tinyML® Meetups

Enabling Ultra-low Power Machine Learning at the Edge

“Deploying tinyml to industrial equipments to increase processes efficiency: signal processing for predictive maintenance”

Daniele Gamba - AISent Srl

October 29, 2021

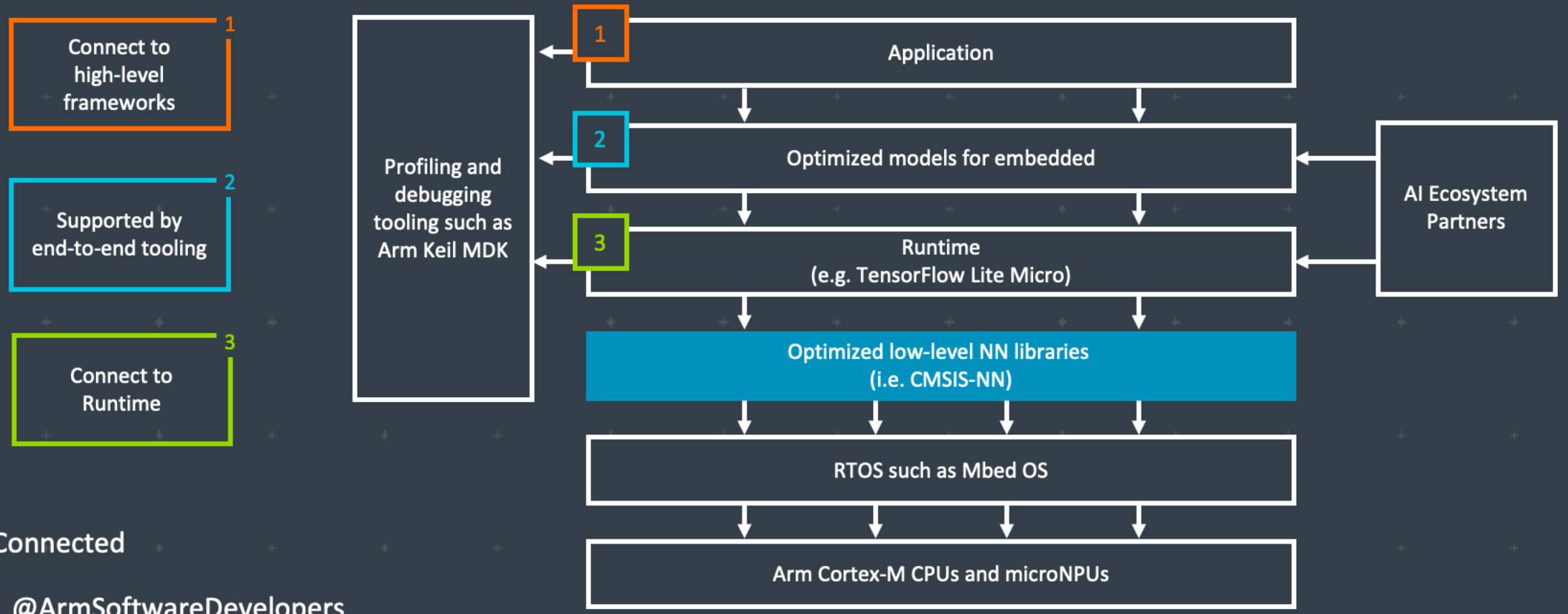
www.tinyML.org

TINY ML TALKS webcast

tinyML Talks Strategic Partners

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

© 2020 Arm Limited (or its affiliates)

