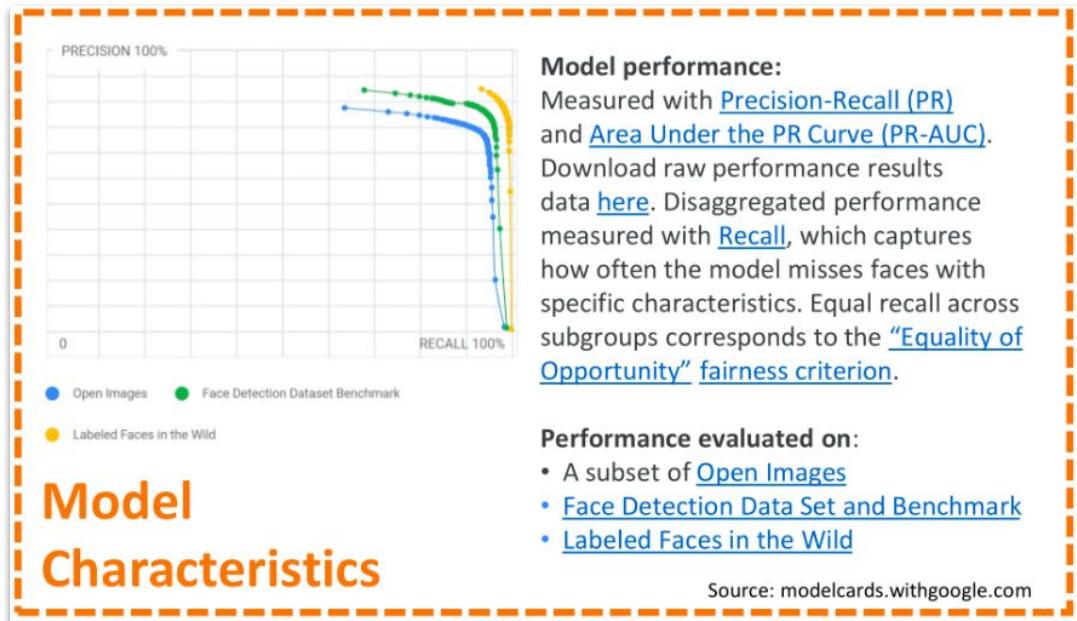
**Responsible  
Machine Learning  
Analysis**

