tinyML® Talks

Enabling Ultra-low Power Machine Learning at the Edge

"An Introduction to TinyML for all backgrounds with hands on introduction to Edge Impulse"

Peter Ing - Edge Impulse

September 24, 2021

www.tinyML.org

tinyML Talks Sponsors and Strategic Partners

AONdevIces

tinyML Strategic Partner

arm

tinyML Strategic Partner

Deeplite

EDGE IMPULSE

tinyML Strategic Partner

emza visual sense

tinyML Strategic Partner

GREENWAVES TECHNOLOGIES

tinyML Strategic Partner

LatentAI
Adaptive AI for a Smarter Edge

tinyML Strategic Partner

HOTG

tinyML Strategic Partner

imagimob

tinyML Strategic Partner

maxim integrated

NOW PART OF ANALOG DEVICES

Qeexo

tinyML Strategic Partner

Qualcomm

tinyML Strategic Partner

RealityAI®

tinyML Strategic Partner

seeed studio
The IoT Hardware Enabler

tinyML Strategic Partner

SensiML™

tinyML Strategic Partner

SynSense

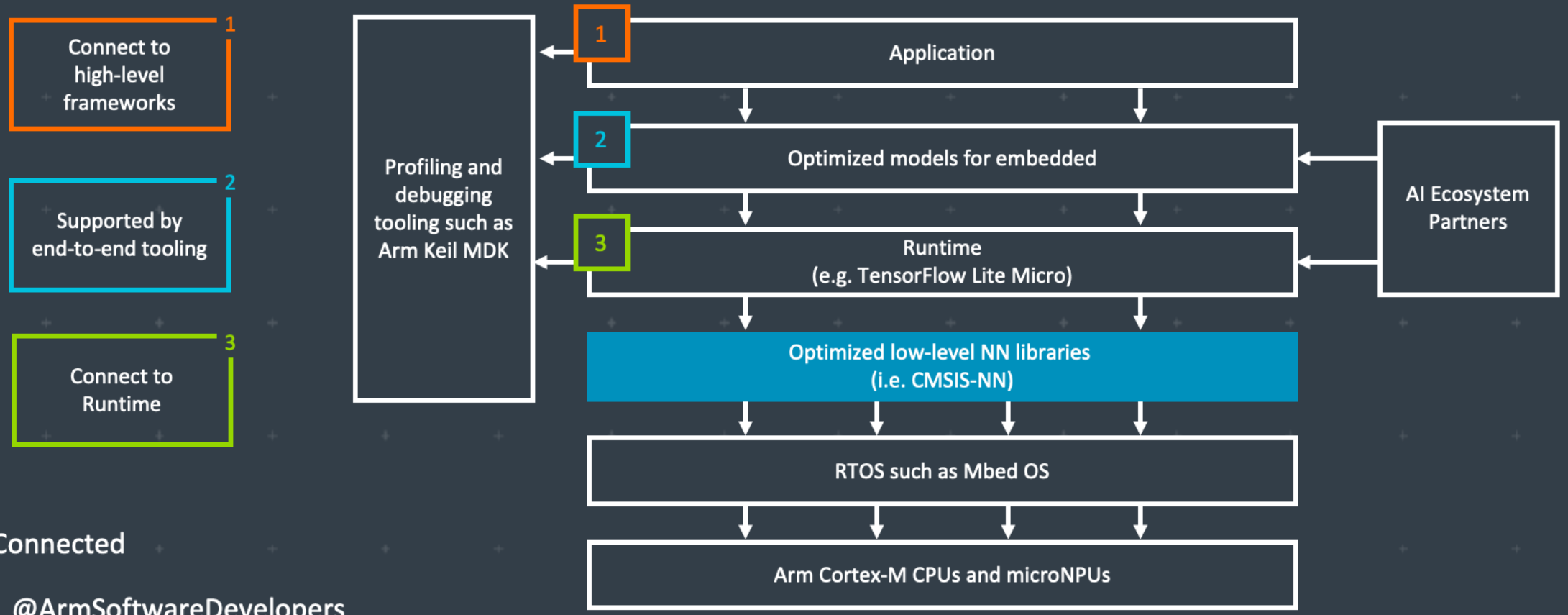
tinyML Strategic Partner

SYNTIANT

tinyML Strategic Partner

Additional Sponsorships available – contact Olga@tinyML.org for info

Arm: The Software and Hardware Foundation for tinyML



Stay Connected

@ArmSoftwareDevelopers

@ArmSoftwareDev

Resources: developer.arm.com/solutions/machine-learning-on-arm

arm

Deeplite

WE USE AI TO MAKE OTHER AI FASTER, SMALLER AND MORE POWER EFFICIENT

Automatically compress SOTA models like MobileNet to <200KB with **little to no drop in accuracy** for inference on resource-limited MCUs

Reduce model optimization trial & error from weeks to days using Deeplite's **design space exploration**

Deploy more models to your device without sacrificing performance or battery life with our **easy-to-use software**

BECOME BETA USER **bit.ly/testdeeplite**

mobilityXlab arm AI 100 CBINSIGHTS

TinyML for all developers

C++ library

Arduino library

WebAssembly

Dataset

Acquire valuable training data securely

Enrich data and train ML algorithms

Edge Device

Real sensors in real time

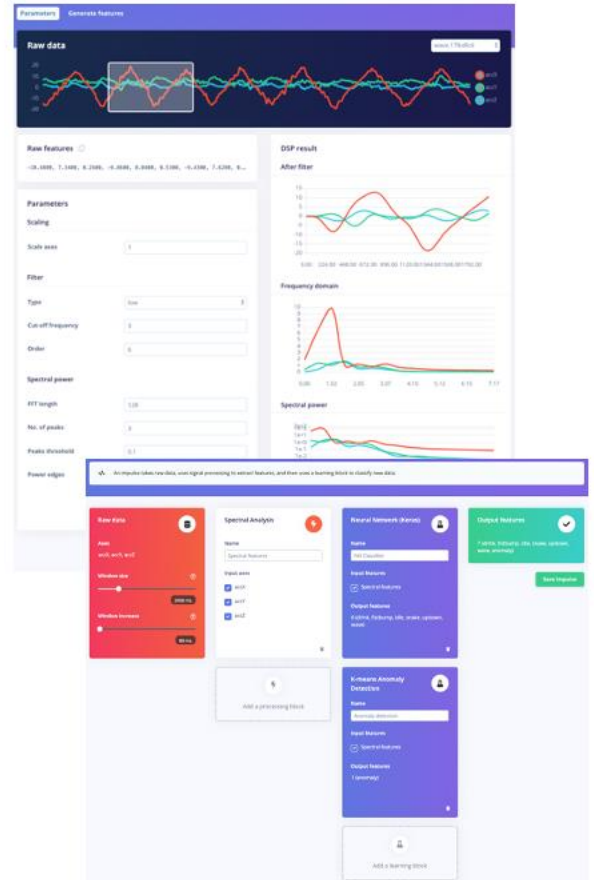
Open source SDK

Impulse

Embedded and edge compute deployment options

Test impulse with real-time device data flows

Test

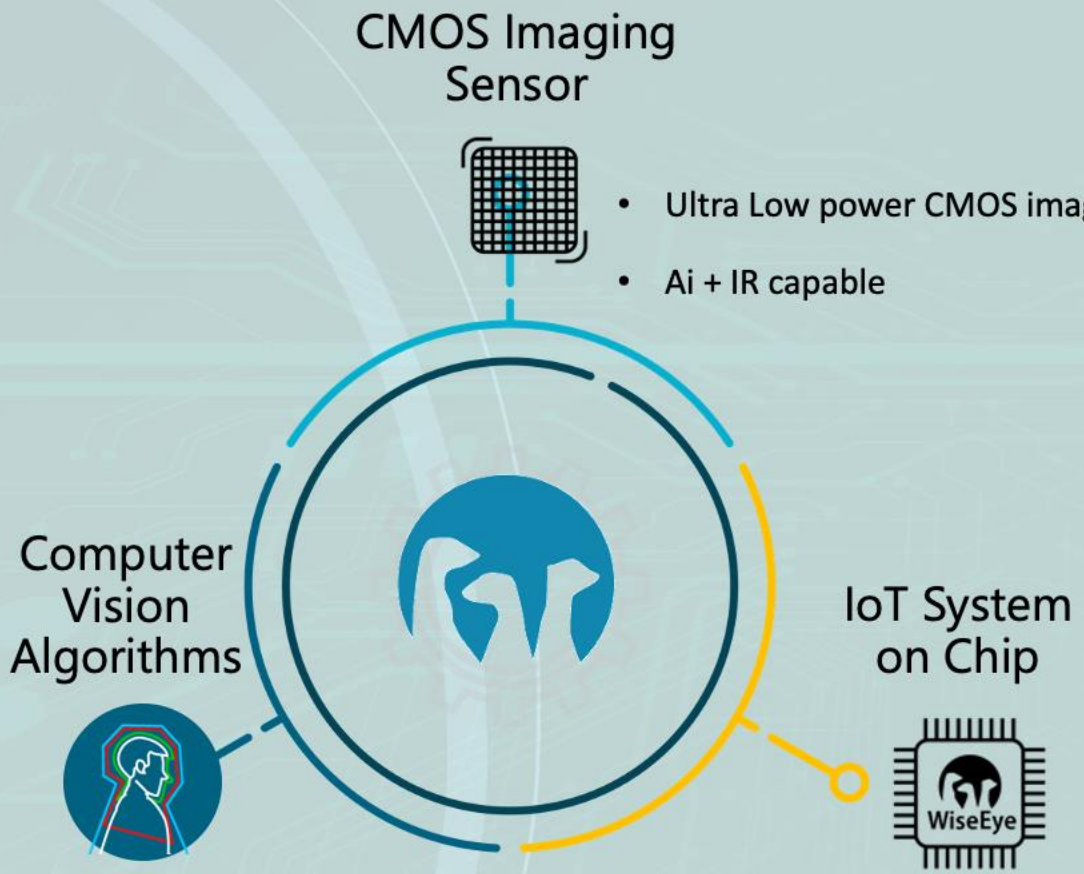


www.edgeimpulse.com

emza
visual sense

The Eye in IoT

Edge AI Visual Sensors



- Ultra Low power CMOS imager
- Ai + IR capable

- Machine Learning algorithm
- <1MB memory footprint
- Microcontrollers computing power
- Trained algorithm
- Processing of low-res images
- Human detection and other classifiers

- Machine Learning edge computing silicon
- <1mW always-on power consumption
- Computer Vision hardware accelerators

info@emza-vs.com

Enabling the next generation of Sensor and Hearable products to process rich data with energy efficiency