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

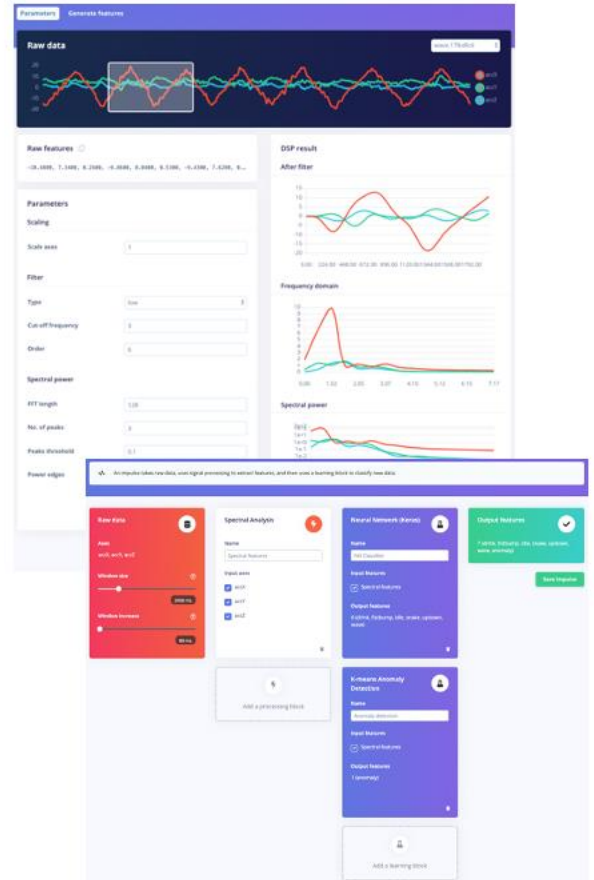
Open source SDK

Impulse

Embedded and edge compute deployment options

Test impulse with real-time device data flows

Test

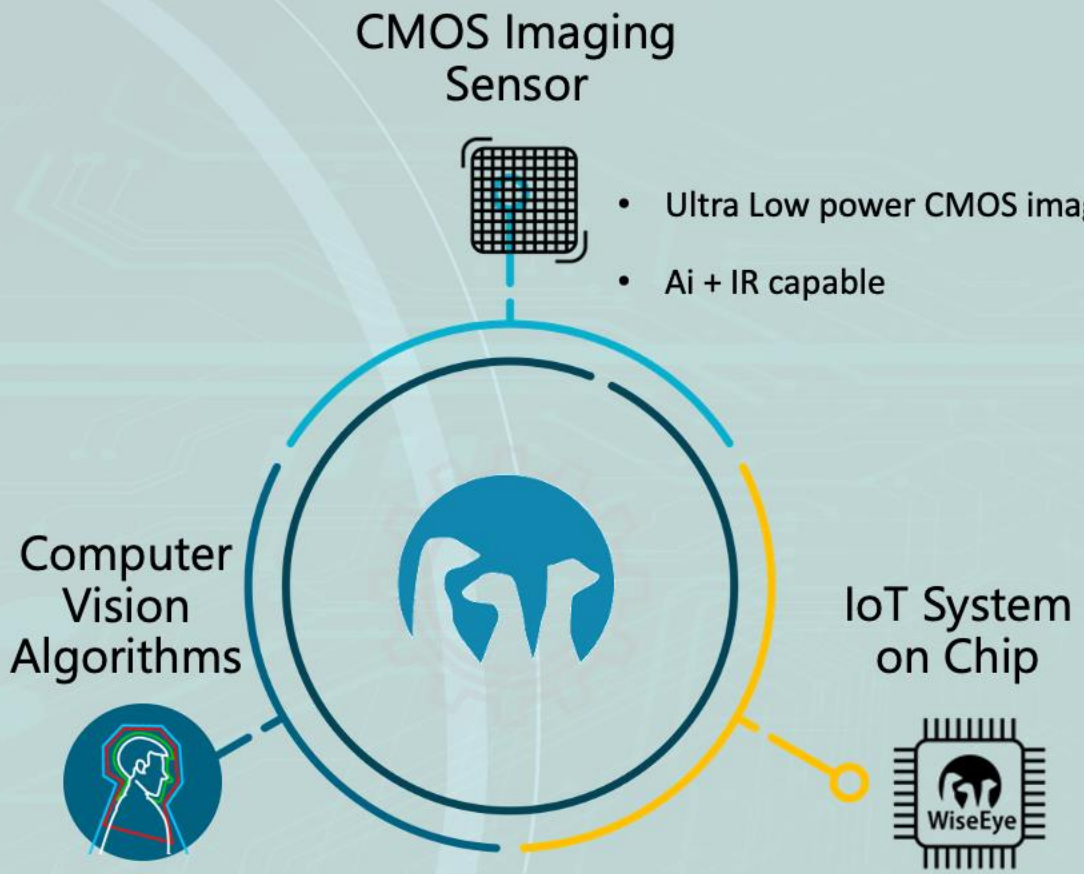


www.edgeimpulse.com

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