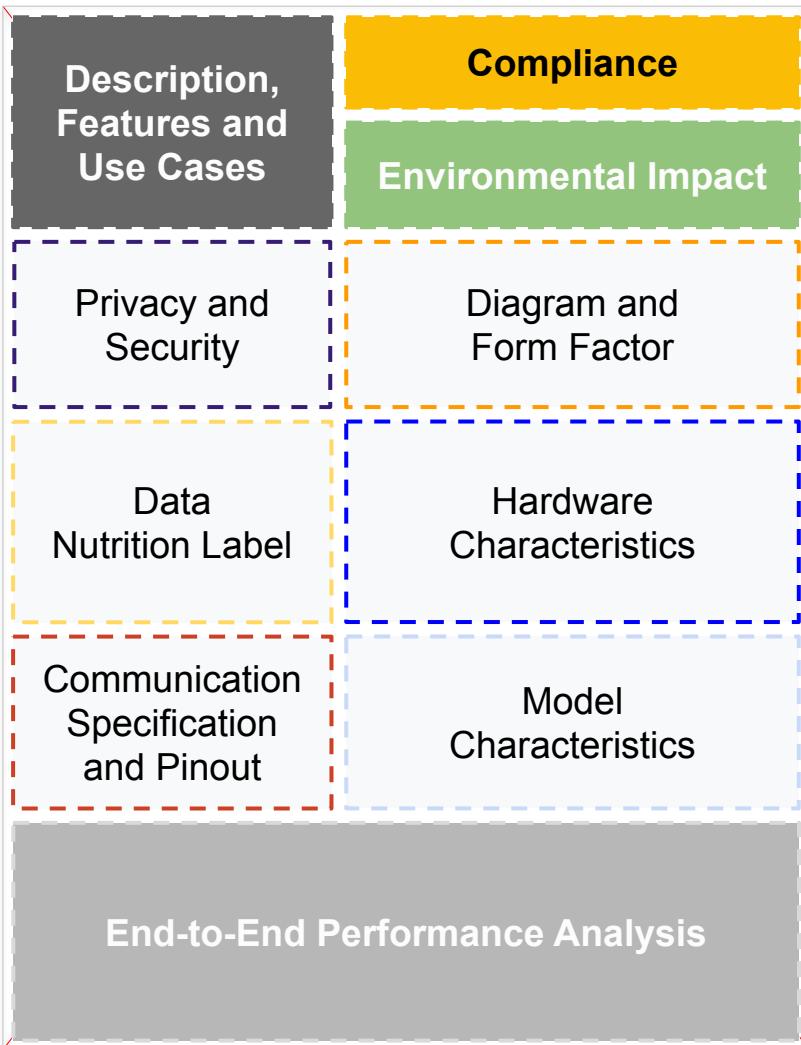
# Responsible Machine Learning Analysis

Source: datanutrition.org

The dashboard includes icons for Aggregated Human Data, Quality Review, Ethical Review, About Humans, Commercial License, Multi-source Funded, Not Actively Updated, Multi-source Data, and No Subpopulations.

## Dataset Nutrition Label



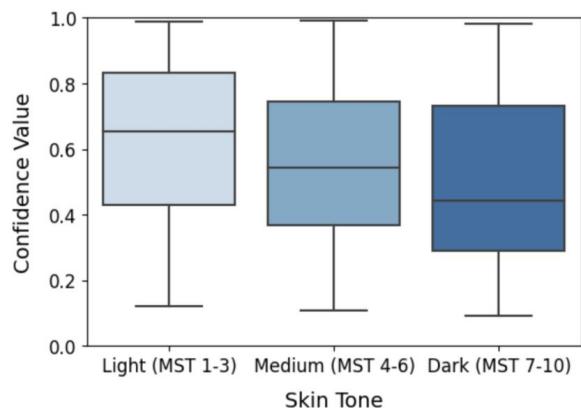
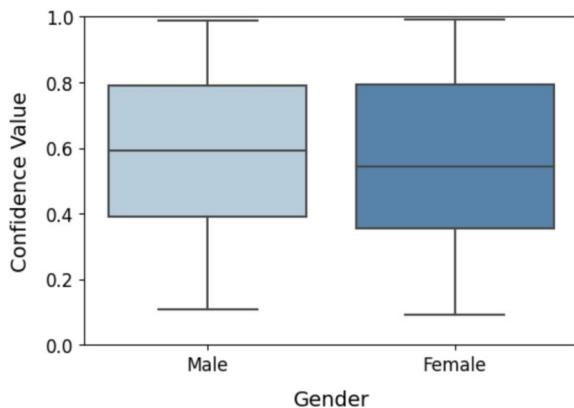
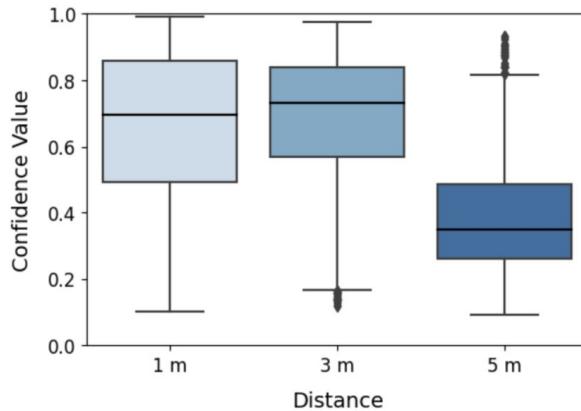
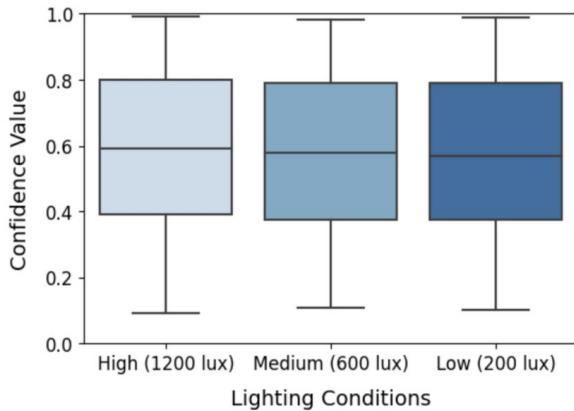