Visible Image

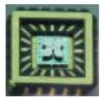
Sound

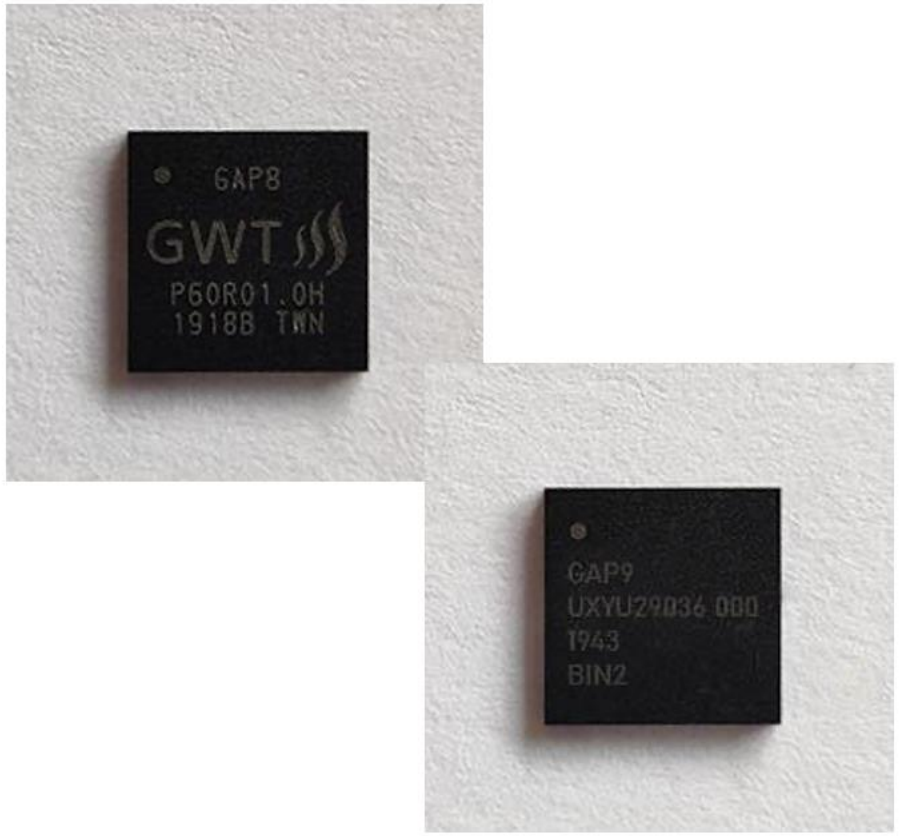
IR Image

Radar

Bio-sensor

Gyro/Accel

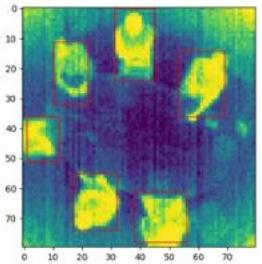




Wearables / Hearables

Battery-powered consumer electronics

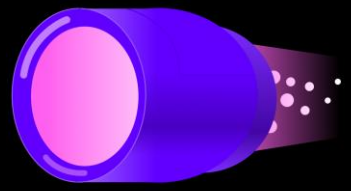
IoT Sensors



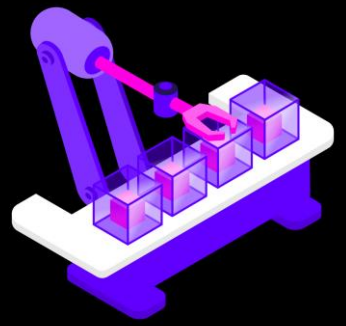
GREENWAVES TECHNOLOGIES

Distributed infrastructure for TinyML apps

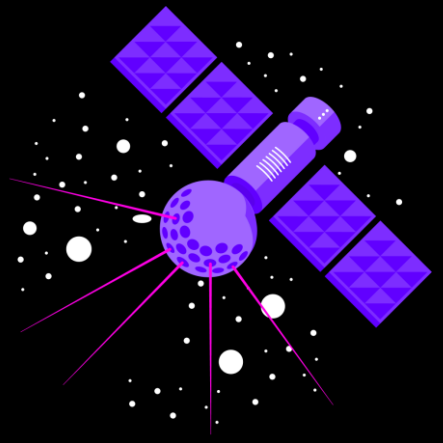
HOTG
Decoupling intelligence



Develop at warp speed



Automate deployments



Device orchestration

HOTG is building the distributed infrastructure to pave the way for AI enabled edge applications

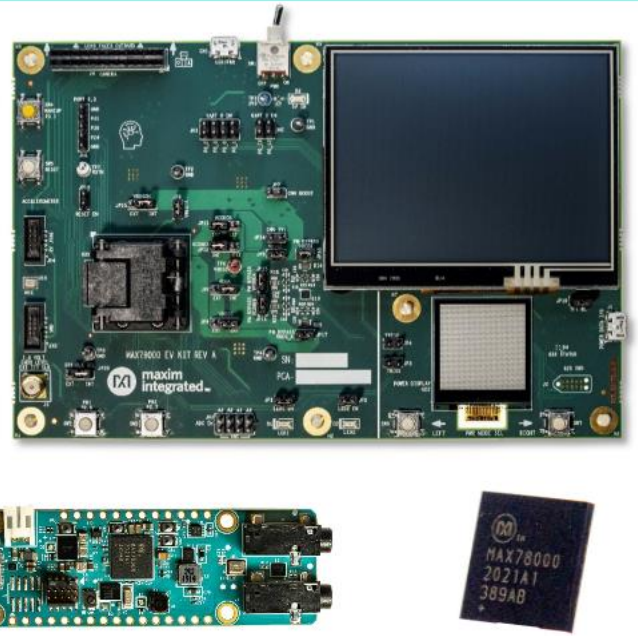
LatentAI

Adaptive AI for the Intelligent Edge

Latentai.com

Maxim Integrated: Enabling Edge Intelligence

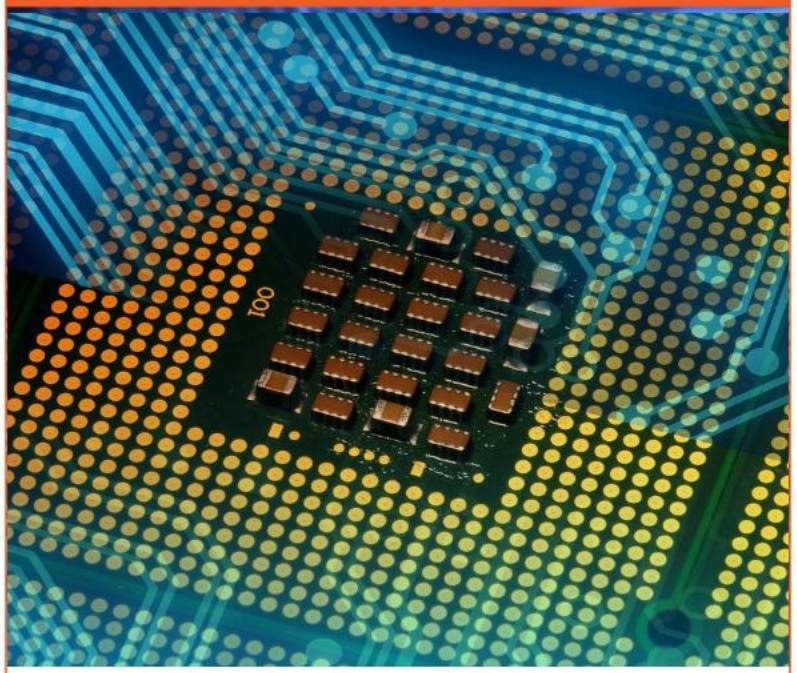
Advanced AI Acceleration IC



The new MAX78000 implements AI inferences at low energy levels, enabling complex audio and video inferencing to run on small batteries. Now the edge can see and hear like never before.

www.maximintegrated.com/MAX78000

Low Power Cortex M4 Micros



Large (3MB flash + 1MB SRAM) and small (256KB flash + 96KB SRAM, 1.6mm x 1.6mm) Cortex M4 microcontrollers enable algorithms and neural networks to run at wearable power levels.

www.maximintegrated.com/microcontrollers

Sensors and Signal Conditioning

Health sensors measure PPG and ECG signals critical to understanding vital signs. Signal chain products enable measuring even the most sensitive signals.

www.maximintegrated.com/sensors

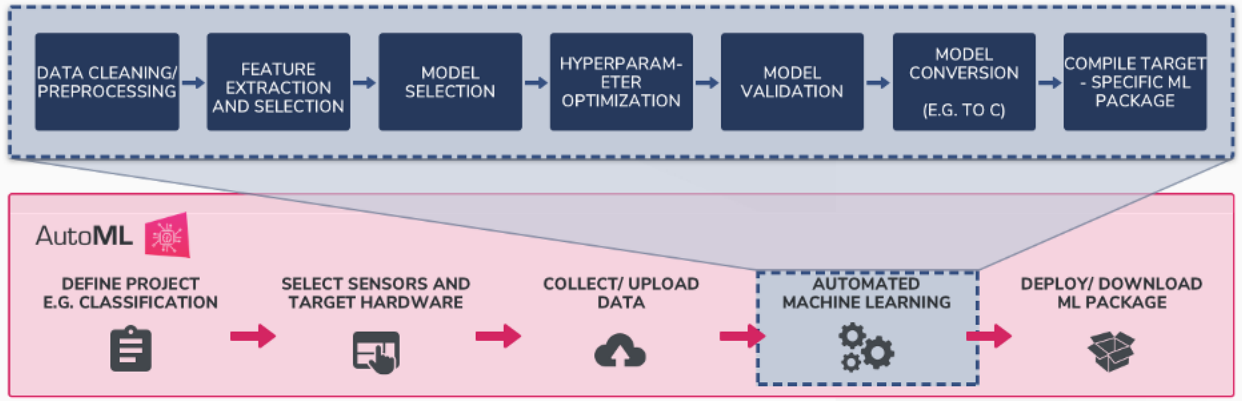
Qeexo AutoML

Automated Machine Learning Platform that builds tinyML solutions for the Edge using sensor data

Key Features

- Supports 17 ML methods:
  - Multi-class algorithms: GBM, XGBoost, Random Forest, Logistic Regression, Gaussian Naive Bayes, Decision Tree, Polynomial SVM, RBF SVM, SVM, CNN, RNN, CRNN, ANN
  - Single-class algorithms: Local Outlier Factor, One Class SVM, One Class Random Forest, Isolation Forest
- Labels, records, validates, and visualizes time-series sensor data
- On-device inference optimized for low latency, low power consumption, and small memory footprint applications
- Supports Arm® Cortex™- M0 to M4 class MCUs

End-to-End Machine Learning Platform



For more information, visit: www.qeexo.com

Target Markets/Applications

- Industrial Predictive Maintenance
- Smart Home
- Wearables
- Automotive
- Mobile
- IoT

Qualcomm
AI research

Advancing AI research to make efficient AI ubiquitous

Power efficiency

Model design, compression, quantization, algorithms, efficient hardware, software tool

Personalization

Continuous learning, contextual, always-on, privacy-preserved, distributed learning

Efficient learning

Robust learning through minimal data, unsupervised learning, on-device learning

A platform to scale AI across the industry

Perception

Object detection, speech recognition, contextual fusion

Reasoning

Scene understanding, language understanding, behavior prediction

Action

Reinforcement learning for decision making

Edge cloud

Cloud

IoT/IIoT

Automotive

Mobile

Reality AI®

Add Advanced Sensing to your Product with Edge AI / TinyML

https://reality.ai info@reality.ai @SensorAI Reality AI

Pre-built Edge AI sensing modules, plus tools to build your own

Reality AI solutions

Prebuilt sound recognition models for indoor and outdoor use cases

Solution for industrial anomaly detection

Pre-built automotive solution that lets cars "see with sound"

Reality AI Tools® software

Build prototypes, then turn them into real products

Explain ML models and relate the function to the physics

Optimize the hardware, including sensor selection and placement

SensiML™

Build Smart IoT Sensor Devices From Data

SensiML pioneered TinyML software tools that auto generate AI code for the intelligent edge.

