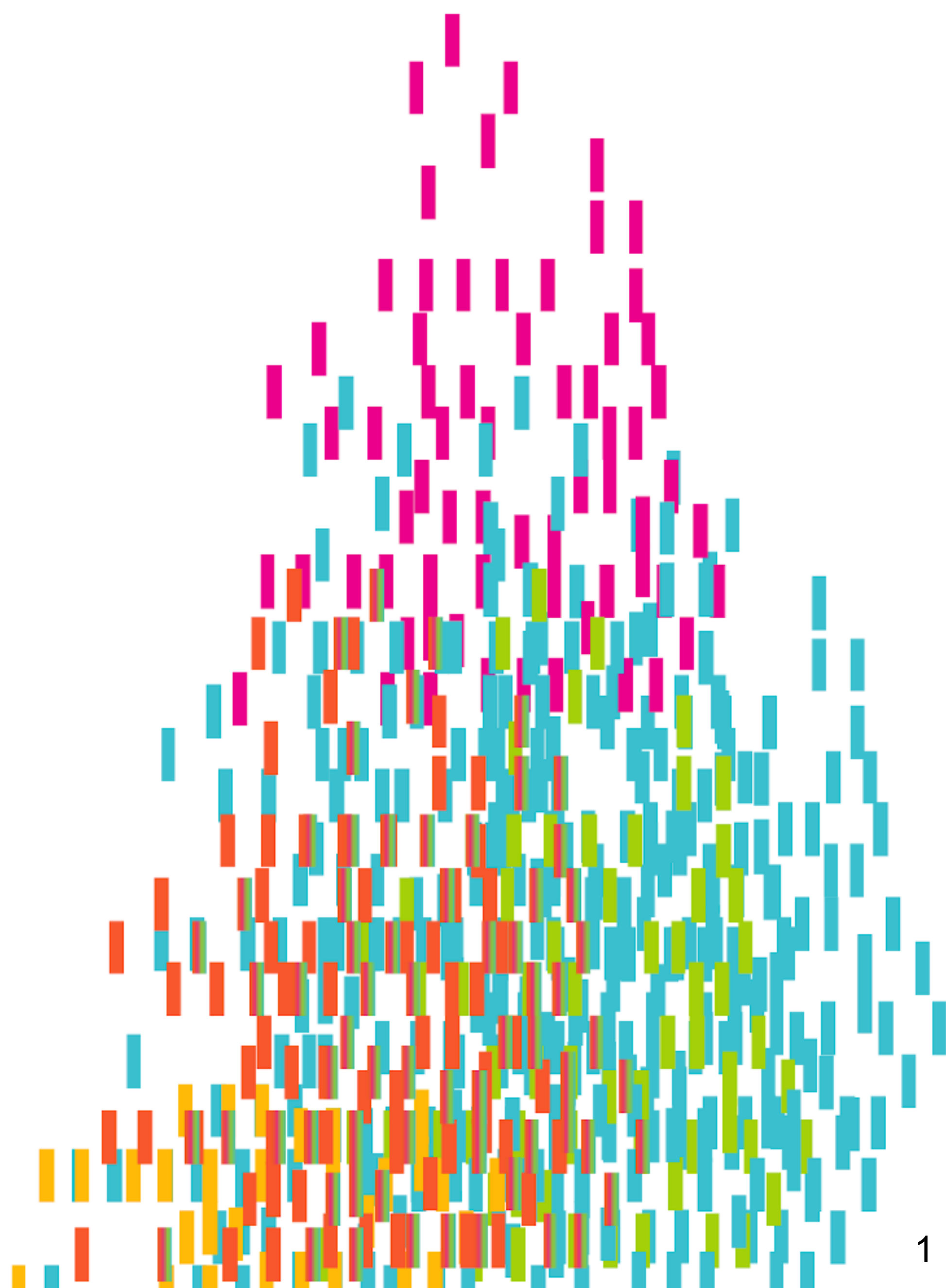


EON Tuner

AutoML for real-world embedded devices



edge ai + vision
ALLIANCE
Formerly the Embedded Vision Alliance



Hi, I'm Jan!

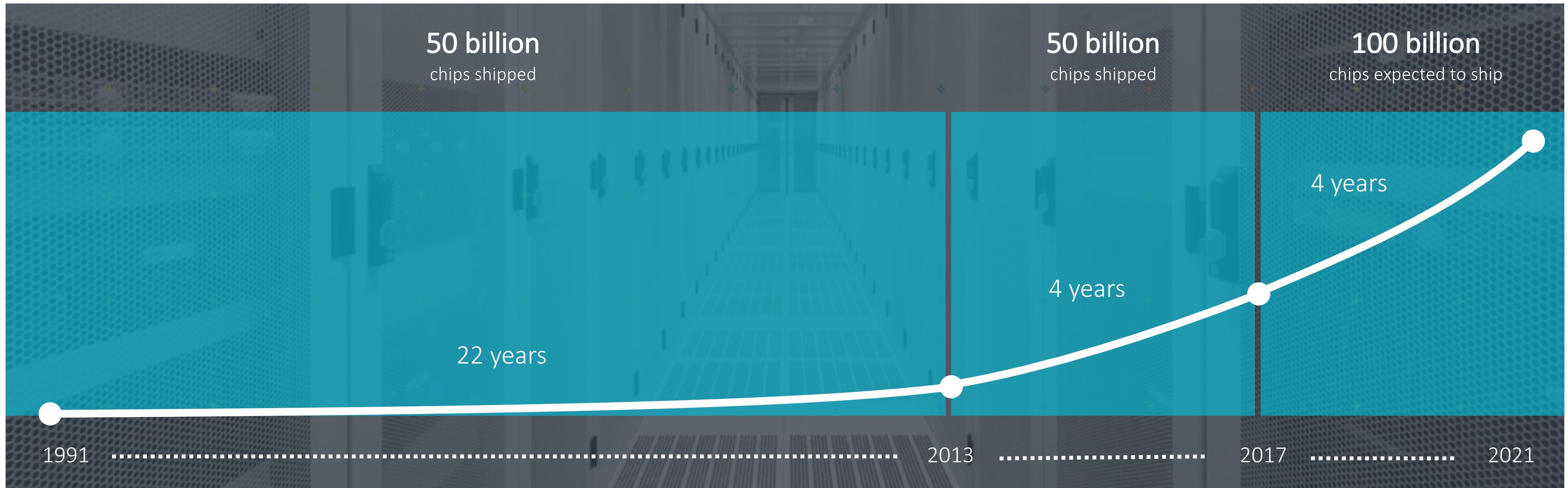
Cofounder and CTO at **Edge Impulse**

Building connected cameras since 2014

Ex-Firefox OS, ex-IoT at Arm



SO. MANY. DEVICES!



I was promised insight in the world, but I get...



↓ -10°C



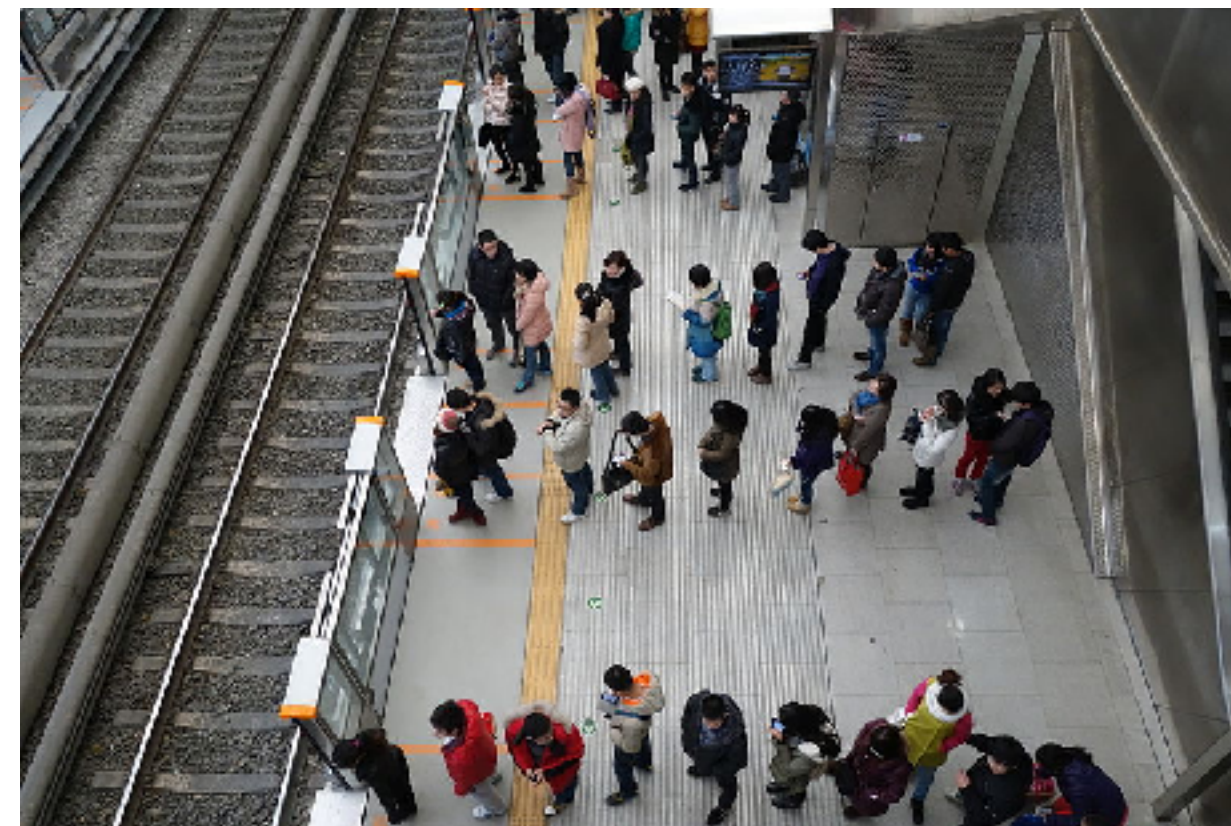
Lots of visual information in the world



Does this bottle have the right label?