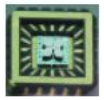
Sound

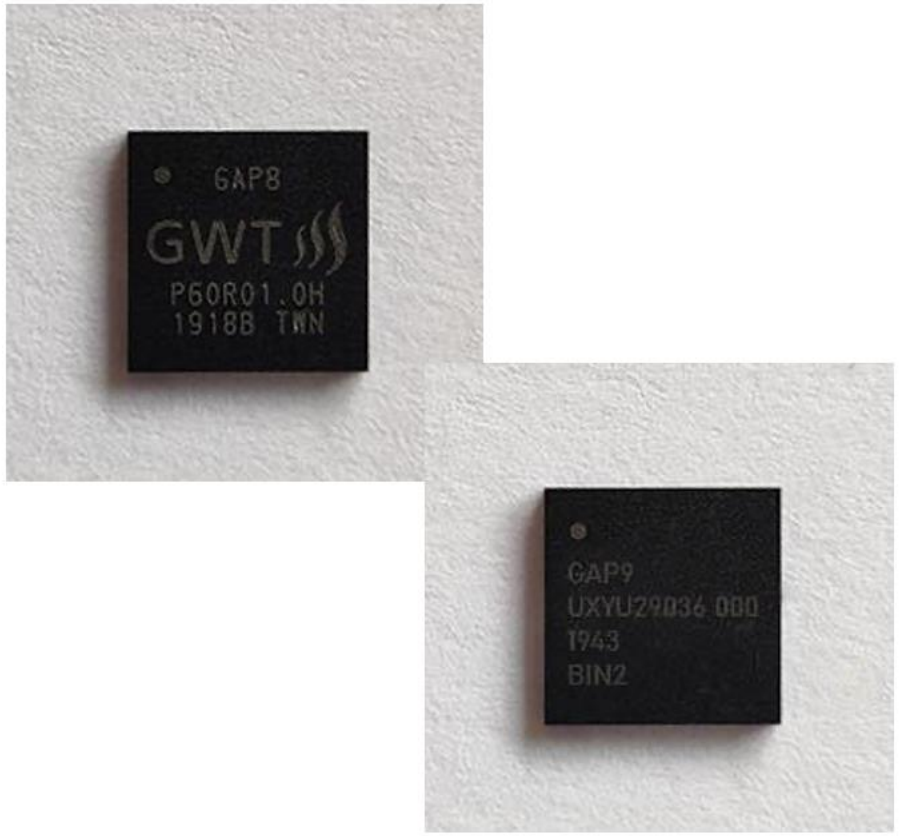
IR Image

Radar

Bio-sensor

Gyro/Accel

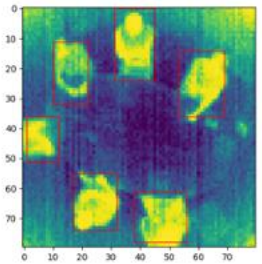




Wearables / Hearables

Battery-powered consumer electronics

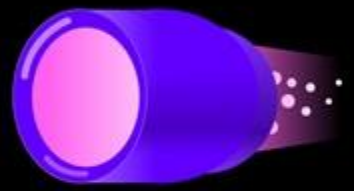
IoT Sensors