- End-to-end AI workflow
- Multi-user auto-labeling of time-series data
- Code transparency and customization at each step in the pipeline

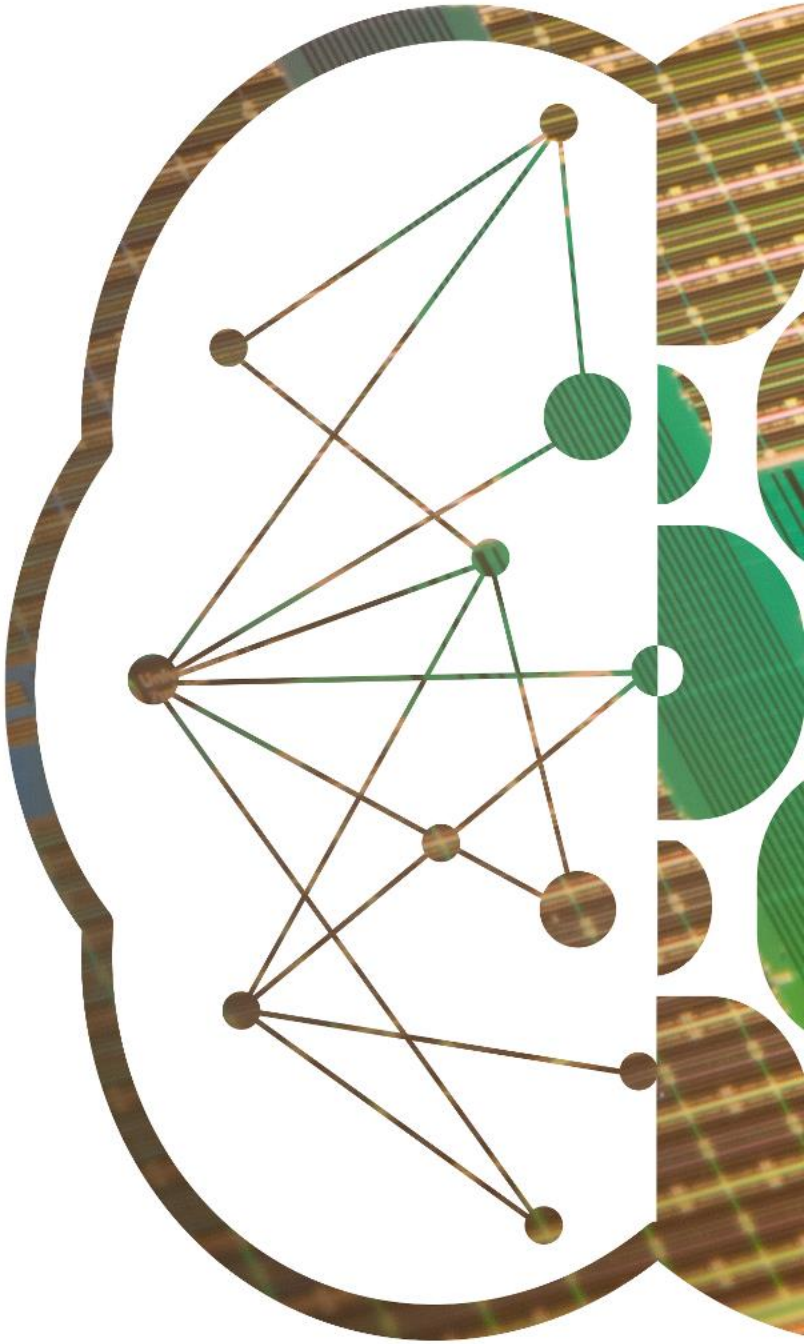
We enable the creation of production-grade smart sensor devices.

sensiml.com

SynSense

SynSense builds **sensing and inference** hardware for **ultra-low-power** (sub-mW) **embedded, mobile and edge** devices. We design systems for **real-time always-on smart sensing**, for audio, vision, IMUs, bio-signals and more.

https://SynSense.ai

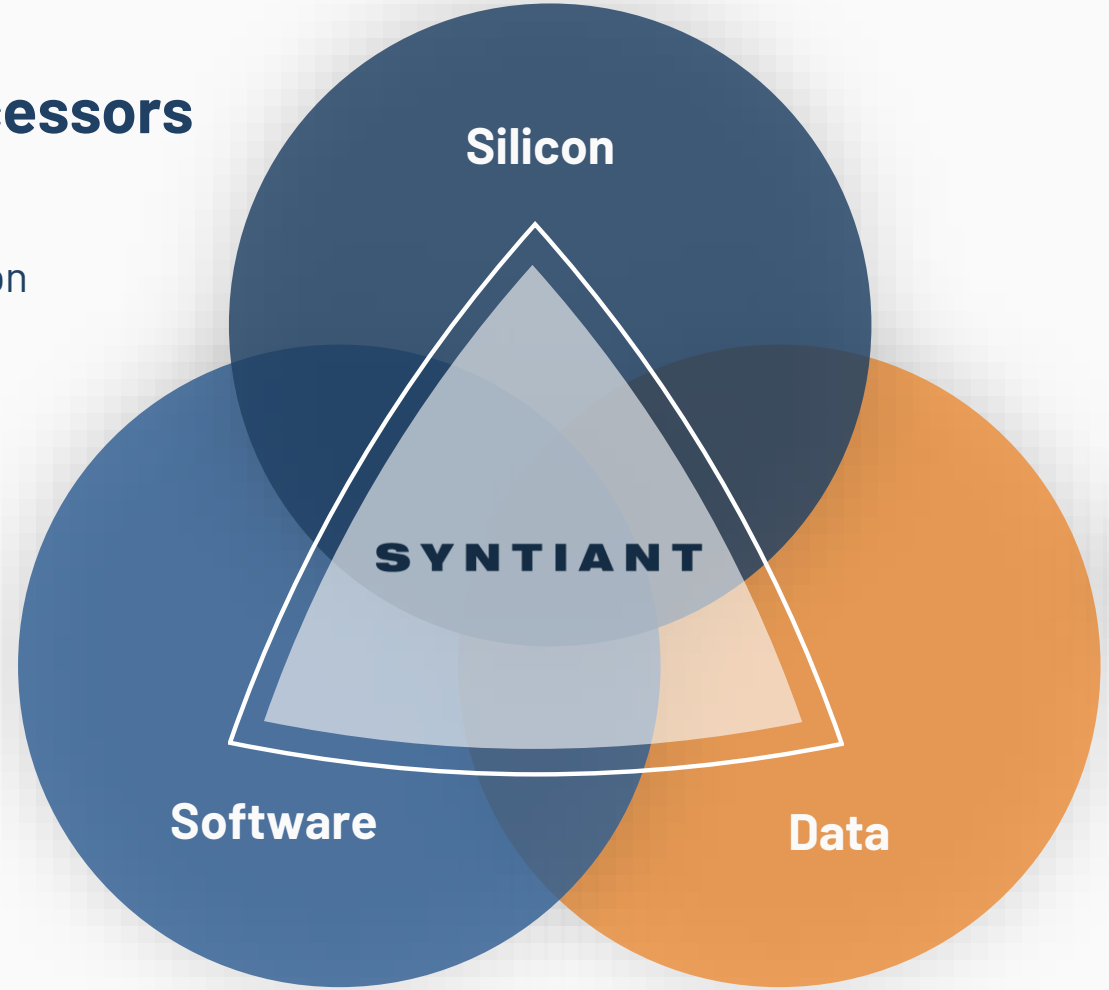


SYNTIANT

End-to-End Deep Learning Solutions for TinyML & Edge AI

Neural Decision Processors

- At-Memory Compute
- Sustained High MAC Utilization
- Native Neural Network Processing