Do I see an elephant?



How many people are in line?



But requiring a human is not great...

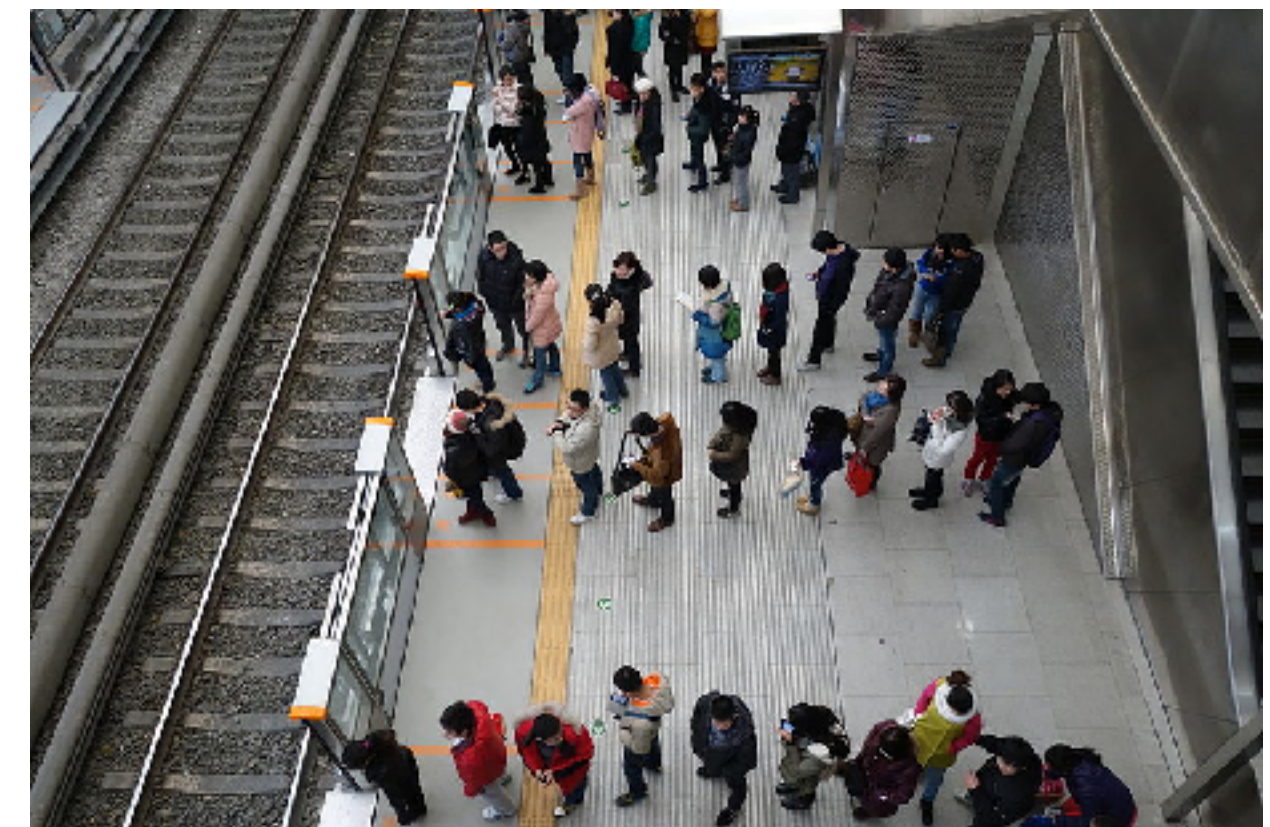
Humans like to sleep

Humans cost money (*)

Being so close to an elephant that you can see it might be dangerous (e.g. poachers)



Do I see an elephant?



How many people are in line?



Sensors need to be more like us

Infinite sounds, images, motions all around us

We know there's a correlation

We can probably collect the data, but mostly throw it away: devices are cost, bandwidth or power constraint

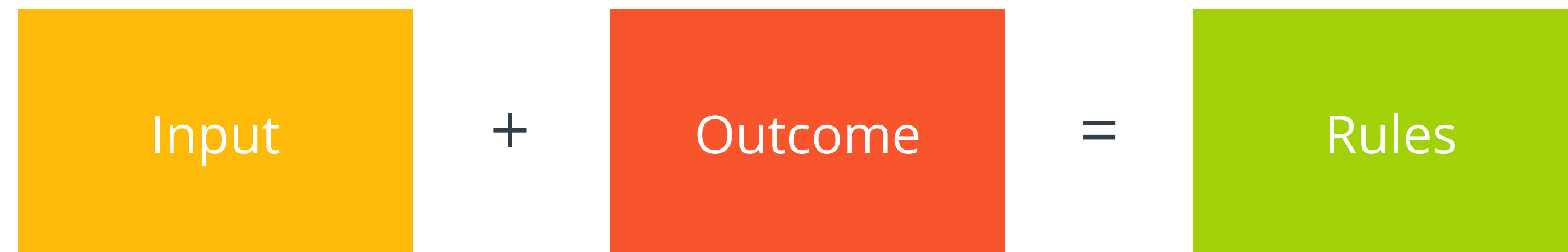


Machine learning helps you find the rules

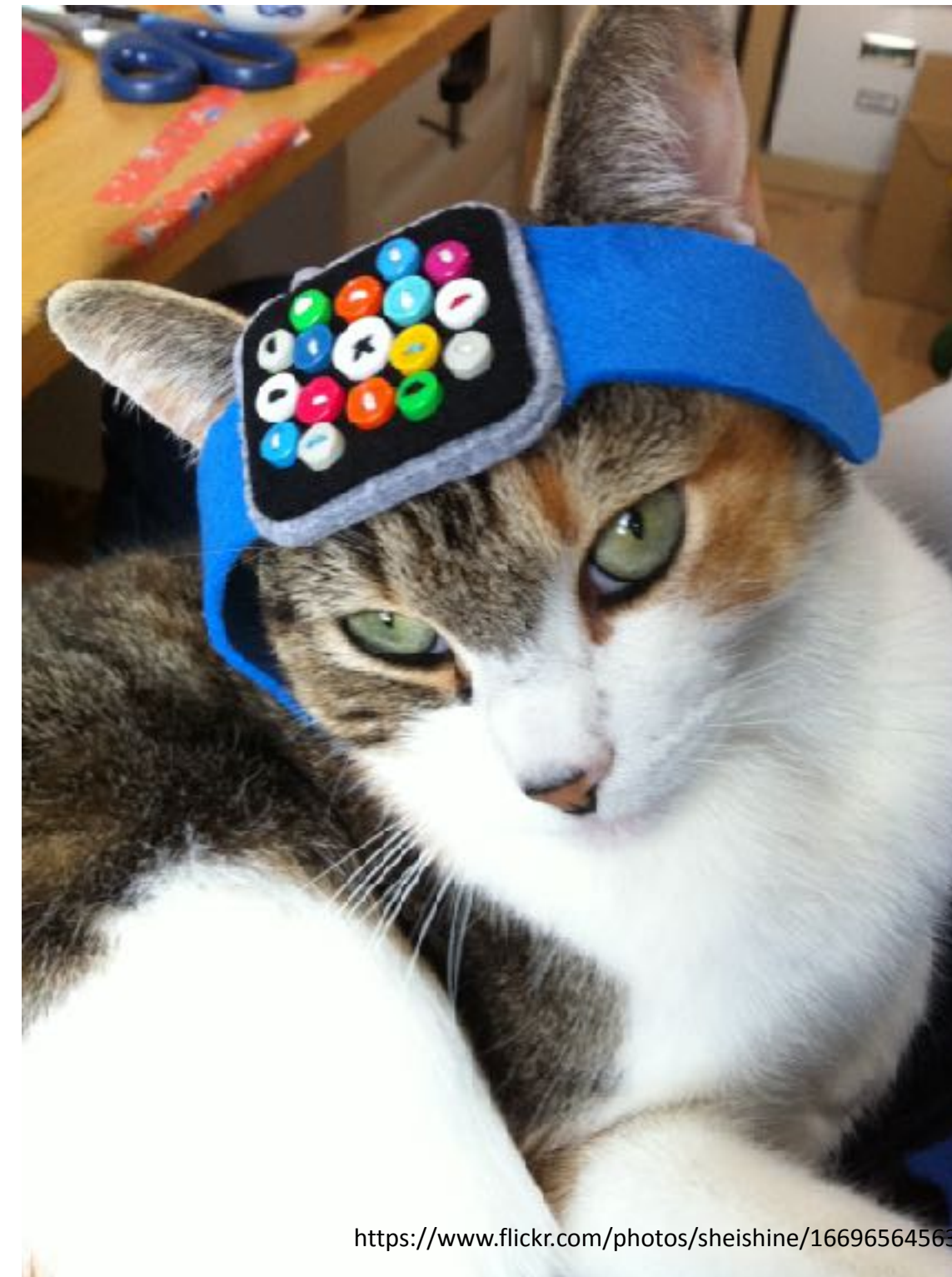
Normal programming



Machine learning



Not just vision



Biosignal analysis



Detecting abnormal vibration



Anything with messy, high-resolution sensor data





Machine learning?