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**

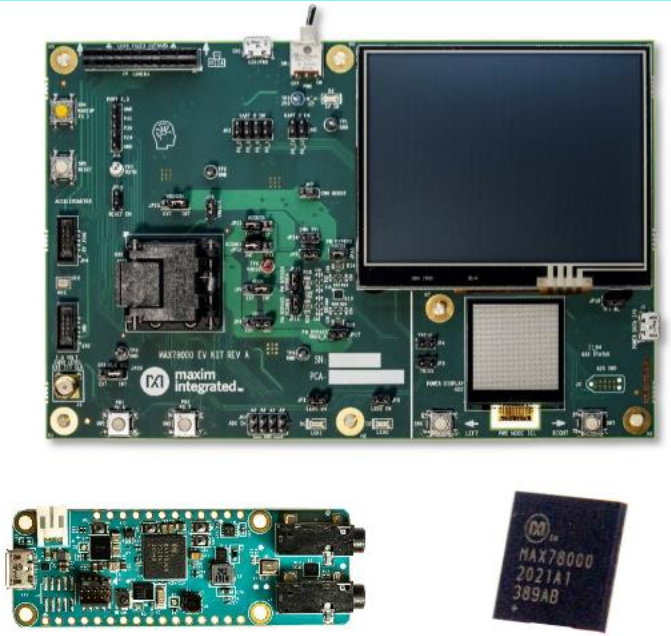
LatentAI

Adaptive AI for the Intelligent Edge

Latentai.com

Maxim Integrated: Enabling Edge Intelligence

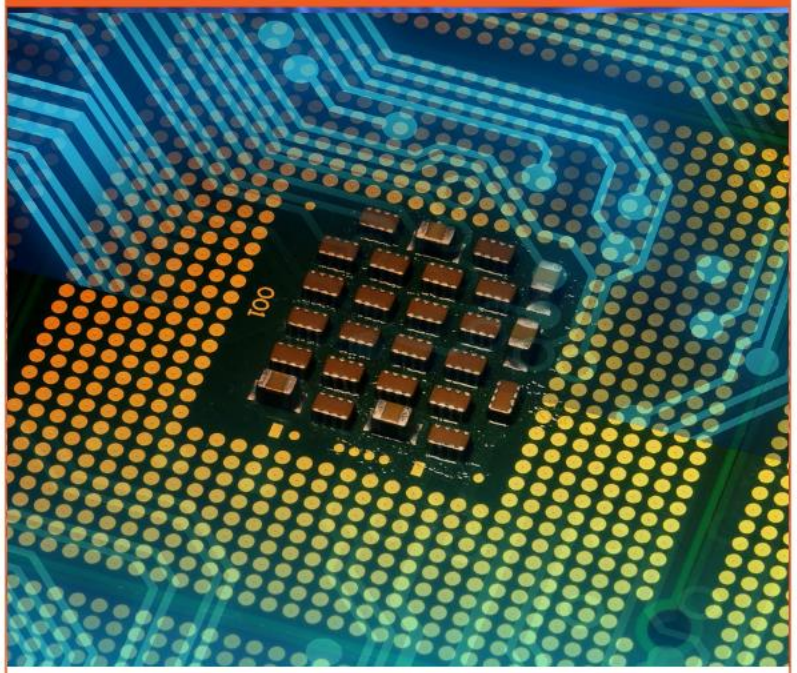
Advanced AI Acceleration IC



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