ML Training Pipeline

- Enables Production Quality Deep Learning Deployments

Data Platform

- Reduces Data Collection Time and Cost
- Increases Model Performance

partners@syntiant.com

www.syntiant.com

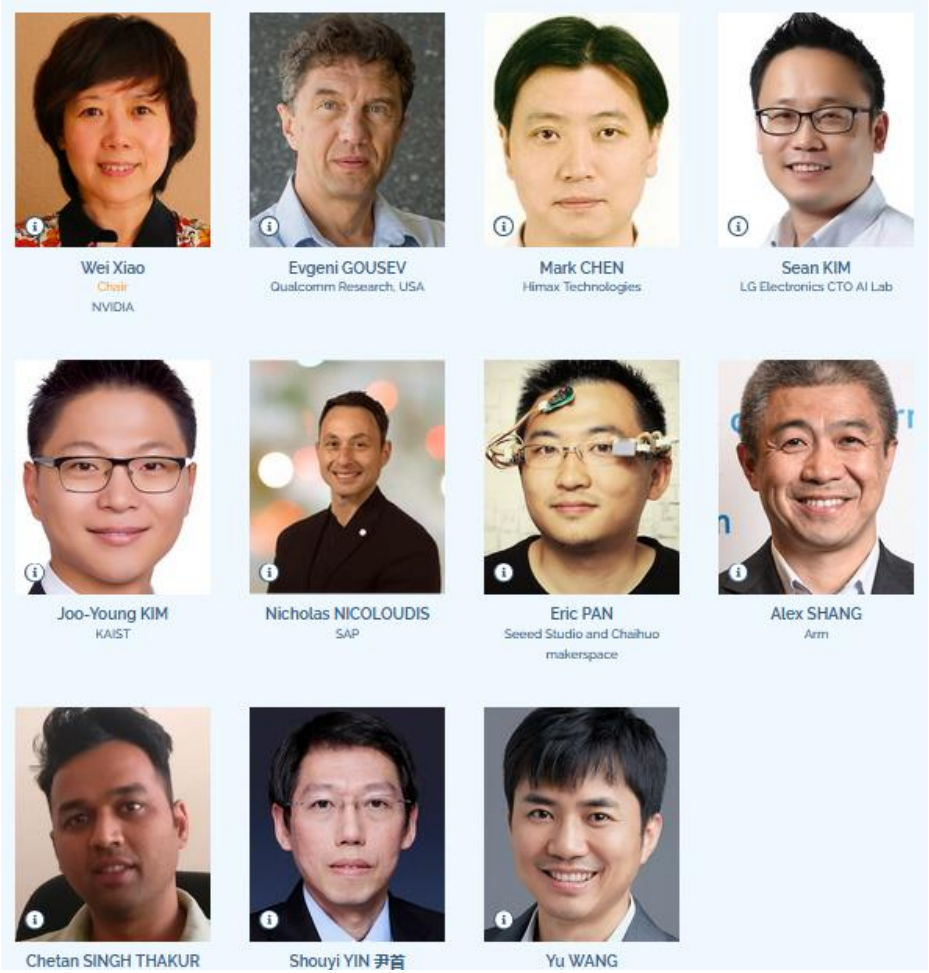
TINY ML ASIA

LIVE ONLINE **November 2-5, 2021**

(9-11:30 am China Standard time)

https://www.tinyml.org/event/asia-2021/

Technical Programm Committee



Register today!

Free event courtesy of our sponsors and strategic partners

arm · EDGE IMPULSE · emza visual sense · GREENWAVES TECHNOLOGIES

HOTG · imagimob · LatentAI Adaptive AI for a Smarter Edge · Qualcomm

Qeexo · RealityAI · seeed studio The IoT Hardware Enabler · SensiML

SynSense · SYNTIANT

More sponsorships are available: sponsorships@tinyML.org

collaboration with hackster.io

AN AVNET COMMUNITY

Focus on:
(i) developing new use cases/apps for tinyML vision; and (ii) promoting tinyML tech & companies in the developer community

Submissions accepted until September 17th, 2021

Winners announced on October 5th, 2021 ($6k value)

Sponsorships available: *sponsorships@tinyML.org*

https://www.hackster.io/contests/tinyml-vision

Next tinyML Talks

Date	Presenter	Topic / Title
Tuesday, September 28	**Marios Fournarakis,** Qualcomm Technologies, Netherlands	A Practical Guide to Neural Network Quantization

Webcast start time is 8 am Pacific time

Please contact talks@tinyml.org if you are interested in presenting

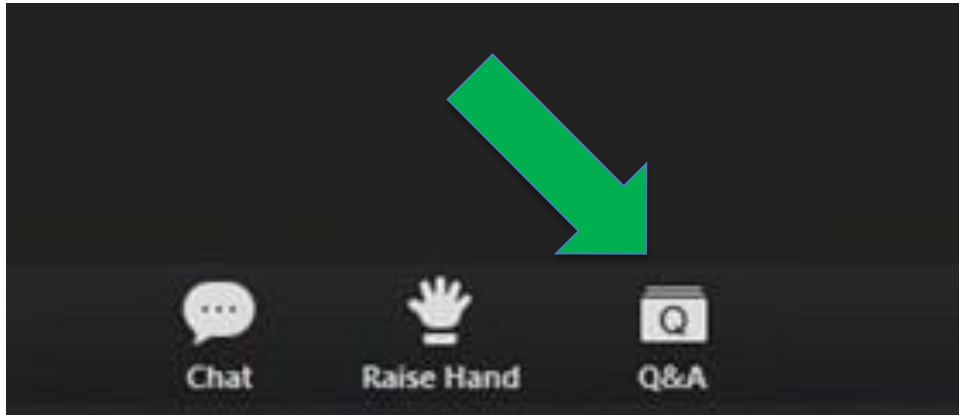
Reminders

Slides & Videos will be posted tomorrow

tinyml.org/forums

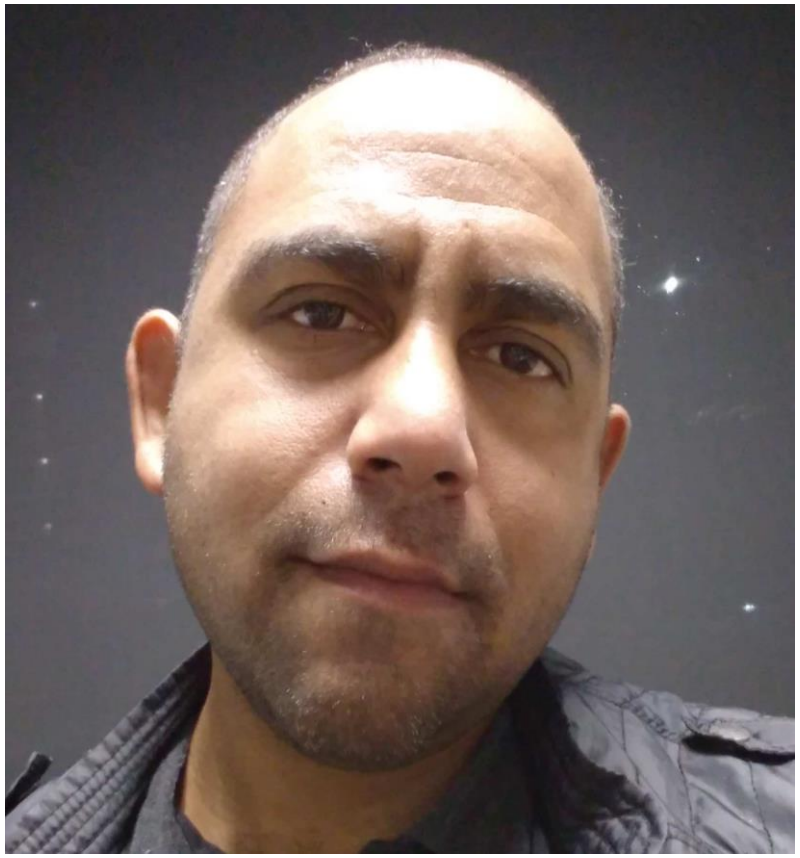
youtube.com/tinyml

Please use the Q&A window for your questions



TINY ML TALKS webcast

Peter Ing



Peter is from Cape Town and sees all technology and science as continuum but had to select one discipline and choose to complete a NDip in Electrical Engineering. He has tinkered in many different areas and has worked formally in the Retail, Transport and Automotive sectors integrating different systems and technologies together. His work interests and experience include Embedded Systems, Industrial Automation, IoT and software development and more recently Machine Learning which is what makes TinyML the ideal landing point.