# ML Sensor Datasheets

## Have 3 Goals:

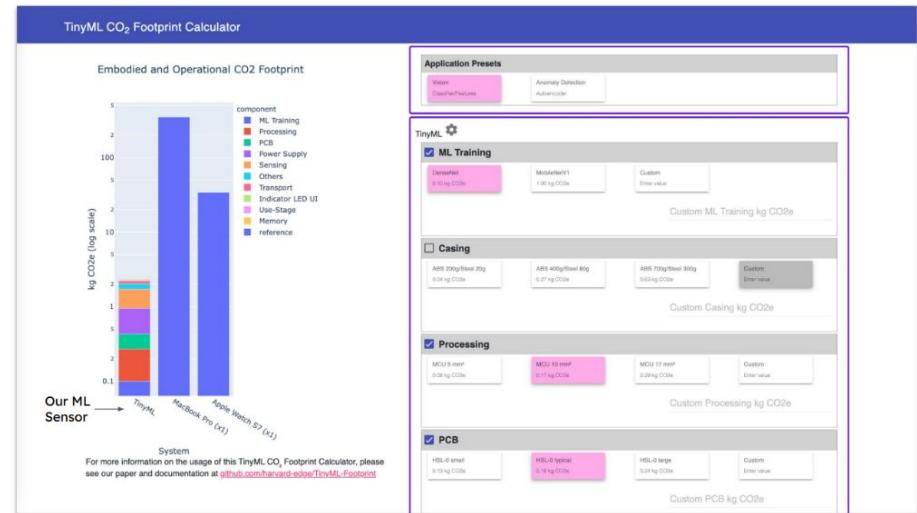
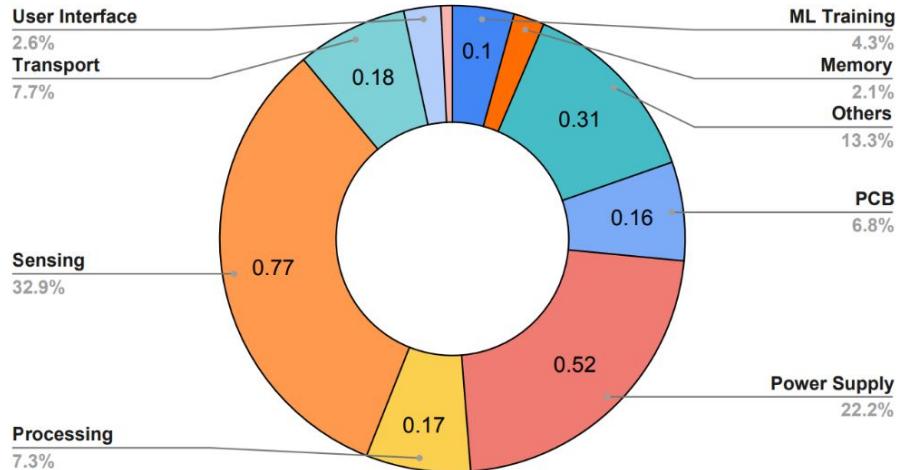
1. Raise the level of **abstraction**
2. Transparent at the hardware, data, model, and **end-to-end layers**
3. Support **Responsible** Use