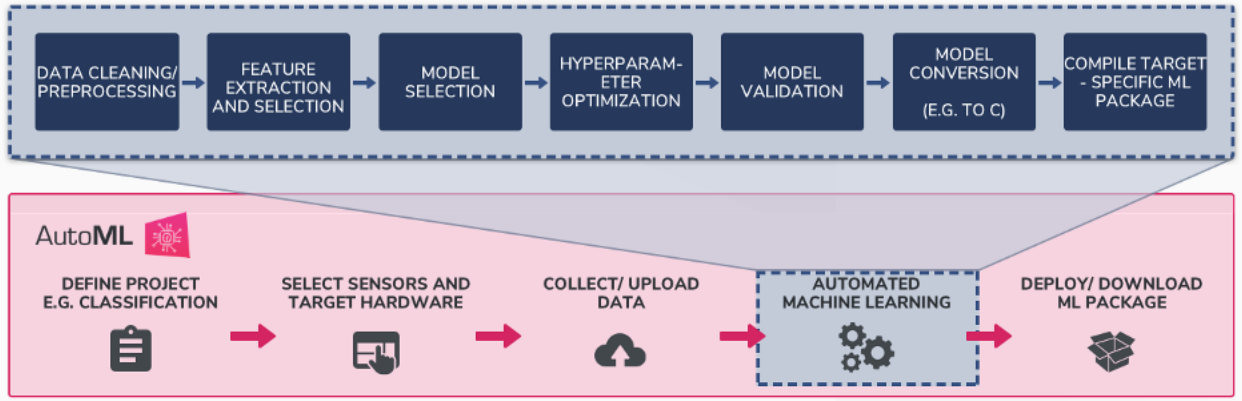
Qeexo AutoML

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

Qualcomm AI Research is an initiative of Qualcomm Technologies, Inc.

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

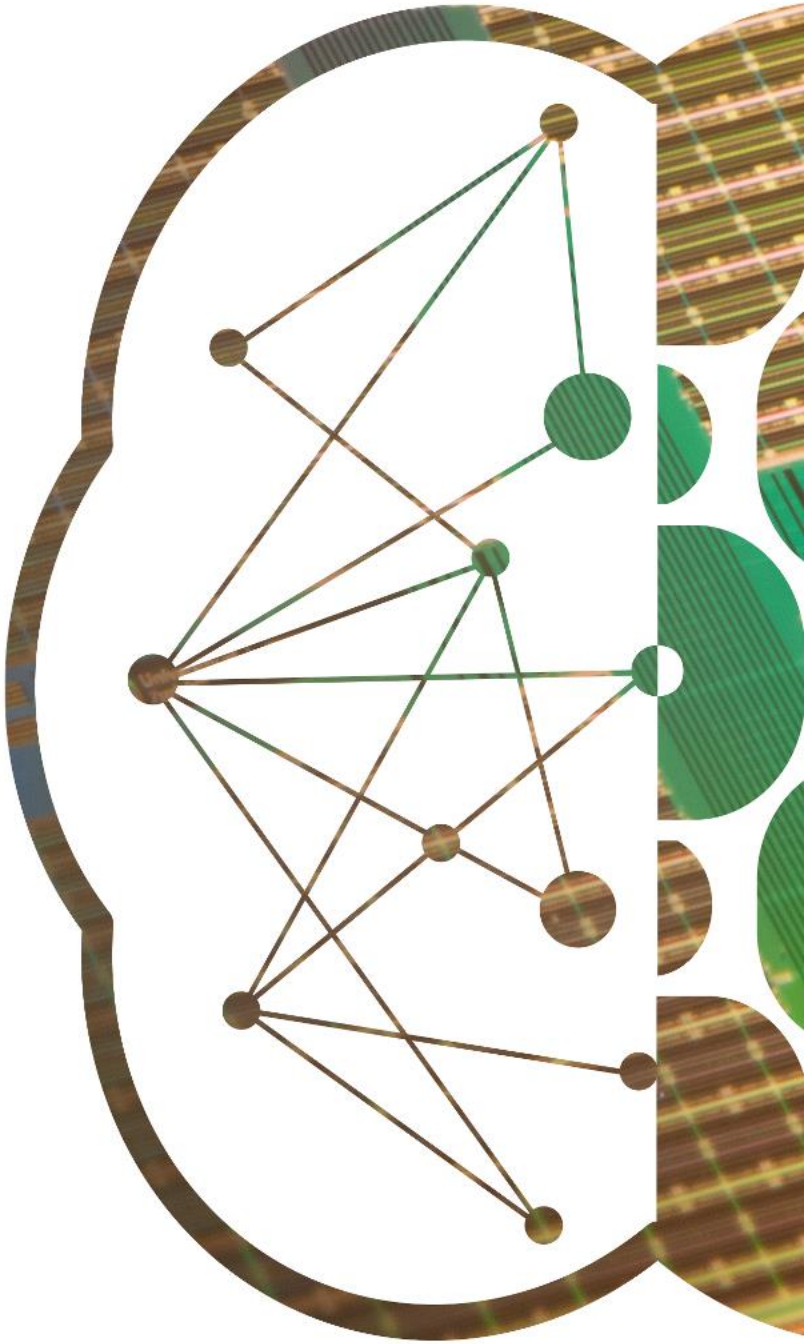
We enable the creation of production-grade smart sensor devices.