TINY ML TALKS webcast

Getting Started with ML in General

TINY ML TALKS webcast

Formalized in 1956

Research in 60s & 70s

AI Winter

Revival in 2000s

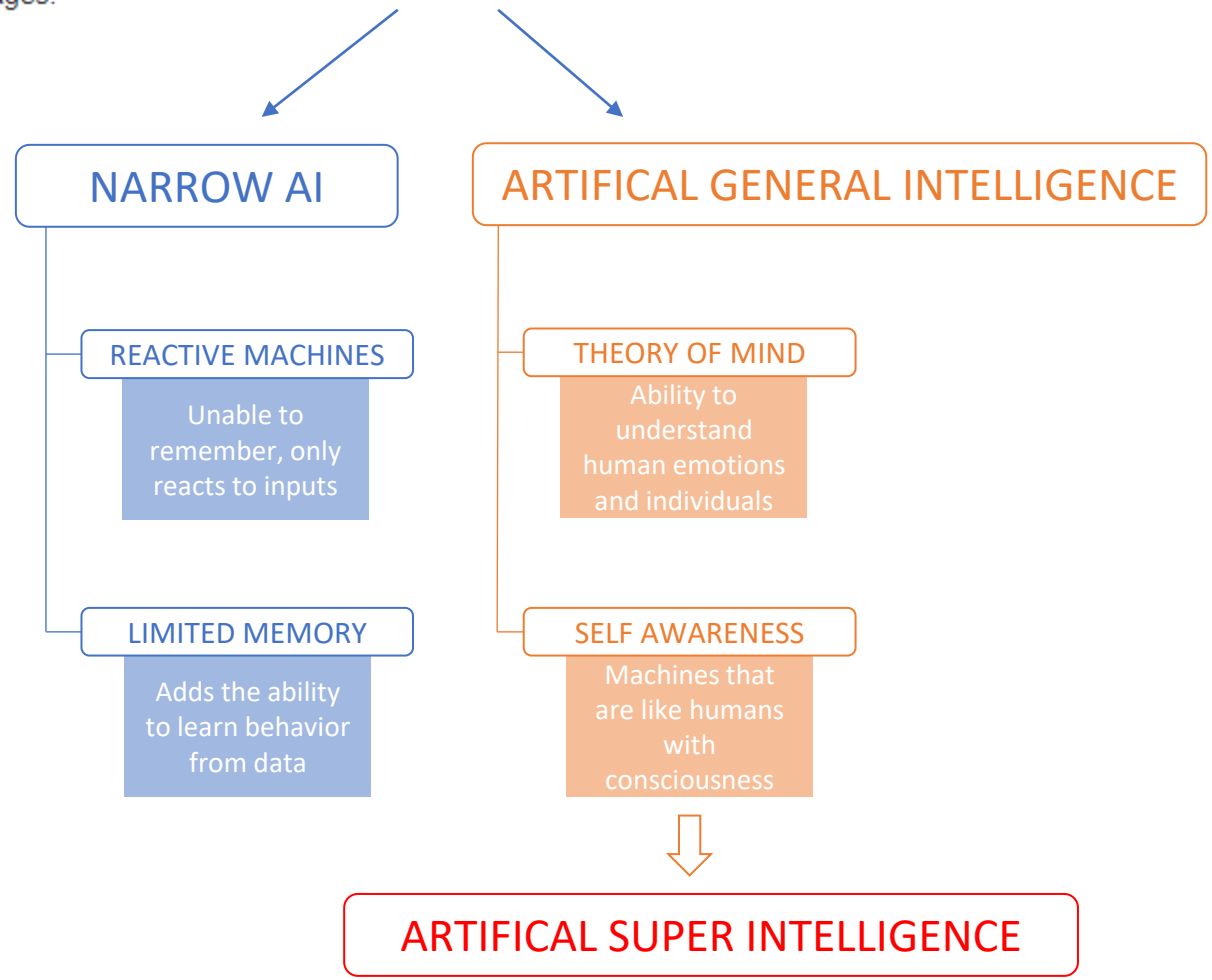
Low Cost Computing & Big Data

AI Spring

artificial intelligence

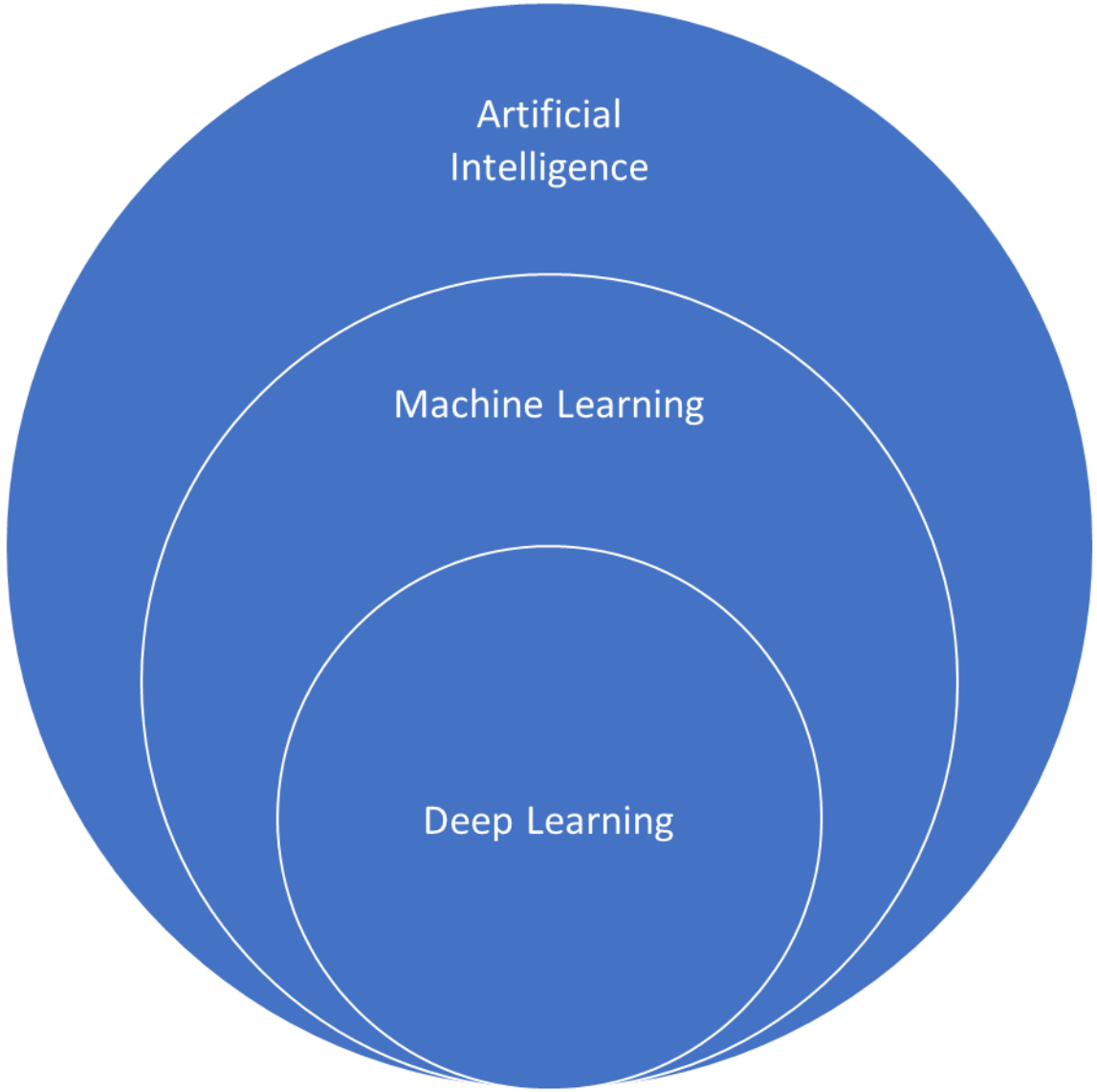
noun

the theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages.



TINY ML TALKS webcast

AI vs ML vs DL



TINY ML TALKS webcast

ARTIFICIAL INTELLIGENCE

MACHINE LEARNING

FUZZY LOGIC
EXPERT SYSTEMS
SYMBOLIC
…

SUPERVISED LEARNING

Trained against dataset:

- Regression
- Classification

UNSUPERVISED LEARNING

No training, processes input data:

- Dimensionality reduction
- Clustering
- Anomaly detection

REINFORCMENT LEARNING

Agent learns optimal behavior through positive and negative rewards i.e reinforcing desired reaction to environment

DEEP LEARNING

Artificial Neural Networks

- *Perceptrons*
- *CNN*
- *RNN…*

Displaying intelligence – Learning from Experience

DETECTING PEOPLE IN A SCENE

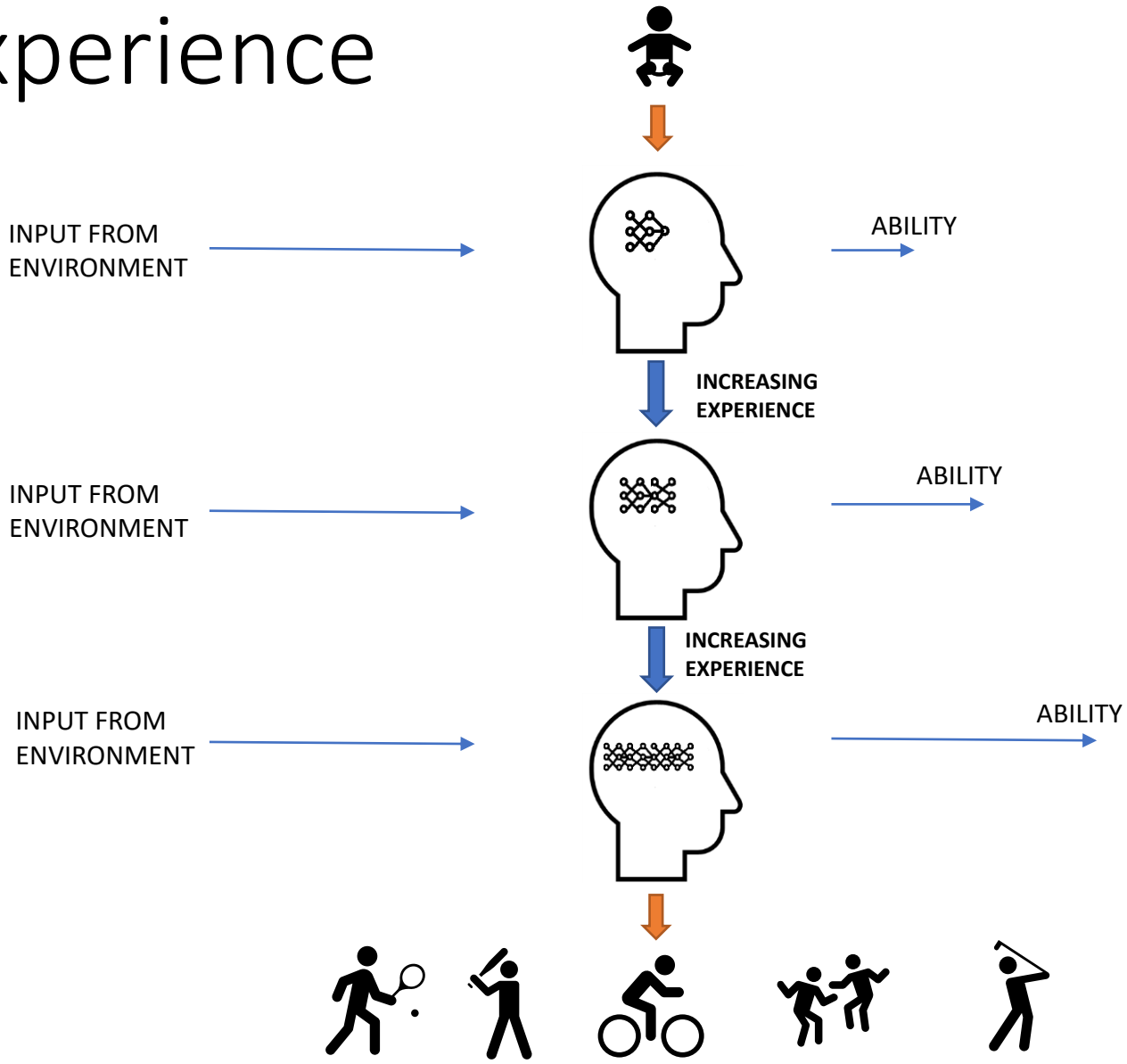
RECOGNIZING SPOKEN WORDS

DANGER

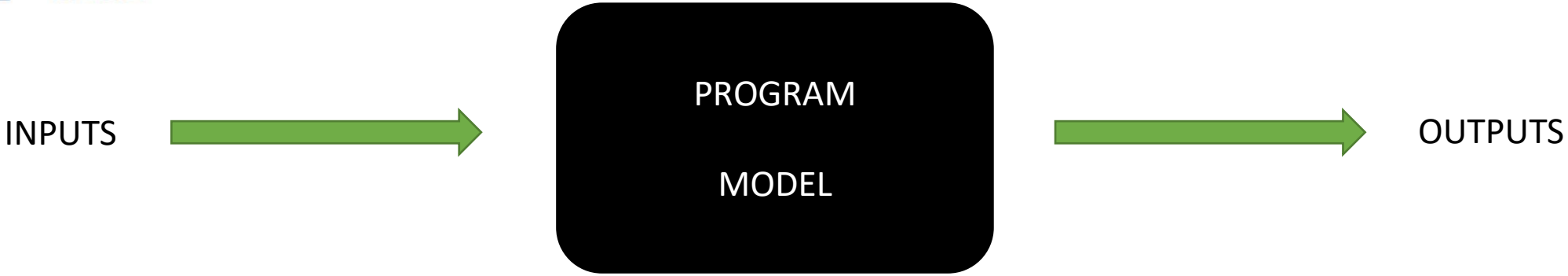
RECOGNIZING WHEN SOMETHING IS ABOUT TO FAIL :

- Sound cues
- Visual cues
- Temperature changes

SENSES



Machine Learning – Programs that learn



LEARNING – Training the model

- Mapping Inputs to Outputs
- Change Internal structure(weight values/parameters)
- Iterative process – Epochs
- Training dataset and Test dataset
- Hyperparameters – control training process
- Overfitting

INFERENCE - Executing program/model on new data

- Running the model
- Generalizing and making predictions on new data
- Output confidence

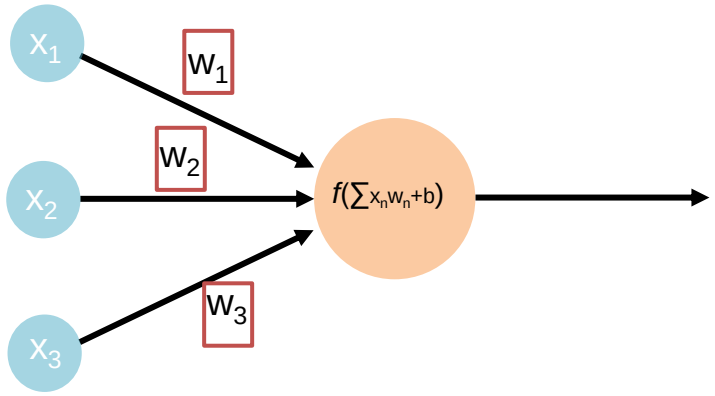
"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E."