Lots of development!

Running on device is key

Constrained devices, but also constrained usecases

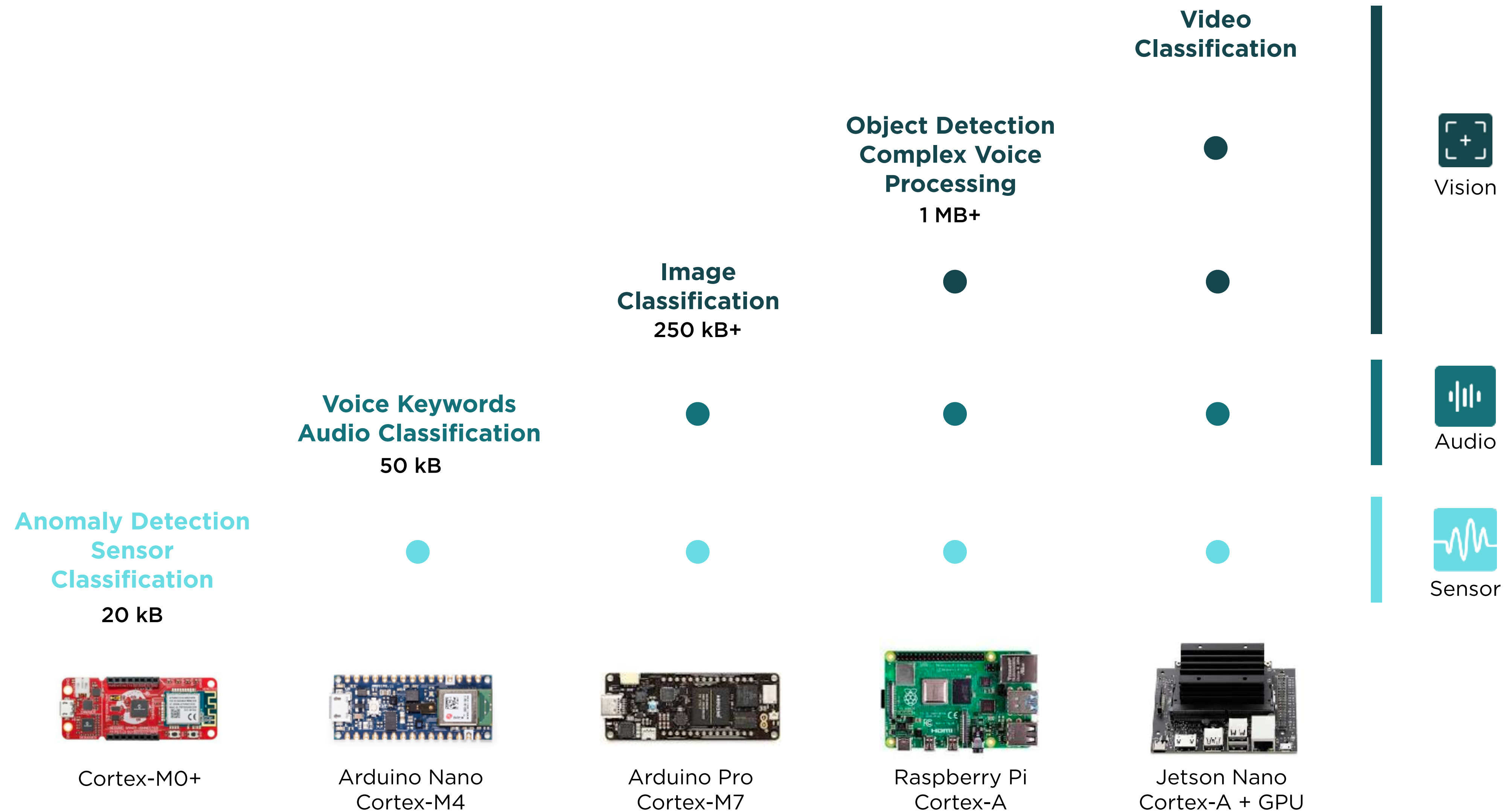
Quantization, knowledge distillation, neural network compilers, hardware acceleration (even on MCU)

Design within the constraints of the device

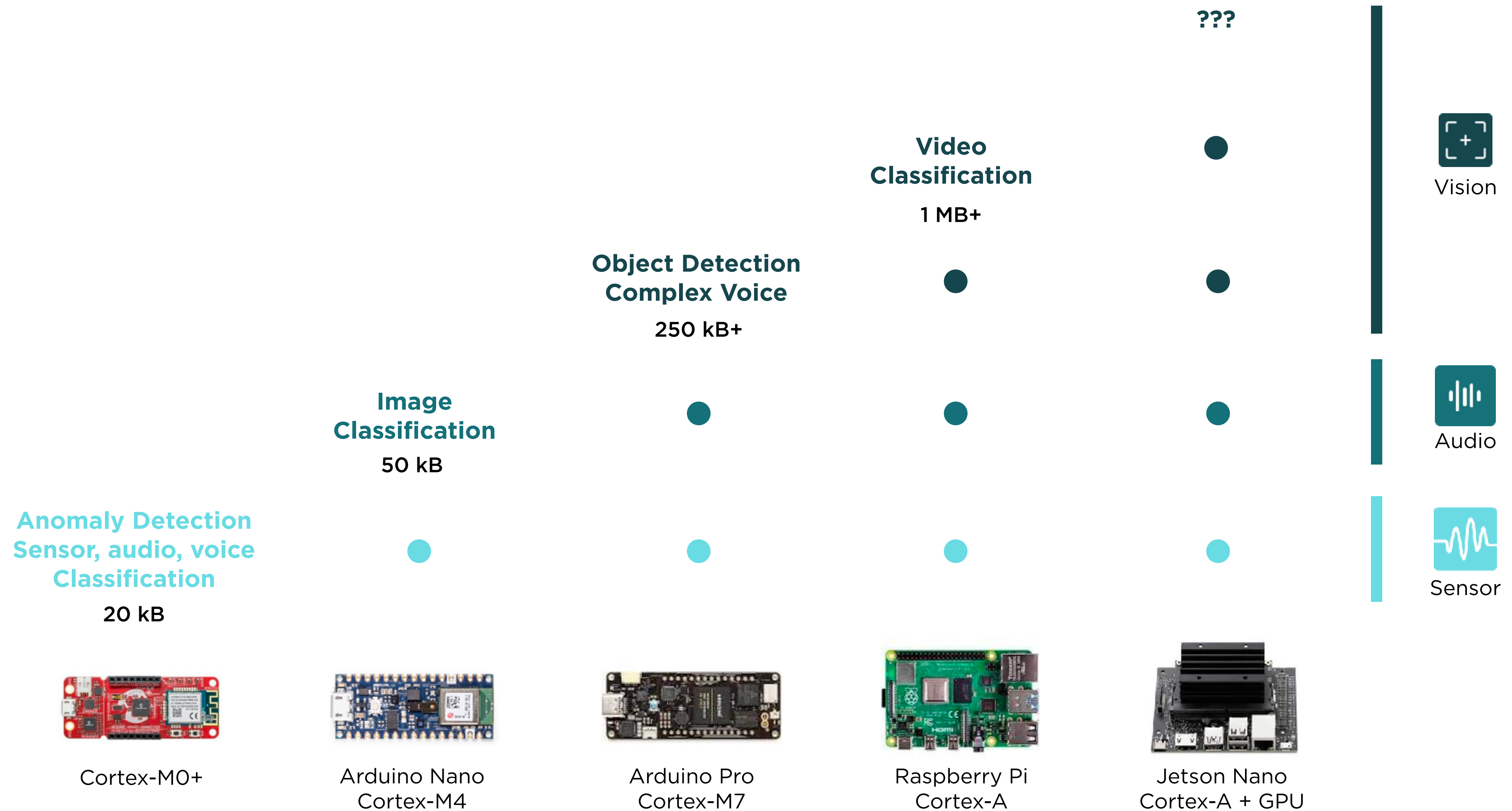
You can actually do a lot!