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

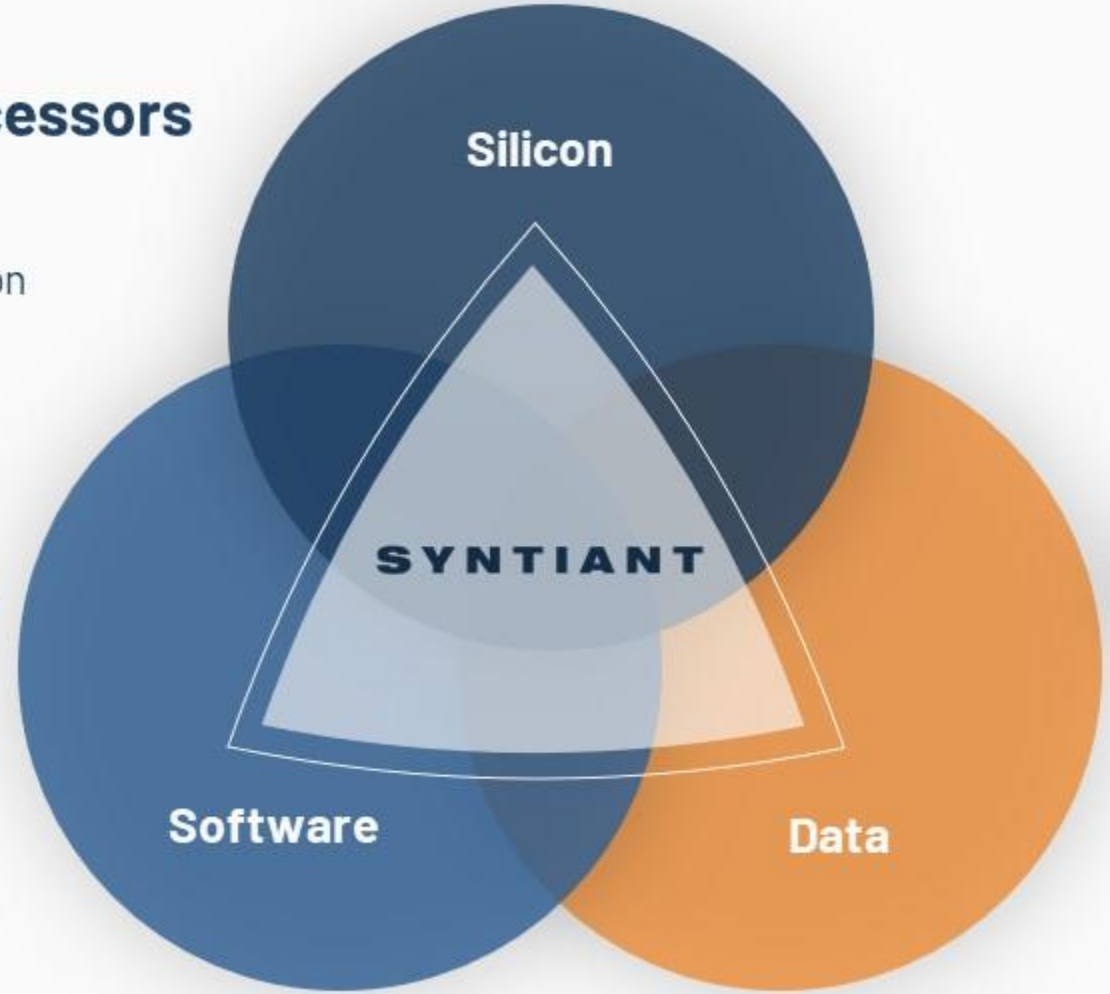


SYNTIANT

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