-Tom Mitchell

tinyML = Inference on low power (<1mW) on small i.e. Tiny embedded devices

TINY ML TALKS webcast

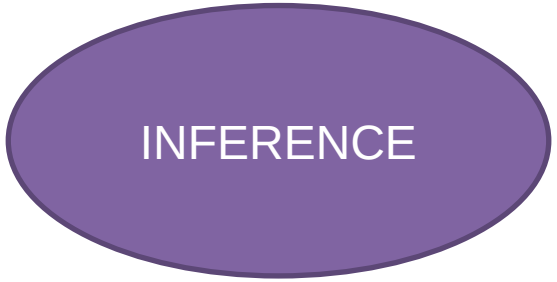
Classical Machine Learning & Deep Learning

CLASSICAL	DEEP LEARNING
$y = \beta_0 + x\beta_1$	
○ Based on statistics ○ Parameters are learned during training	○ Based on Artificial Neurons ○ Deep Learning – multiple hidden layers ○ Weights are learned during training

COMPUTATIONAL COMLPLEXITY

TINY ML TALKS webcast

Software Frameworks

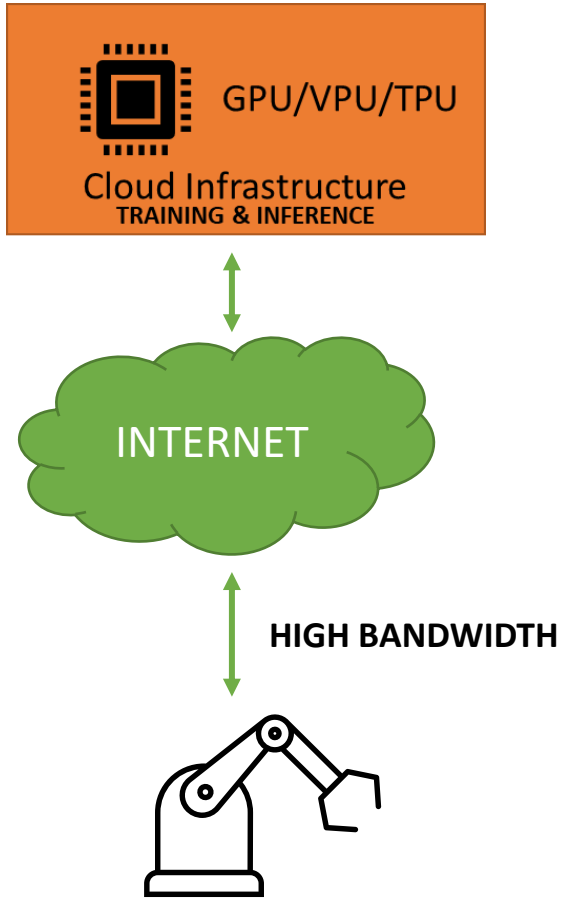


PyTorch

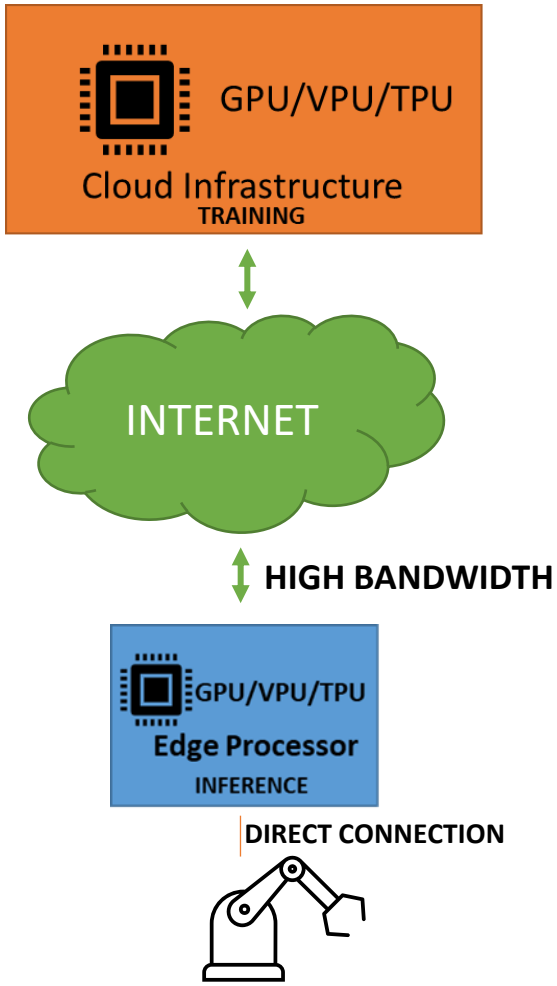
TensorFlow

TINY ML TALKS webcast

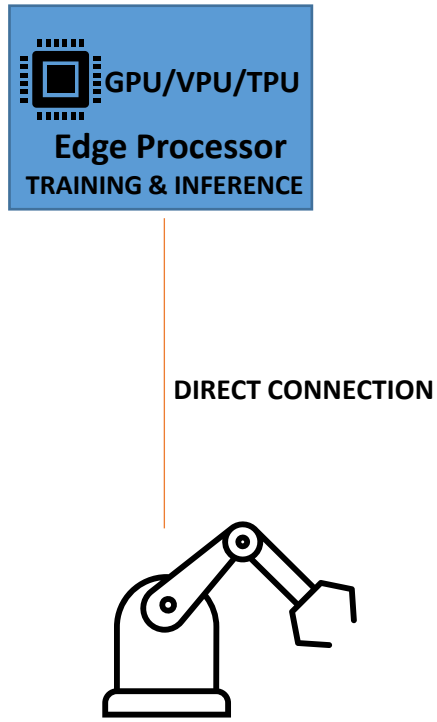
Typical AI/ML Architectures



CLOUD AI



CLOUD & EDGE AI



EDGE AI

TINY ML TALKS webcast

Overview of the Edge

CLOUD

Backend Processing
Storage/Database
Web application/Platform

INTERNET

EDGE

Gateways
Aggregation/Internet Connectivity
Computing close to application
Storage/Preprocessing

BLUETOOTH/ZIGBEE/THREAD/LORA

THINGS

Sensors
Processing capability
Internet Connectivity

Image Source: Wikipedia , Creative commons

TINY ML TALKS webcast

Hardware Spectrum

MPU VS MCU

MPU
MICROPROCESSOR UNIT

1. Operating System Support
2. External memory and peripherals(PCB)
3. High Power Consumption
4. Memory in Megabytes to Gigabytes
5. Multicore/Multitasking
6. Not always Realtime
7. Optional GPU

Raspberry Pi
(Single Board Computer)

MCU
MICROCONTROLLER UNIT

1. Barebones/No Operating System/RTOS
2. Memory and Peripherals included
3. Low Power Consumption
4. Memory in Kilobytes to Megabytes
5. Single Core/Dual Core – limited multitasking
6. Realtime & DSP capabilities
7. Lower Cost

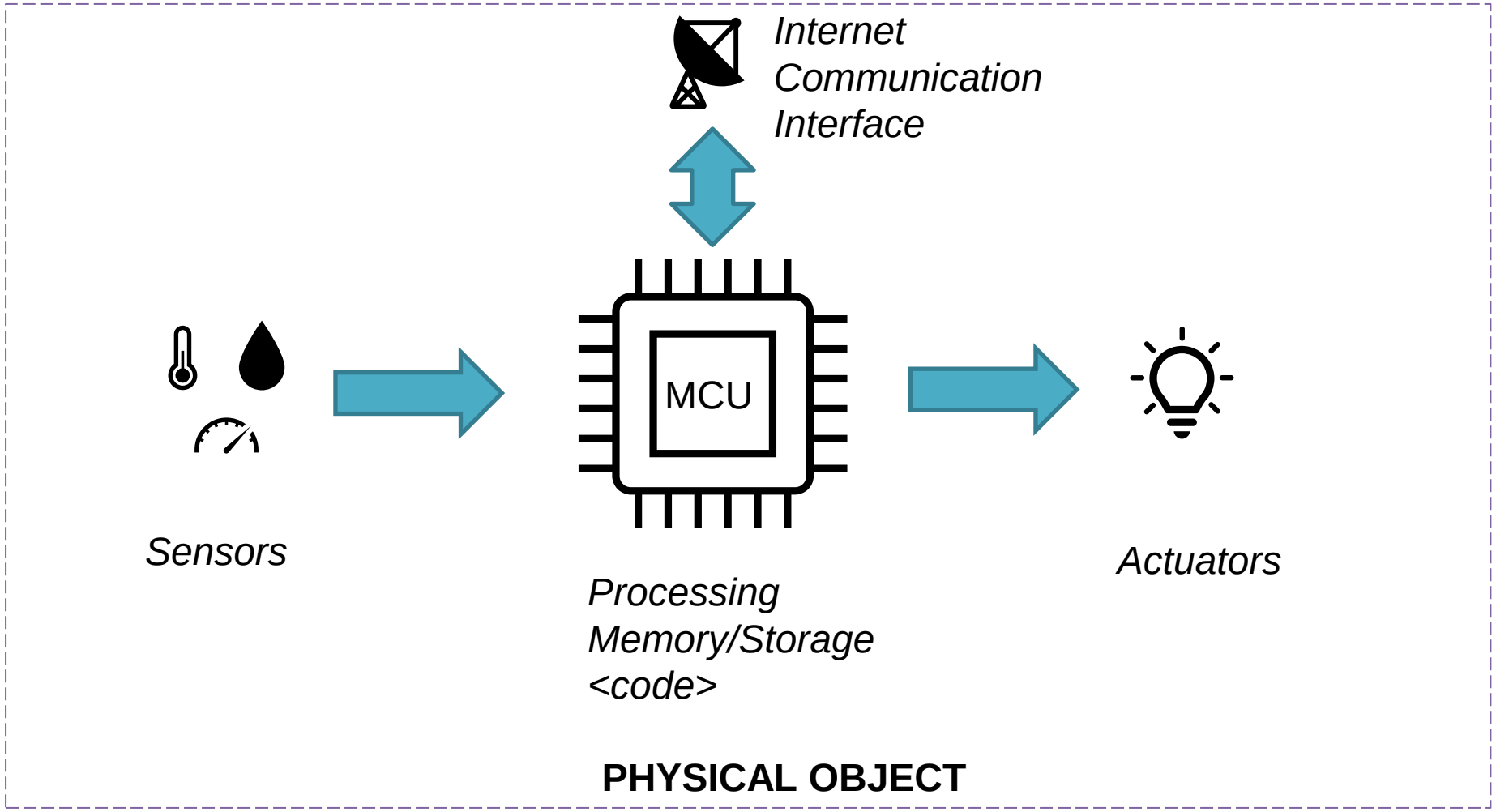
Arduino
(Embedded Platform)