Where we are today



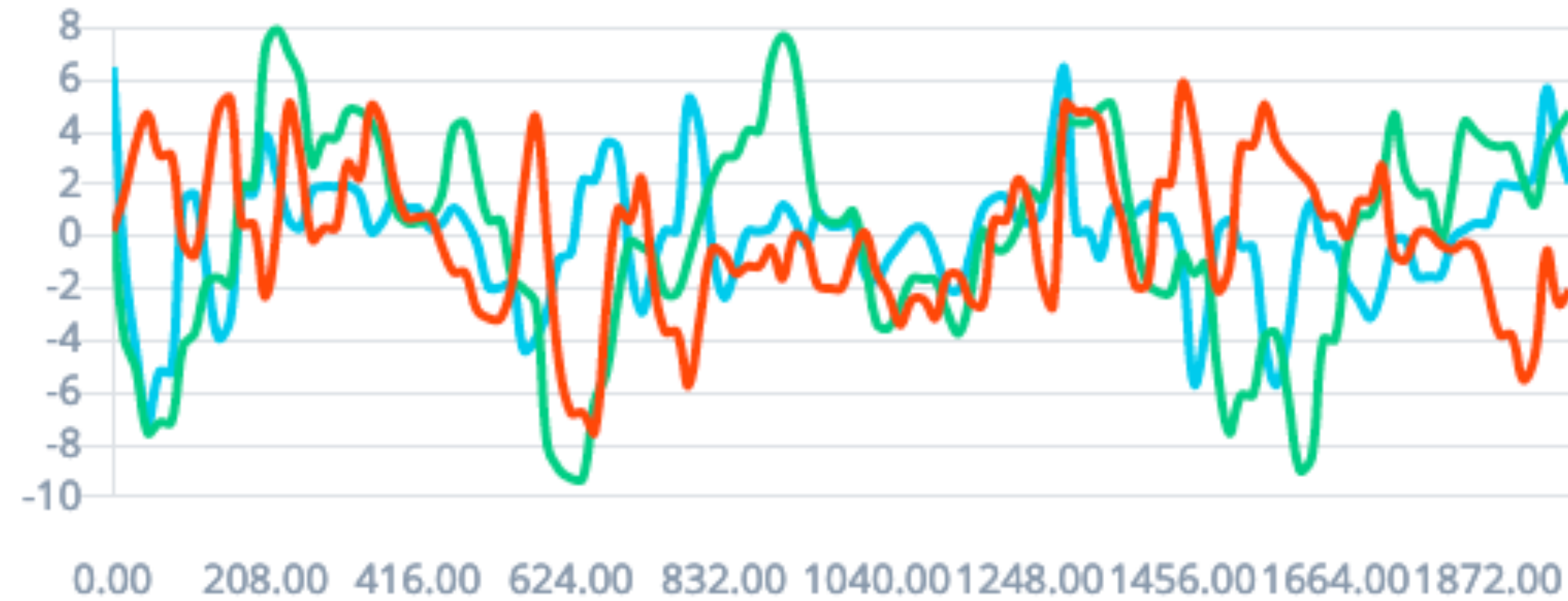
Where we'll be tomorrow



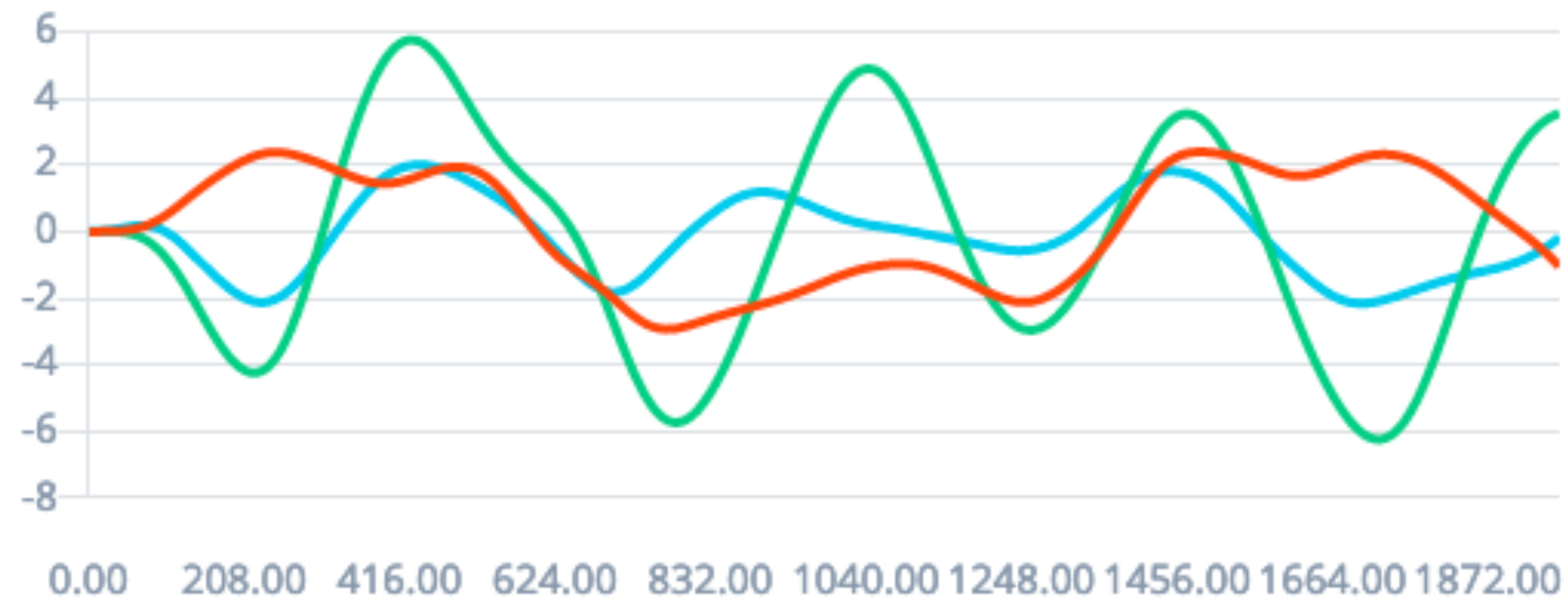
Side-step to non-vision



Signal processing + ML = ❤️



Apply low-pass filter...