Overall System  
Analysis



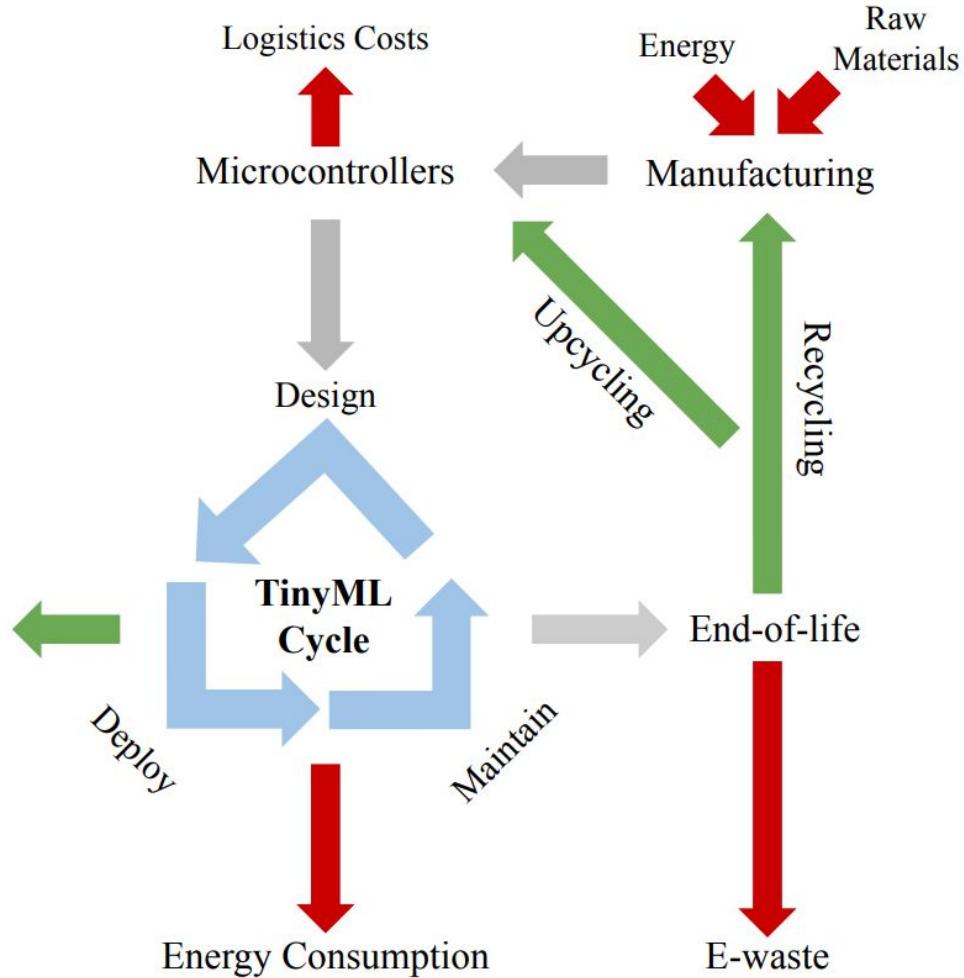
# End-to-End Responsible Performance Analysis

# Environmental Impact



Lets Explore this Impact  
in More Detail

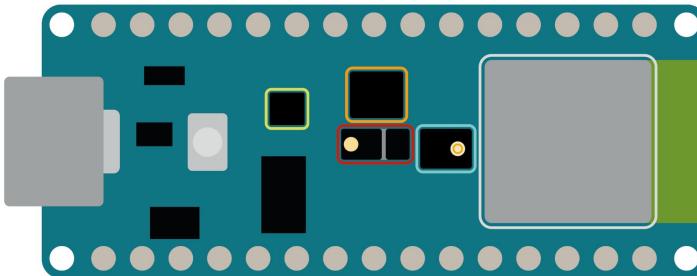
## Sustainable Development Goals



TinyML can support the SDGs but comes with costs. **What is the net impact?**

[arxiv.org/abs/2301.11899](https://arxiv.org/abs/2301.11899)

# Real TinyML Systems are more than just an MCU!

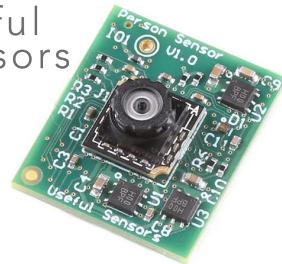


- ◆ Color, brightness, proximity and gesture sensor
- ◆ Digital microphone
- ◆ Motion, vibration and orientation sensor
- ◆ Temperature, humidity and pressure sensor
- ◆ Arm Cortex-M4 microcontroller and BLE module