COST/COMPLEXITY/POWER

Image Source: Wikipedia

TINY ML TALKS webcast

What are Things

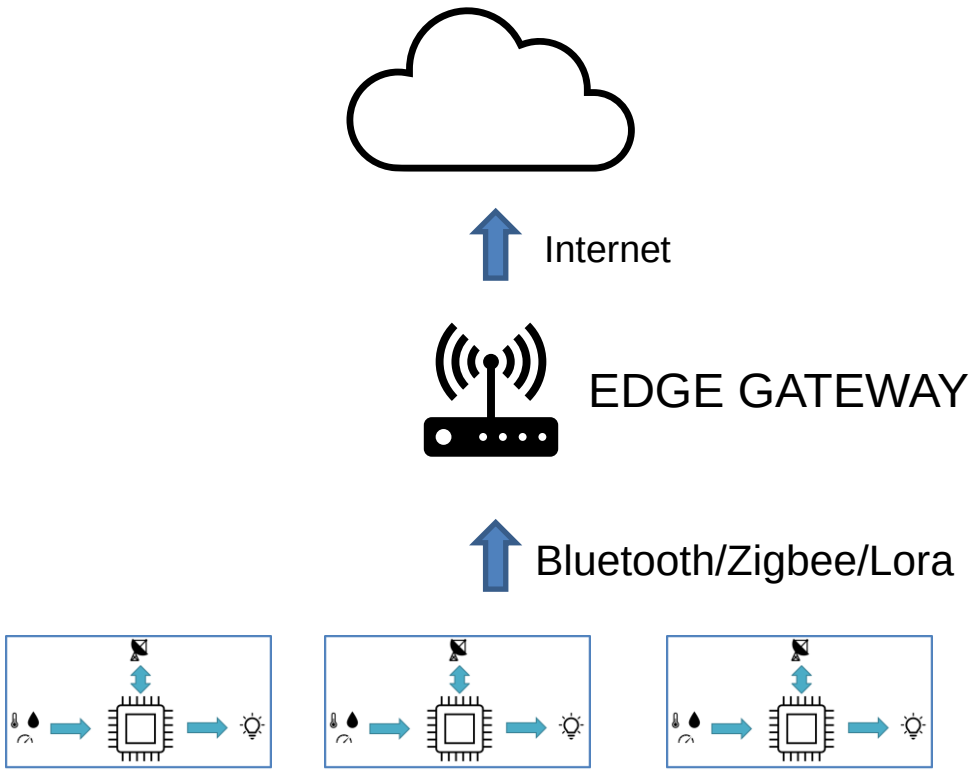


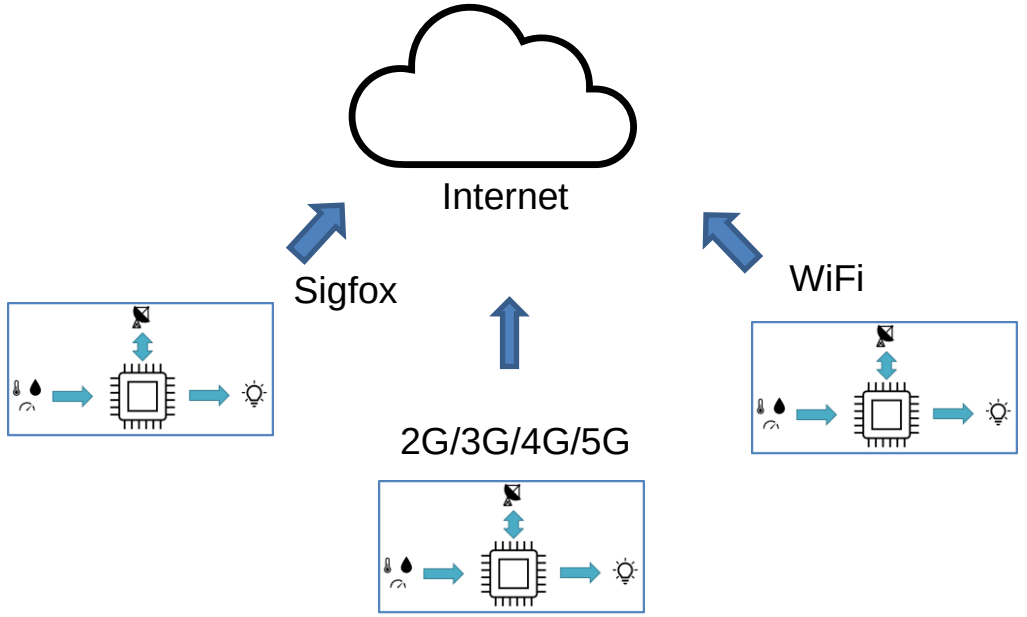
Turning objects into intelligent “Things”

TINY ML TALKS webcast

What is the Internet of Things

Physical things connected to the internet



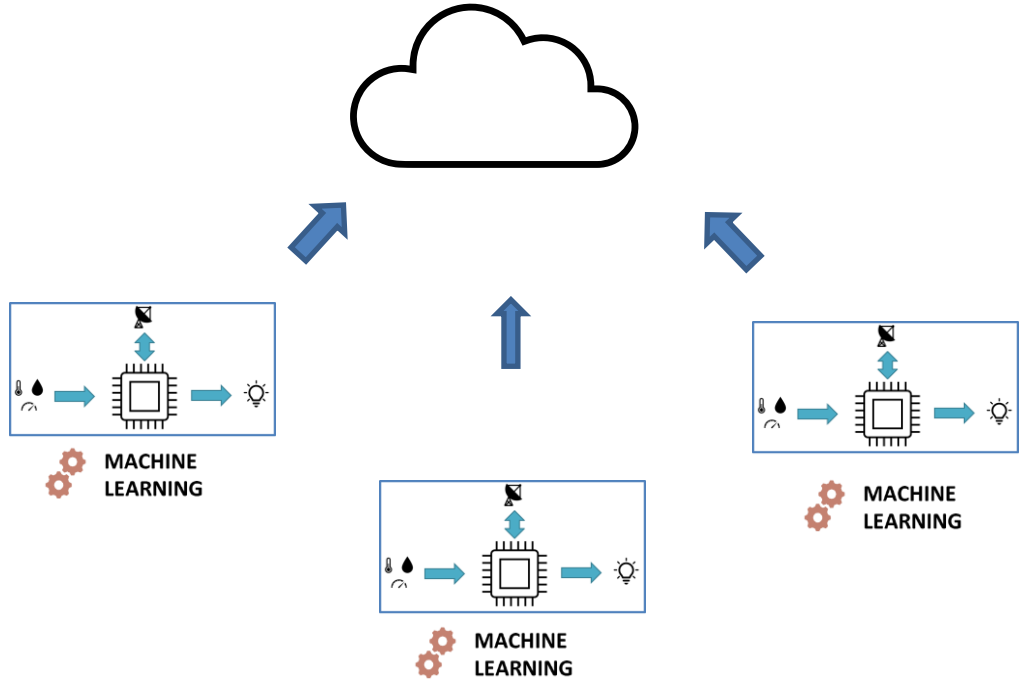


TINY ML TALKS webcast

Artificial Intelligence of Things

AI + IoT = AIoT

TensorFlow Lite



TINY ML

Inference on Edge and IoT devices

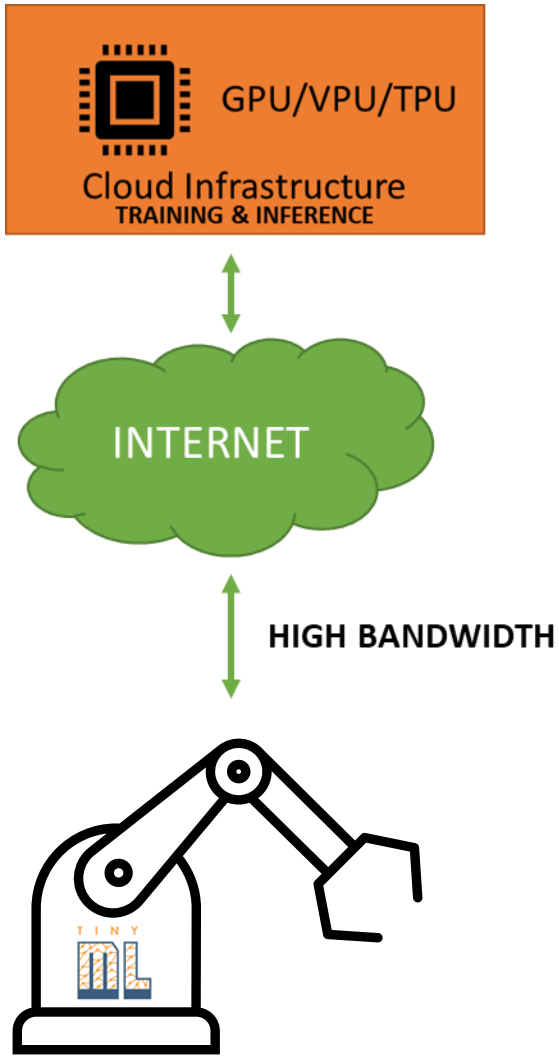
Introducing tinyML

Field of machine learning technologies utilizing optimized Machine Learning to perform inference on extremely lower power (mW range) embedded systems.