= much easier job
for the ML algo



Leveraging signal processing

On-device intelligence is not new

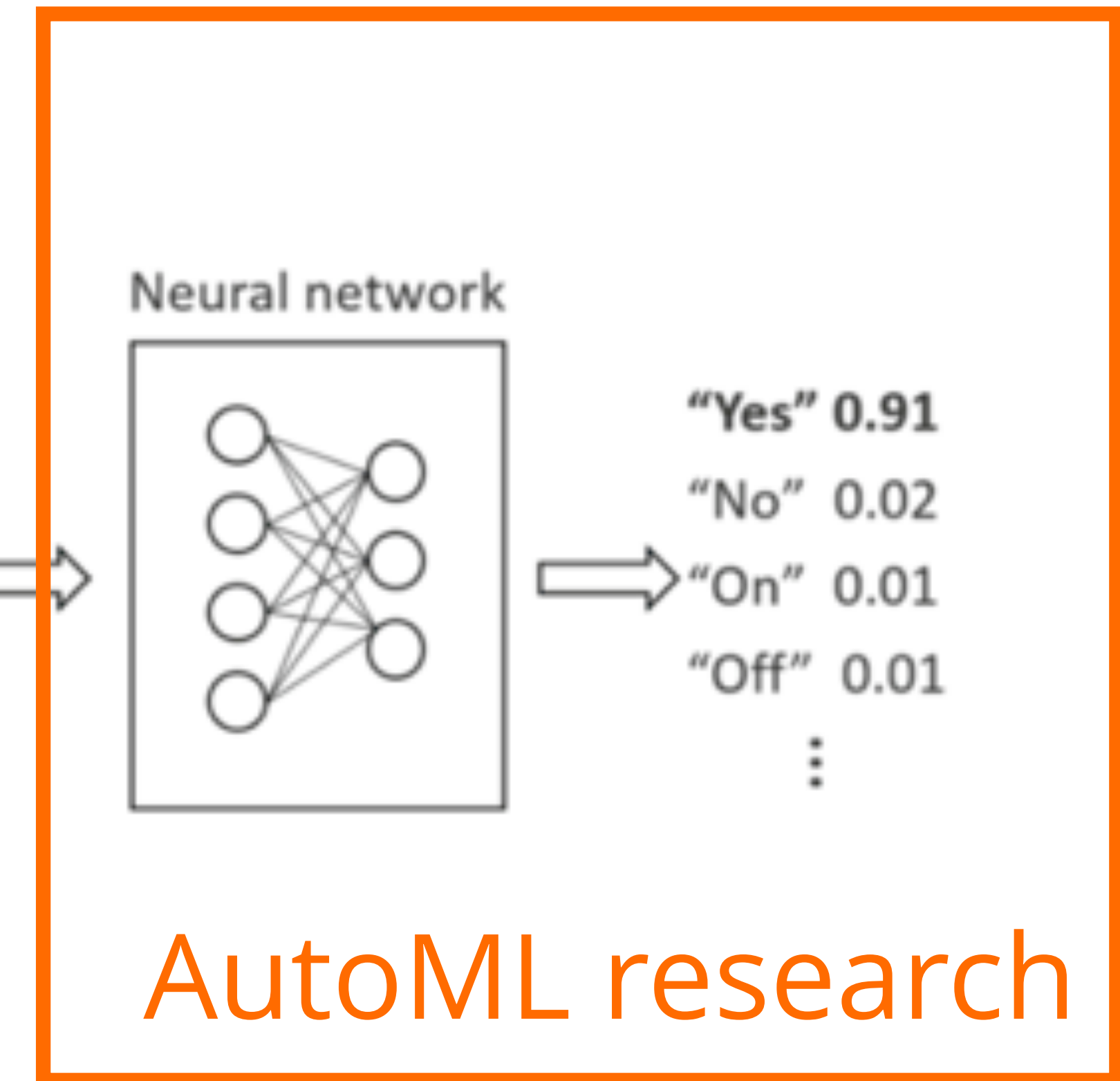
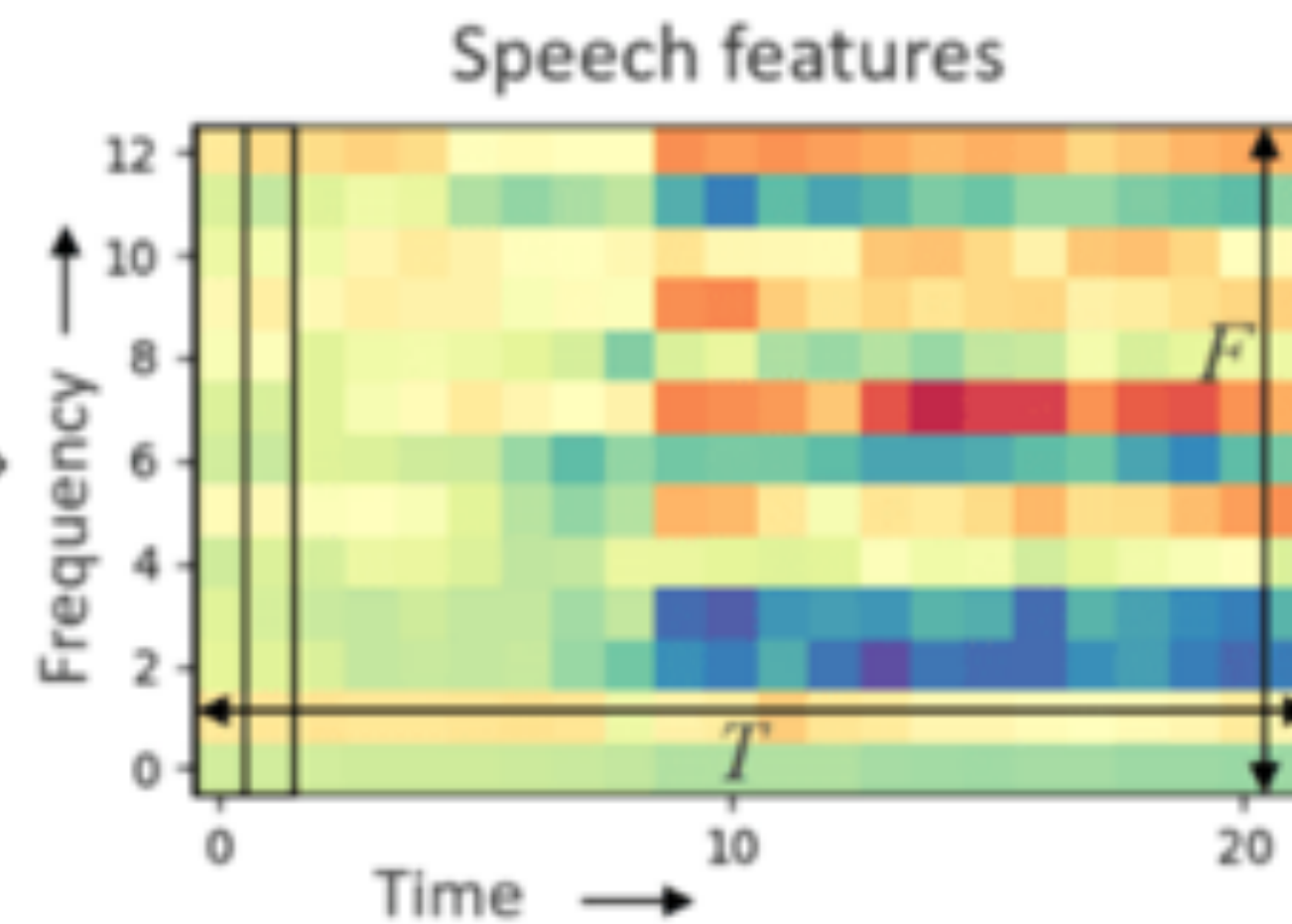
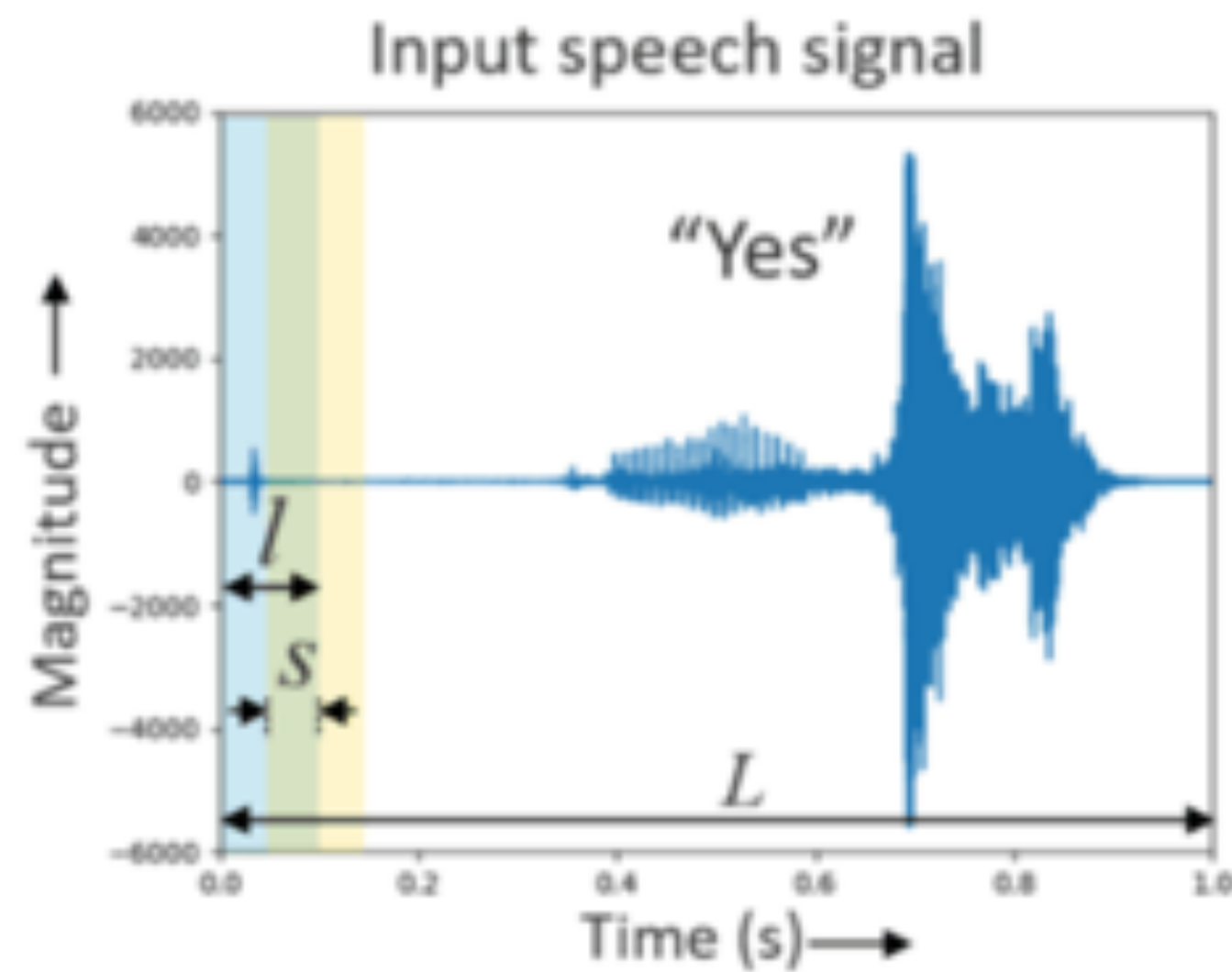
Neural networks are inefficient, if you can preprocess? Do so!

Significantly reduce input features, leading to smaller networks.