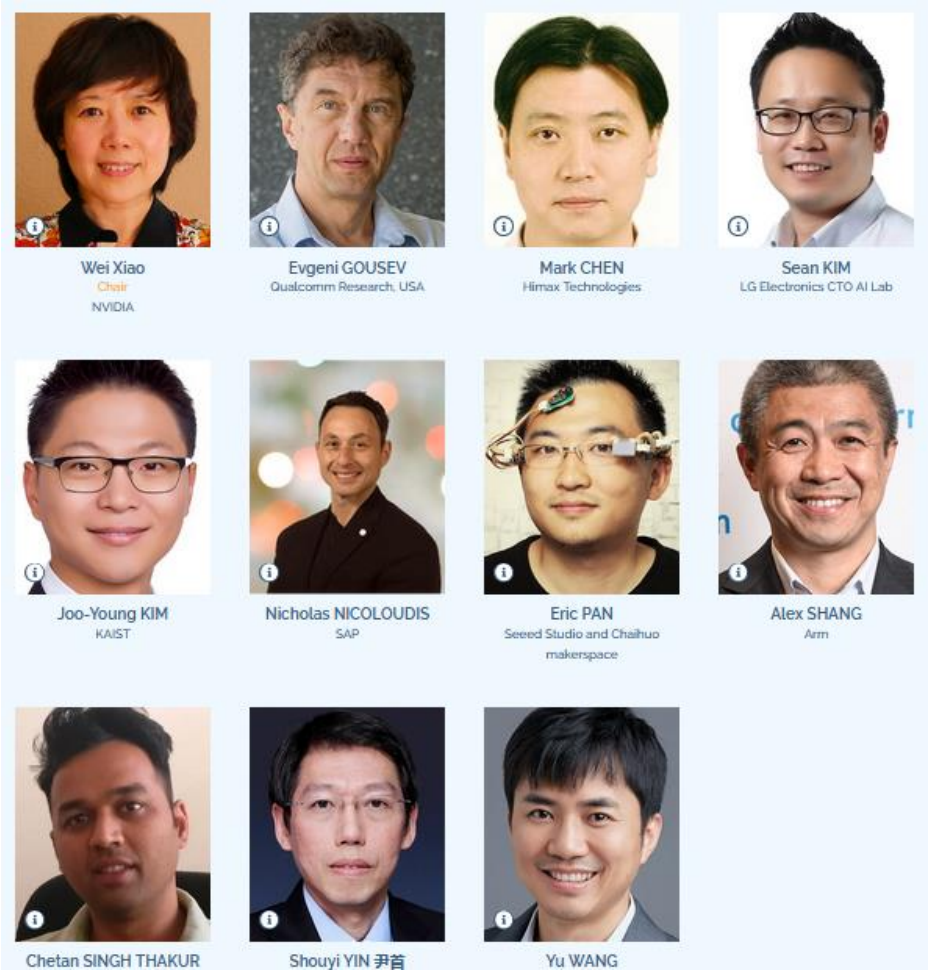
TINY ML ASIA

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

(9-11:30 am China Standard time)