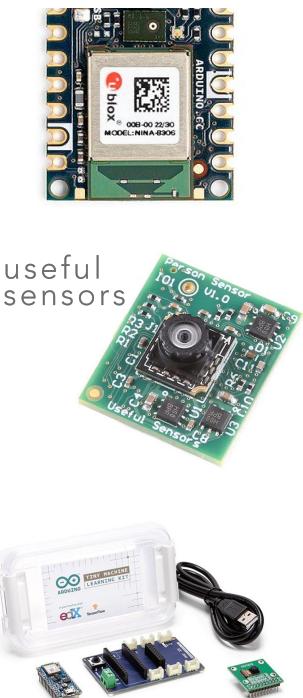
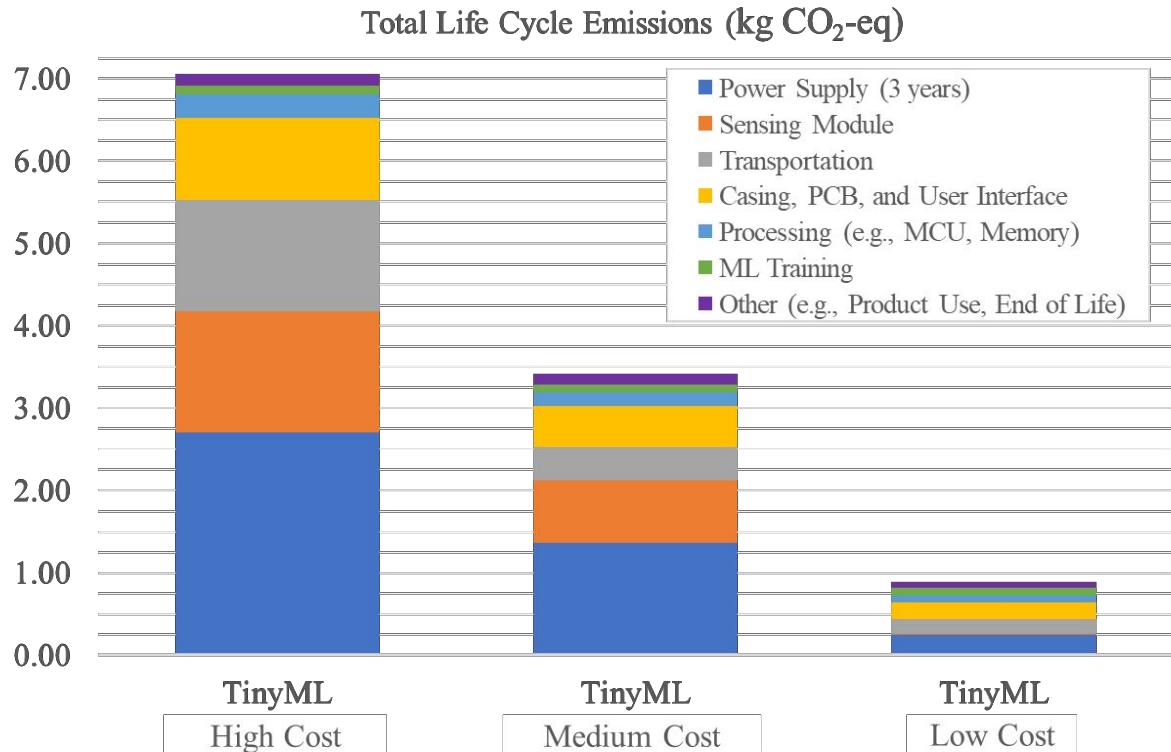


# Building Representative Systems

Cost Level	High Cost	Medium Cost	Low Cost
Application	Image Classification		Keyword Spotting
Size	Large	Compact	Compact



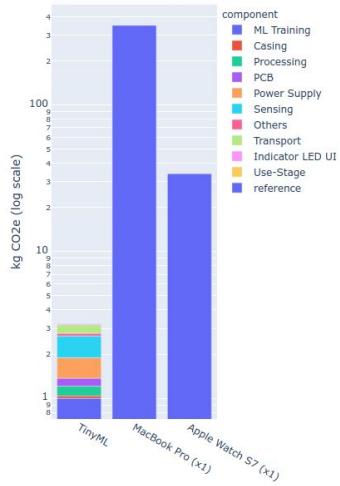
# Building Representative Systems



# [harvard-edge.github.io/TinyML-Footprint/](https://harvard-edge.github.io/TinyML-Footprint/)

## TinyML CO<sub>2</sub> Footprint Calculator

Embodied and Operational CO<sub>2</sub> Footprint



For more information on the usage of this TinyML CO<sub>2</sub> Footprint Calculator, please see our paper and documentation at [github.com/harvard-edge/TinyML\\_Footprint](https://github.com/harvard-edge/TinyML_Footprint)