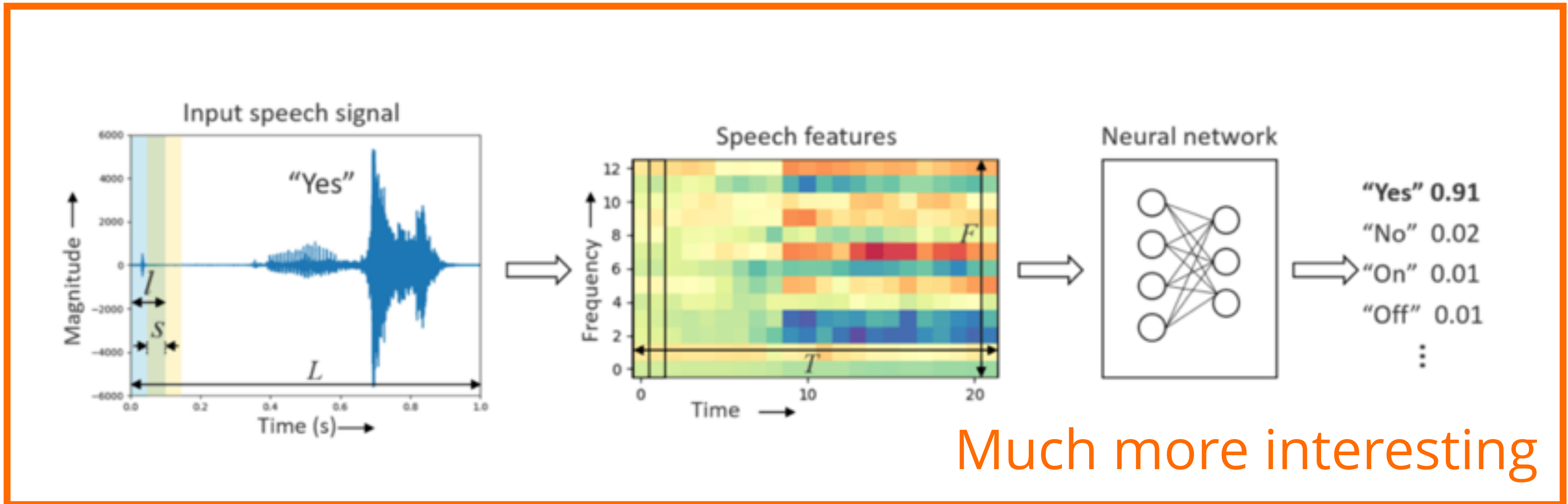
Cleans up input



ML Sensor pipeline is often combination



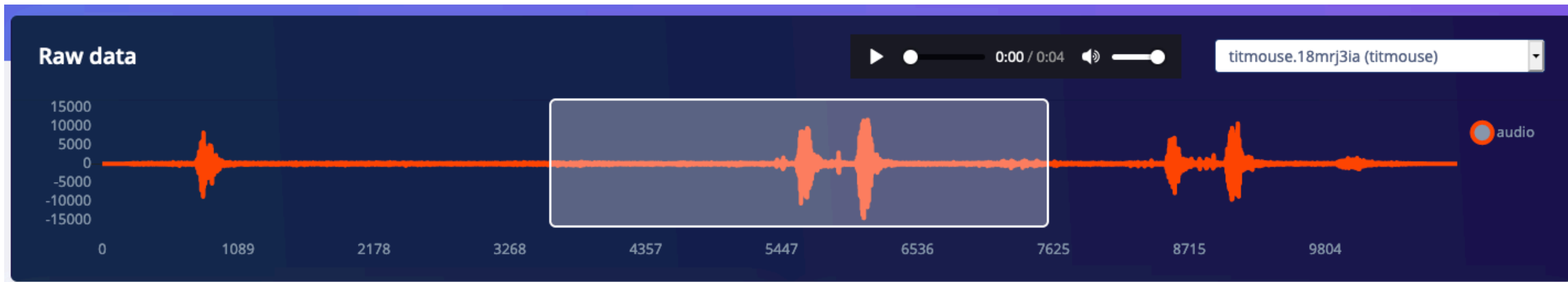
ML Sensor pipeline is often combination



Much more interesting



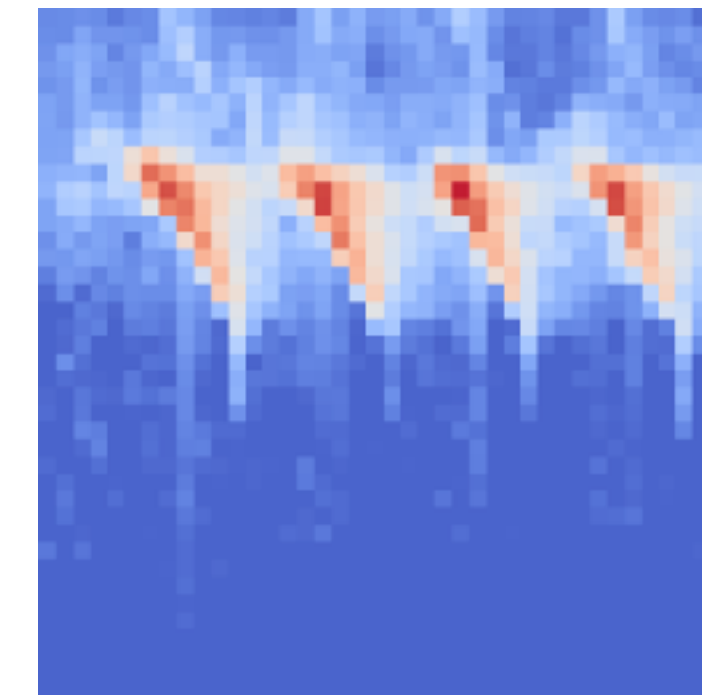
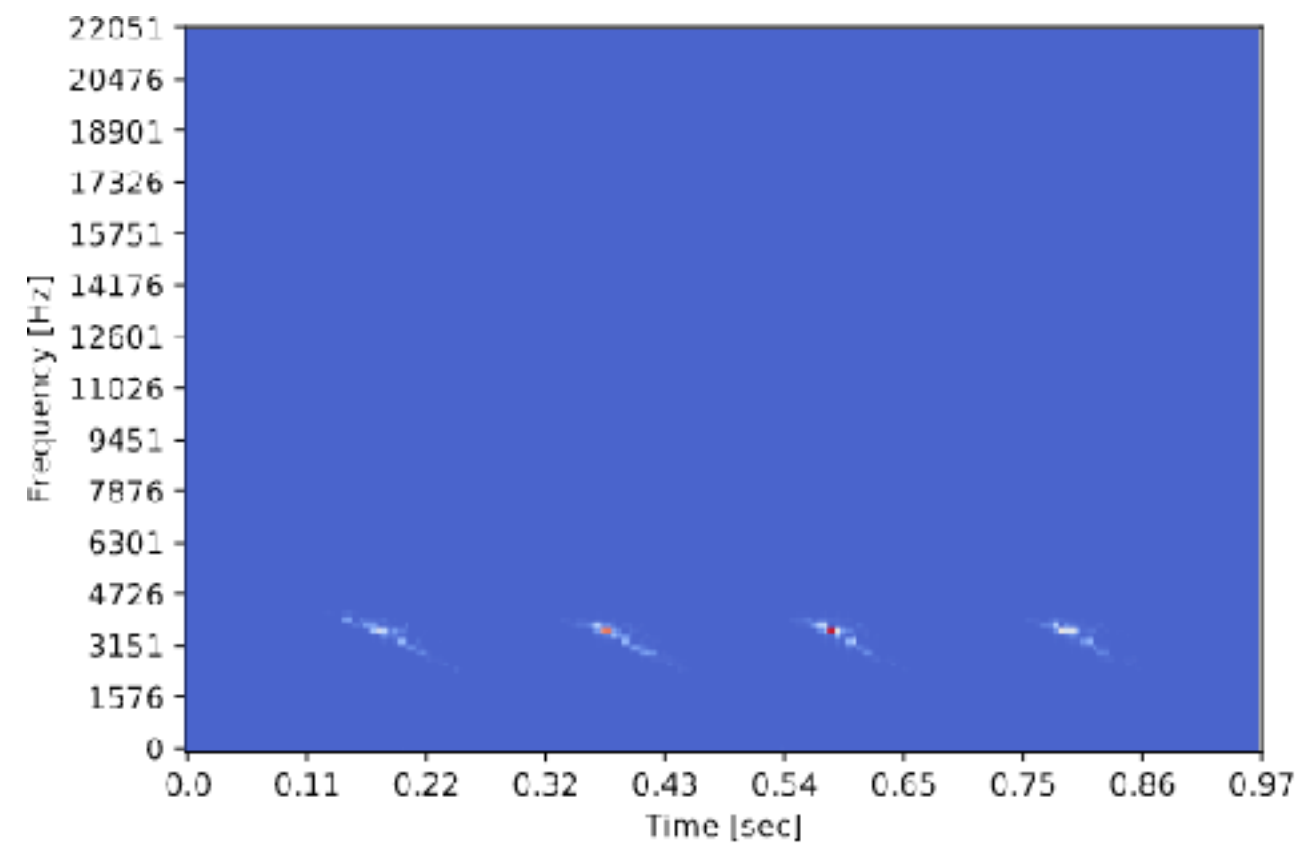
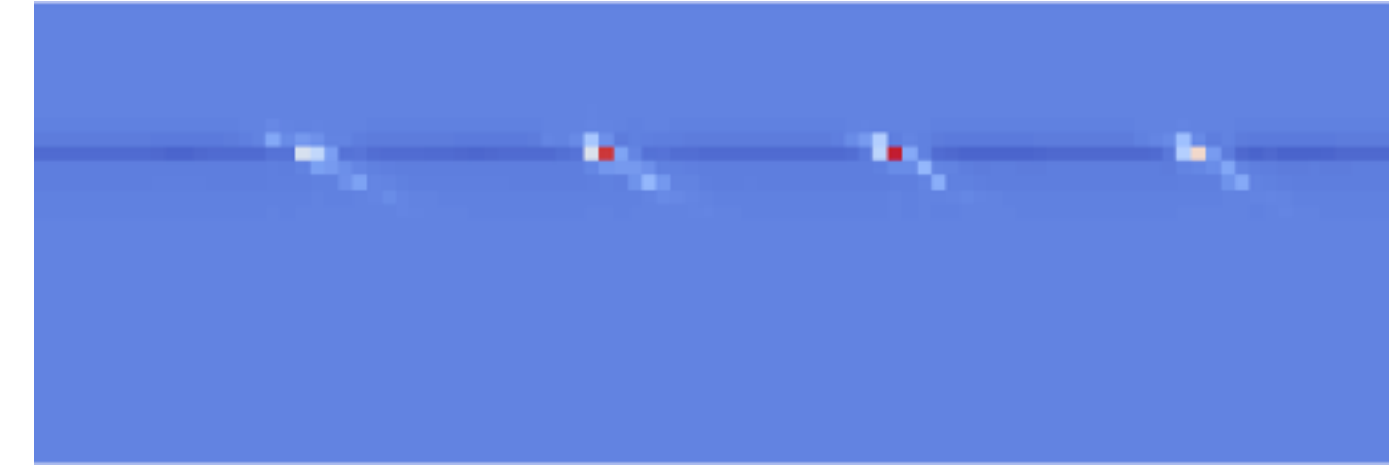
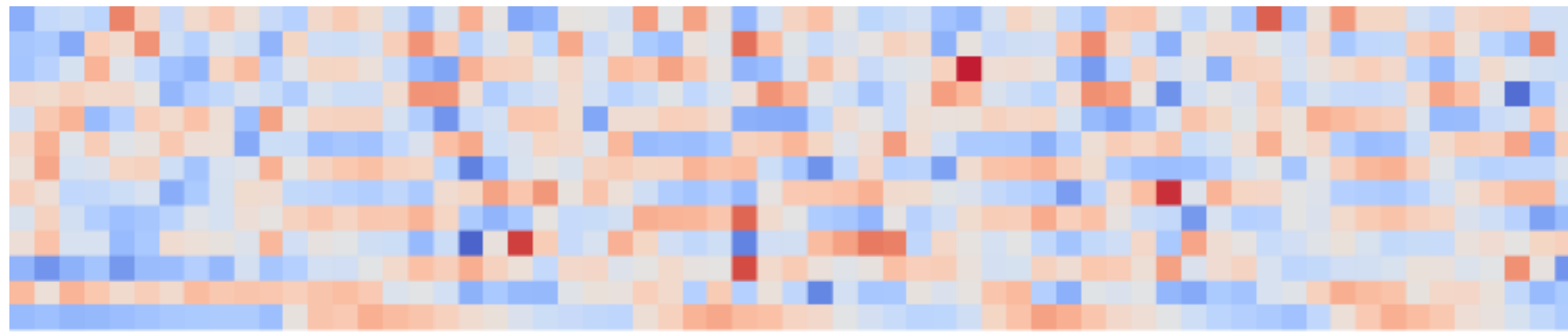
Wide range of parameters