Power in mW range

Limited memory and processing power

Low Cost hardware



"Machine Intelligence next to the physical world"

Made possible by model optimization techniques and tools to optimize neural networks for inference on constrained devices

TINY ML TALKS webcast

Benefits of tinyML

LOW POWER	Targeting battery powered and portable applications
CONNECTIVITY	No Internet connectivity required for on device inference
COST	Low cost hardware no need for expensive GPU's/NPU's
PRIVACY	No connectivity means higher security and data privacy
LATENCY	Lowest latency due to efficient inference at data collection point

TINY ML TALKS webcast

Use cases for tinyML

Healthcare	Agriculture	Industrial	Retail	Transport
Disease Detection	Soil condition	Predictive Maintenance	Automated checkout	Self Driving vehicles
Acute distress	Adaptive lighting	Smart Sensors	Smart Mirrors	Fuel usage optimization
Sleep disorders	Crop diseases	Smart Manufacturing	Loss Prevention	Driver monitoring
Fitness & Exercise	Yield prediction	Process Control		Logistics

Computer Vision

Safety and Security

Autonomous Robots

Voice recognition applications

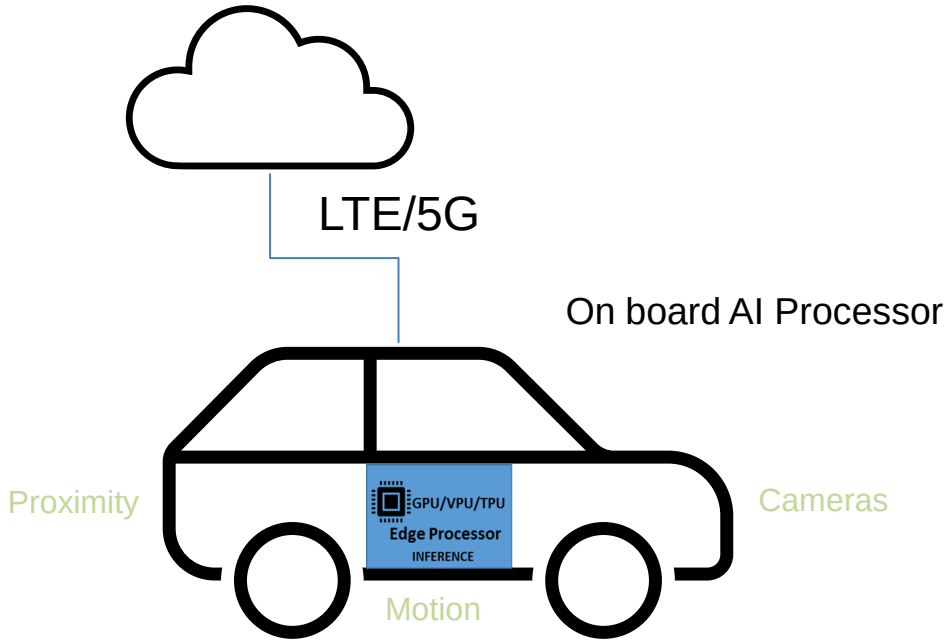
Building Management/Home Automation

…and many more

Diving deeper into an application scenario

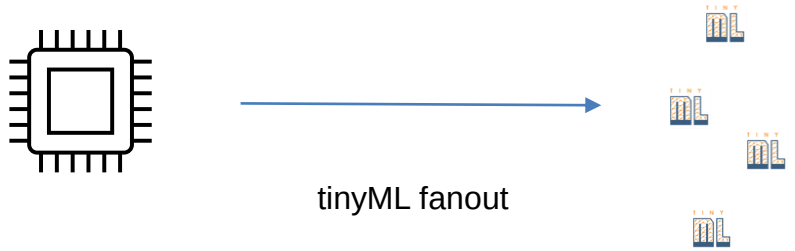
Sensors are Machine Senses

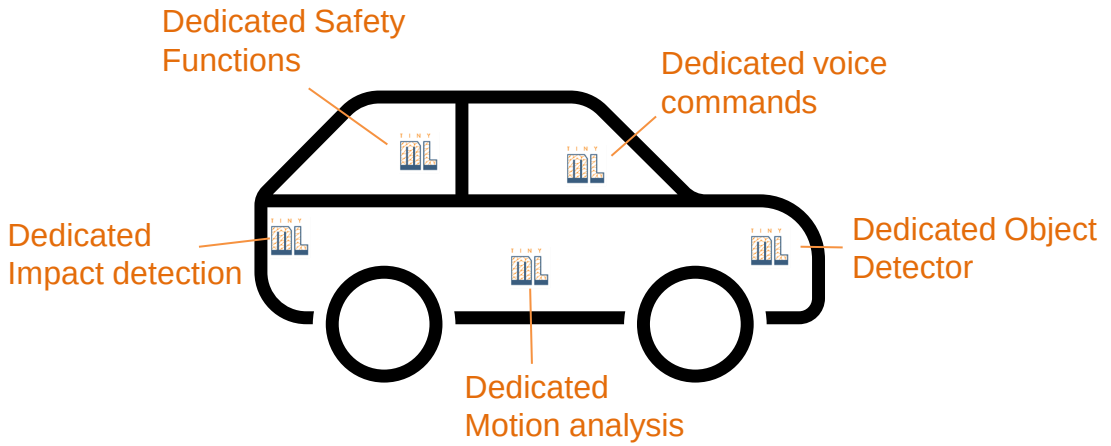
- Motion – IMU(Accelerometer/Gyro)
- Sound – Audio (Microphone)
- Sight – Image Sensor (Camera)
- Environment – Temperature/Humidity/Pressure
- Proximity - distance



Microprocessors/GPU –
General Purpose Multitasking

Microcontrollers –
Dedicated single function





The tinyML Movement

tinyML Foundation: non-profit organization creating and driving a Global Community around low power edge ML

- Bringing together a diverse community
- Non-discriminatory
- Multidisciplinary
- Open and transparent
- Highly Technical

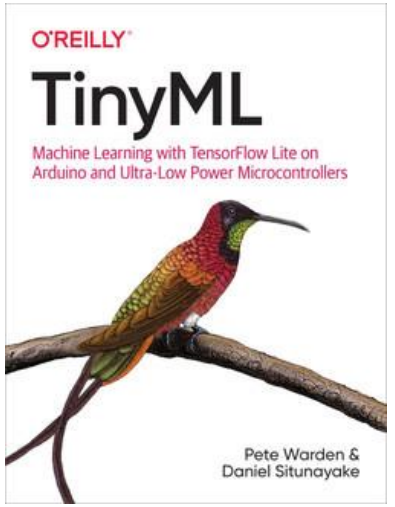
Summits

tinyML for Good

Vision Challenge

Meetup Groups

tinyML Talks



tinyML is becoming a key philosophy, technological approach and ML ecosystem as part of the continuation of the 4IR

Edge Impulse Demo

Getting started with ML

ML Devops ~ MLOps

EDGE IMPULSE

Machine Learning Workflow

- Understand the Problem
- Collect Data
- Prepare Data for ML
- Choose a model & train
- Validate & retrain
- Deploy

Jupyter Notebooks

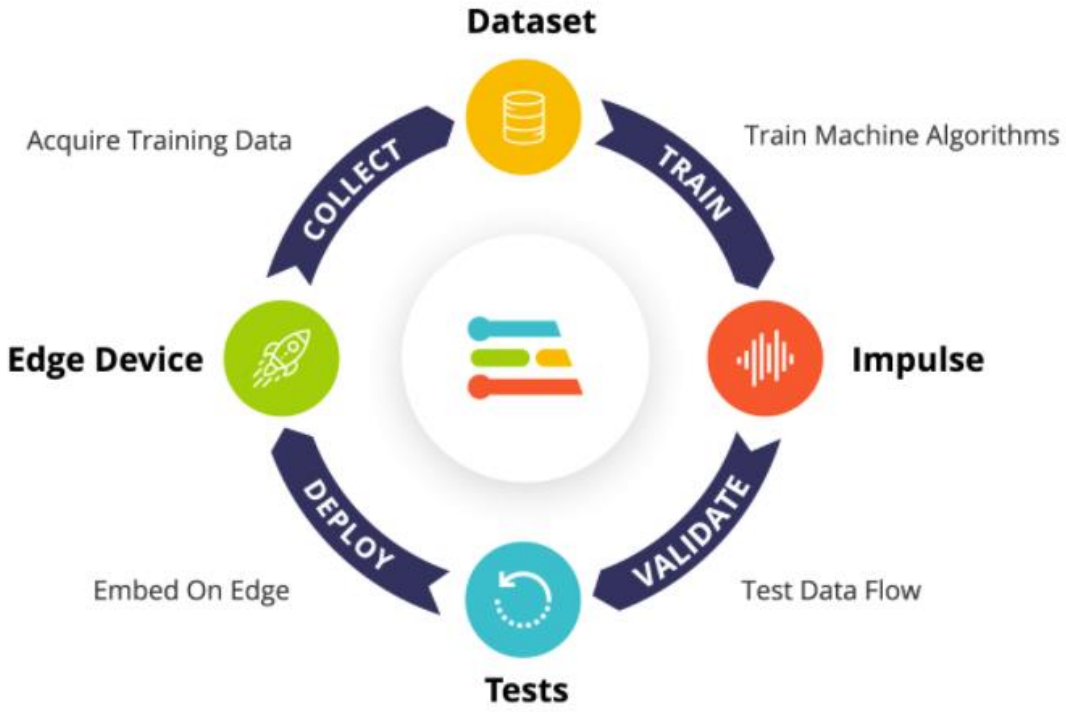
Custom code/interfacing

Managing Data

Feature extraction

Understanding Frameworks

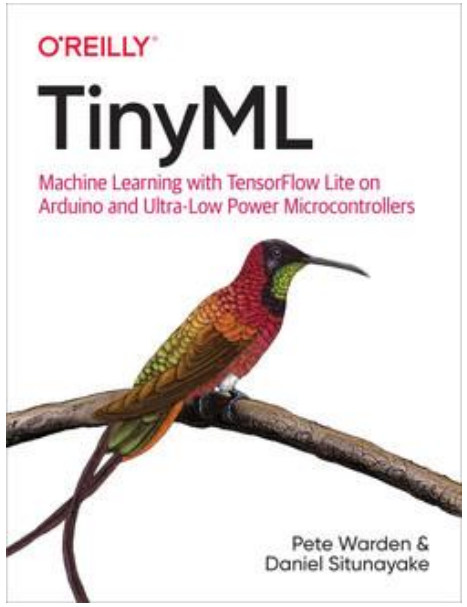
Splitting data



EON COMPILER

TINY ML TALKS webcast

Exploring Further



www.edgeimpulse.com/imagine

Copyright Notice

www.tinyml.org

Copyright Notice

www.tinyML.org