Application Presets

Vision  
Classifier/Features      Anomaly Detection  
Autoencoder

**TinyML**

**ML Training**

DenseNet 0.10 kg CO <sub>2</sub> e	MobileNetV1 1.00 kg CO <sub>2</sub> e	Custom Enter value
---------------------------------------	--	-----------------------

Custom ML Training kg CO<sub>2</sub>e

**Casing**

ABS 200g/Steel 20g 0.04 kg CO <sub>2</sub> e	ABS 400g/Steel 80g 0.27 kg CO <sub>2</sub> e	ABS 700g/Steel 300g 0.63 kg CO <sub>2</sub> e	Custom Enter value
---	---	--	-----------------------

Custom Casing kg CO<sub>2</sub>e

**Processing**

MCU 5 mm* 0.08 kg CO <sub>2</sub> e	MCU 10 mm* 0.17 kg CO <sub>2</sub> e	MCU 17 mm* 0.29 kg CO <sub>2</sub> e	Custom Enter value
--	---	---	-----------------------

Custom Processing kg CO<sub>2</sub>e

**PCB**

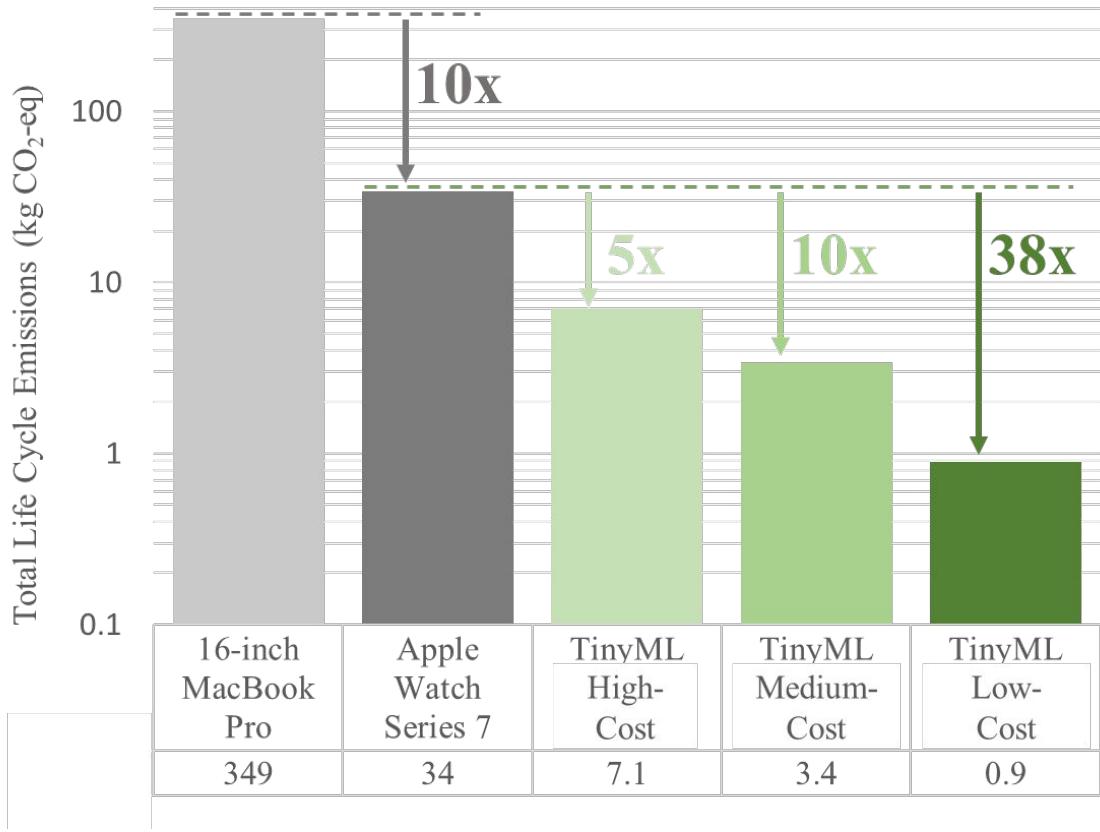
HSL-0 small 0.13 kg CO <sub>2</sub> e	HSL-0 typical 0.16 kg CO <sub>2</sub> e	HSL-0 large 0.24 kg CO <sub>2</sub> e	Custom Enter value
--	--	--	-----------------------