Window length, window step, downsample?



Wide range of parameters



+ endless configuration options



Constrained targets - what's worth it?

EDGE IMPULSE SPECTROGRAM (BIRD SOUND CLASSIFIER) Jan Jongboom

Parameters Generate features

Raw data: dtmouse16khz.wav.27gf72mg (dtmouse1)

Raw features: 28, -41, -184, -32, -257, -137, 63, -54, -238, -228, -13, ...

Parameters: Spectrogram

- Frame length: 0.02
- Frame stride: 0.01
- Frequency bands: 128

Graph options: Show axes

Save parameters

On-device performance

- PROCESSING TIME: 77 ms.
- PEAK RAM USAGE: 27 KB

EDGE IMPULSE SPECTROGRAM (BIRD SOUND CLASSIFIER) Jan Jongboom

Parameters Generate features

Raw data: titmouse16khz.wav.27gf72mg (titmouse1)

Raw features: 28, -41, -184, -32, -257, -137, 63, -54, -238, -228, -13, ...

Parameters: Spectrogram

- Frame length: 0.02
- Frame stride: 0.01
- Frequency bands: 256

Graph options: Show axes

Save parameters

On-device performance

- PROCESSING TIME: 159 ms.
- PEAK RAM USAGE: 51 KB



Edge Impulse

The place to build embedded Machine Learning models (vision & non-vision)

From engineers (not data scientists!) for engineers

Every step of the way, from data collection to deployment

Already >38,000 real ML projects created!

Free for developers: edgeimpulse.com

The screenshot displays the Edge Impulse web interface for a project named 'Keyword spotting'. The interface is divided into several sections:

- Navigation Menu:** Includes Dashboard, Devices, Data acquisition, Impulse design, Retrain model, Live classification, Model testing, Versioning, Deployment, GETTING STARTED, Documentation, and Forums.
- Project Info:** Shows 'EdgeImpulse Inc. / Keyword spotting' and a user profile for 'Jan Jongboom'.
- Creating your first impulse (67% complete):** A progress bar with three main steps:
 - Acquire data:** 'Every Machine Learning project starts with data. You can capture data from a development board or your phone, or import data you already collected.' Includes a button 'LET'S COLLECT SOME DATA'.
 - Design an Impulse:** 'Train the model to interpret previously unseen data, based on historical data. Use this to categorize new data, or to find anomalies in sensor readings.' Includes three buttons: 'GETTING STARTED: CONTINUOUS MOTION RECOGNITION', 'GETTING STARTED: RECOGNIZE SOUNDS FROM AUDIO', and 'GETTING STARTED: ADDING SIGHT TO YOUR SENSORS'.
 - Deploy:** 'Package the complete impulse up, from signal processing code to trained model, and deploy it on your device. This ensures that the impulse runs with low latency and without requiring a network connection.' Includes a button 'DEPLOY YOUR MODEL'.
- Summary:** Shows 'DEVICES CONNECTED: 5' and 'DATA COLLECTED: 28m 46s'.
- Collaborators:** Lists users: Arjan Kamphuis, Daniel, Jan Jongboom (EDGE IMPULSE STAFF), and Zach Shelby (EDGE IMPULSE STAFF).
- Project Info:** Shows 'Project ID: 5' and 'Latency calculator: Cortex-M4F BOM'.
- Download block output:** Shows 'No downloads available yet'.



Introducing the EON Tuner!

EON Tuner

Config

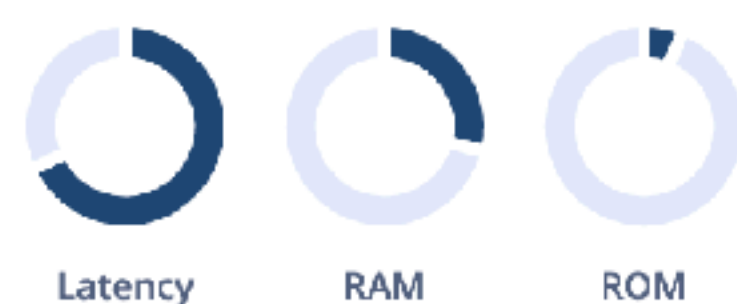
Logs

Finding the most optimal architecture for your model (17 model variants evaluated / 50 variants total)

cortex-m7-216mhz 1000 ms RAM: 128kB • ROM: 1024kB

97% 10x10 | SYNTIA | Clone | a65

PERFORMANCE



ACCURACY

coo	97	3	0	0	0	0
dis	2	94	4	0	0	0
eat	0	0	100	0	0	0
soc	4	0	2	93	0	0
vac	0	0	0	0	100	0

97% 10x10 | SYNTIA | Clone | 133

PERFORMANCE



ACCURACY

coo	98	2	0	0	0	0
dis	4	88	8	0	0	0
eat	0	0	100	0	0	0
soc	2	2	0	96	0	0
vac	0	0	0	0	100	0

96% 10x10 | SYNTIA | Clone | 2f1

PERFORMANCE



ACCURACY

coo	95	5	0	0	0	0
dis	2	92	4	2	0	0
eat	0	3	98	0	0	0
soc	2	0	2	96	0	0
vac	0	0	0	0	100	0

Filters

Status

- Pending
- Running
- Completed
- Failed

View

Data set

- Validation
- Train

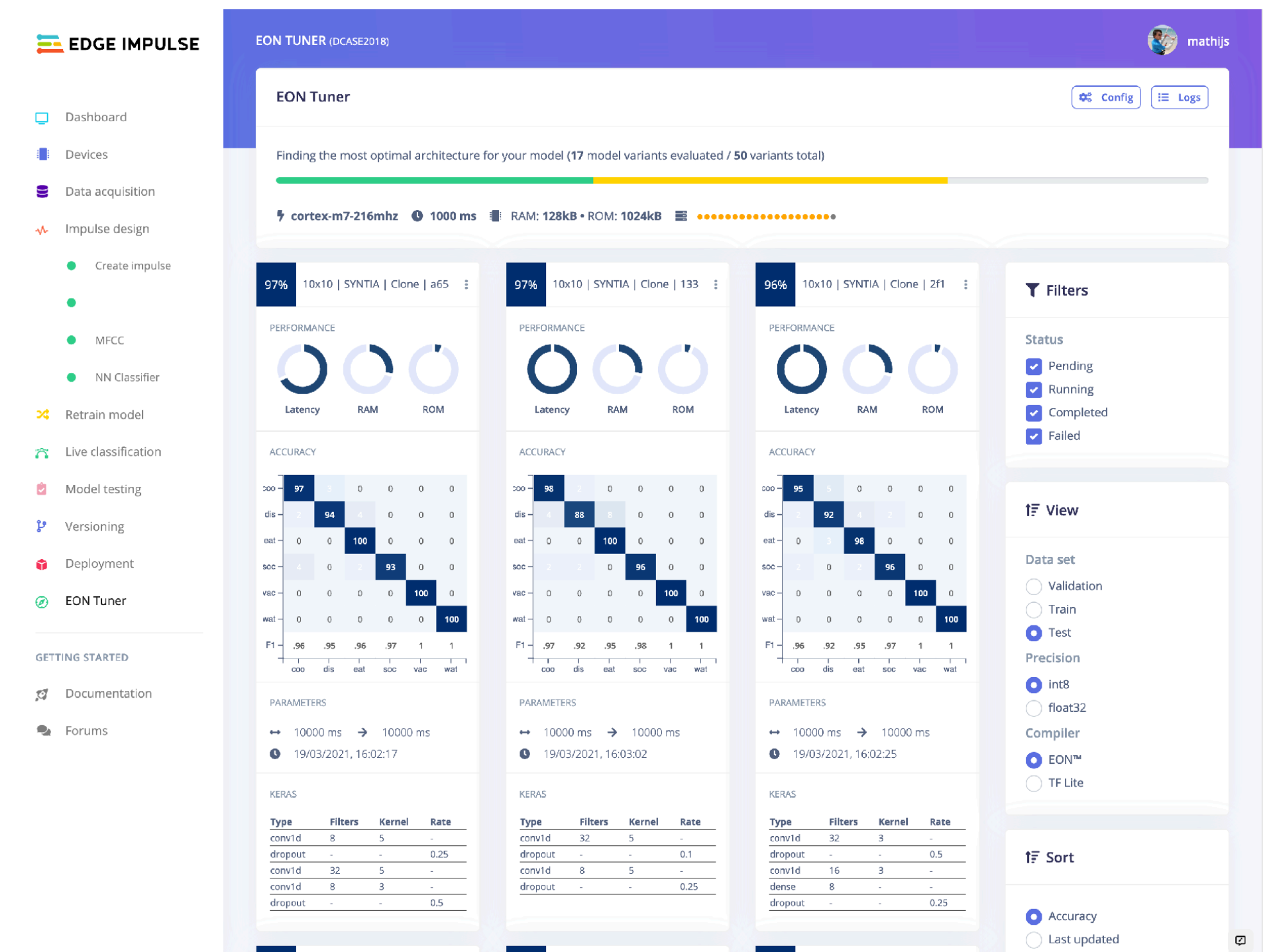
EON Tuner

Find best model for sensor data over mix of input blocks, DSP blocks and ML blocks

Specify device constraints

Extensible with your own DSP and ML blocks

Biggest win: "we found DSP configuration that works so well, we no longer need ML"



Still an engineering tool!

Model search complete! (100 model variants evaluated)

EON Tuner settings

Target | Advanced

Find the optimal architecture for your machine learning model

The EON™ Tuner will evaluate many candidate model architectures (selected based on your target device and latency requirements) concurrently to help you find the best performing architecture for your application.

The search process can take up to 24 hours to complete. We will notify you by e-mail on completion of the search process. While the search is running you can view the progress on this page at any time.

Dataset category

Speech (Keywords)

- Speech (Keywords)
 - Yes/no
 - Hello world
- Speech (Continuous)
 - Emotions
 - Speaker gender
- Audio (Events)
 - Breaking glass
 - Gunshot
- Audio (Continuous)
 - Household activities
 - Traffic level

? Help me choose the right dataset category

0.95 mfcc-keras-9ae

PERFORMANCE

250 ms | 128 kB | 1024 kB

Legend: DSP (Dark Blue), NN (Light Blue), Unused (Grey)

ACCURACY

hel	98	2	0
nol	2	92	6
unk	7	5	88
F1	.95	.93	.91

INPUT

↔ 1000 ms | → 1000 ms

MFCC

↔ 0.02 | → 0.01 | ≡ 13 | ≡ 32

KERAS

Type	Filters	Kernel	Rate
conv1d	8	5	-
dropout	-	-	0.25
conv1d	8	5	-
dropout	-	-	0.25

01/05/2021, 02:33:56

0.95 mfcc-keras-efe

PERFORMANCE

250 ms | 128 kB | 1024 kB

Legend: DSP (Dark Blue), NN (Light Blue), Unused (Grey)

ACCURACY

hel	97	0	3
nol	2	91	7
unk	6	3	91
F1	.95	.94	.9

INPUT

↔ 1000 ms | → 1000 ms

MFCC

↔ 0.04 | → 0.02 | ≡ 13 | ≡ 32

KERAS

Type	Filters	Kernel	Rate
conv1d	8	3	-
dropout	-	-	0.1
conv1d	16	5	-
dense	8	-	-
dropout	-	-	0.25

01/05/2021, 02:35:14



Questions it can answer for vision

What transfer learning block to use?

Grayscale or RGB? Resolution?

Preprocessing of data (e.g. edge finding using CV) yes or no?

Within the constraints of your device



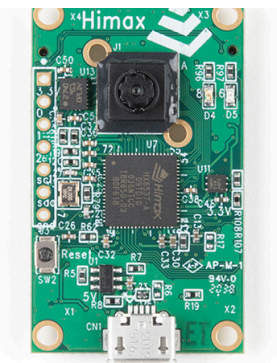
Getting started

<https://docs.edgeimpulse.com/docs>

Very wide range of dev boards, from Cortex-M4F to Jetson Nano

Deploy to any device that has a C++ compiler

Or use your phone!



Demo

www.edgeimpulse.com



 EDGE IMPULSE

SEPT 29 - OCT 01

Imagine

The future of data-driven
engineering starts now.

Recap

The ML hype is real

ML + sensors = perfect fit

Let's make those billions of devices see the world!

edgeimpulse.com



Questions?

Full docs:

<https://docs.edgeimpulse.com>

Performance metrics:

<https://docs.edgeimpulse.com/docs/inference-performance-metrics>

Adding sight to your sensors:

<https://docs.edgeimpulse.com/docs/image-classification>

Object detection:

<https://docs.edgeimpulse.com/docs/object-detection>

More questions:

forum.edgeimpulse.com / jan@edgeimpulse.com



Empowering Product Creators to Harness Edge AI and Vision



The Edge AI and Vision Alliance (www.edge-ai-vision.com) is a partnership of 100+ leading edge AI and vision technology and services suppliers, and solutions providers

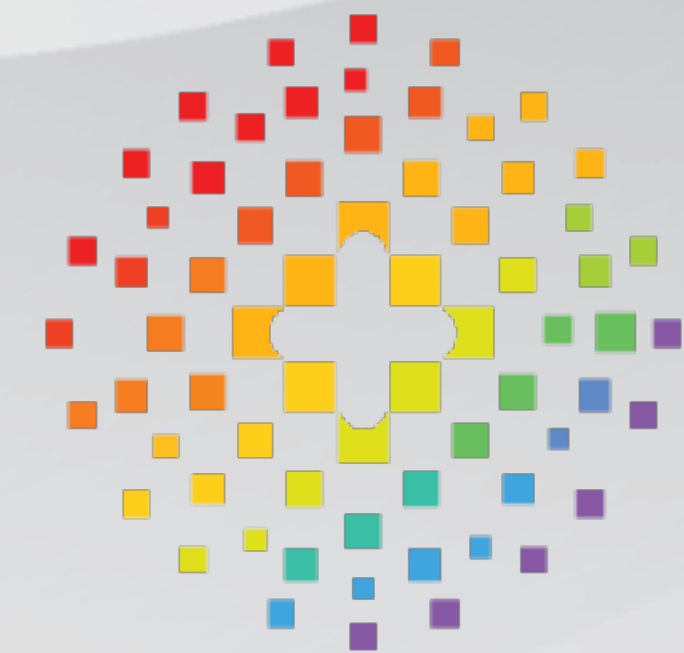
Mission: To inspire and empower engineers to design products that perceive and understand.

The Alliance provides low-cost, high-quality technical educational resources for product developers

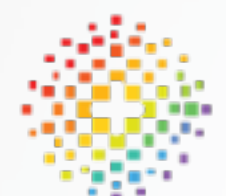
Register for updates at www.edge-ai-vision.com

The Alliance enables edge AI and vision technology providers to grow their businesses through leads, partnerships, and insights

For membership, email us: membership@edge-ai-vision.com



edge ai + vision
ALLIANCE™



Join us at the Embedded Vision Summit

May 16-19, 2022—Santa Clara, California



The only industry event focused on practical techniques and technologies for system and application creators

- *“Awesome! I was very inspired!”*
- *“Fantastic. Learned a lot and met great people.”*
- *“Wonderful speakers and informative exhibits!”*

Embedded Vision Summit 2022 highlights:

- **Inspiring keynotes** by leading innovators
- High-quality, practical **technical, business and product talks**
- Exciting **demos, tutorials** and **expert bars** of the latest applications and technologies



Visit www.EmbeddedVisionSummit.com to learn more