Custom PCB kg CO<sub>2</sub>e

**Power Supply**

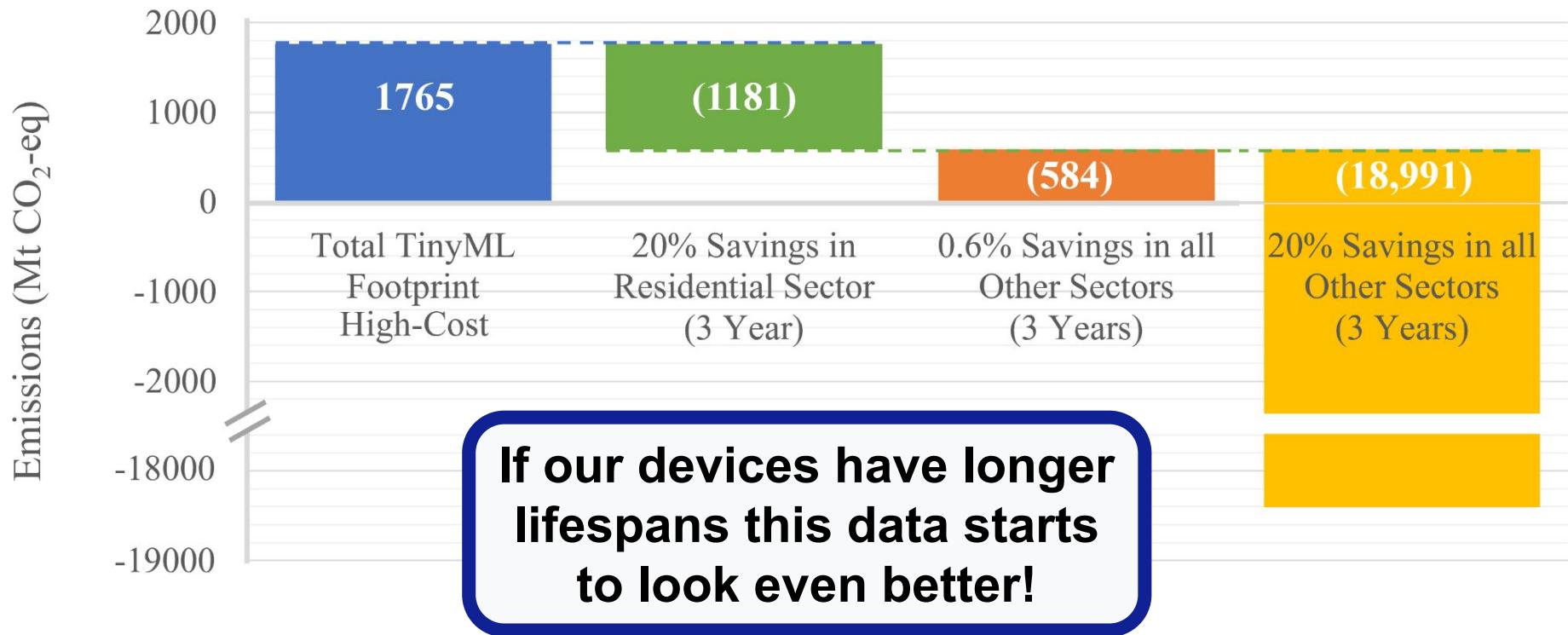


# TinyML Systems in Context



**5x to 38x  
Savings  
over a  
3-year  
lifespan!**

# What if we scale to 250bn devices?



# The Big Impact of TinyML: Embedded Machine Learning at the Extreme Edge



Workshop on Widening Access to TinyML Network  
by Establishing Best Practices in Education | (smr 3851)

ICTP International Centre for Theoretical Physics AL AKAHAWYUN UNIVERSITY

Workshop, Trieste, Italy  
3 - 7 July 2023

[tinyMLedu.org](http://tinyMLedu.org)



Brian Plancher  
Barnard College, Columbia University  
[brianplancher.com](http://brianplancher.com)

جامعة الانهرين  
MoroccoAI  
AI Summer School  
AL AKHAWYUN  
UNIVERSITY