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

Technical Program Committee



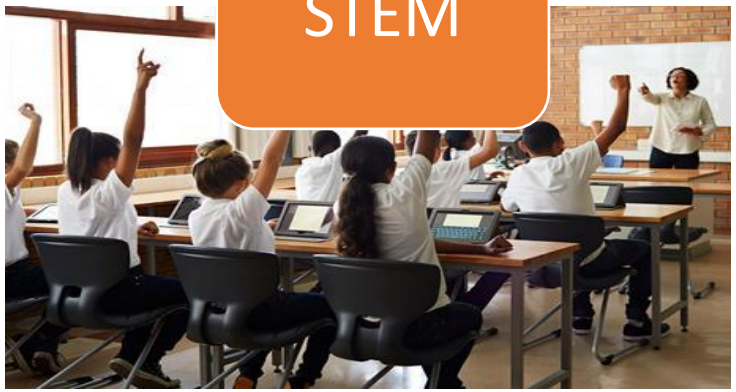
Register today!

Free event courtesy of our sponsors and strategic partners

More sponsorships are available: sponsorships@tinyML.org

tinyML for Good – Workshop, November 17th(7 am PDT)

STEM



TINY ML For Good

Healthcare

Earth
Climate
Conservation

Contact: 4good@tinyML.org

TINY ML VISION Challenge

THE 2021 WINNERS ARE

collaboration with hackster.io AN AVNET COMMUNITY

Team Sol

RANKED WINNERS: 1ST PLACE

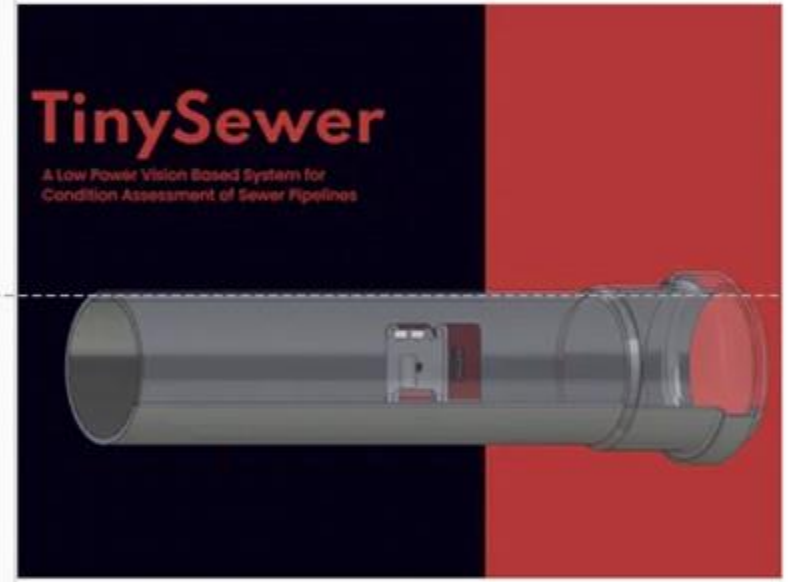
TinyML Aerial Forest Fire Detection

TheBlue Phoenix

RANKED WINNERS: 2ND PLACE

WorkSafe: Computer Vision based multiparameter monitor with

Huy Mai



RANKED WINNERS: 3RD PLACE

TinySewer - Low Power Sewer Faults Detection System

Honorable mention prize winners:

Flat Tire Detection Using Machine Vision by Bob Hammell

Smart Bird Feeder by **Ariela, Anna, Audrey, Nathan, Tianlang, Haoming, Eric, Edward and Tera** Guided by: Chen Feng

More details: tinyml.org/news/tinyml-vision-challenge-winners

arm EDGE IMPULSE Himax intel LATTICE SEMICONDUCTOR LUXonjs POI PixArt Imaging Inc. Qualcomm SynSense

Next tinyML Talks

Date	Presenter	Topic / Title
Tuesday, November 16	Rehan Hafiz, Information Technology University, Lahore	SuperSlash: Unifying Design Space Exploration and Model Compression methodology for design of deep learning accelerators for TinyML

Webcast start time is 8 am Pacific time

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

TINY ML TALKS webcast

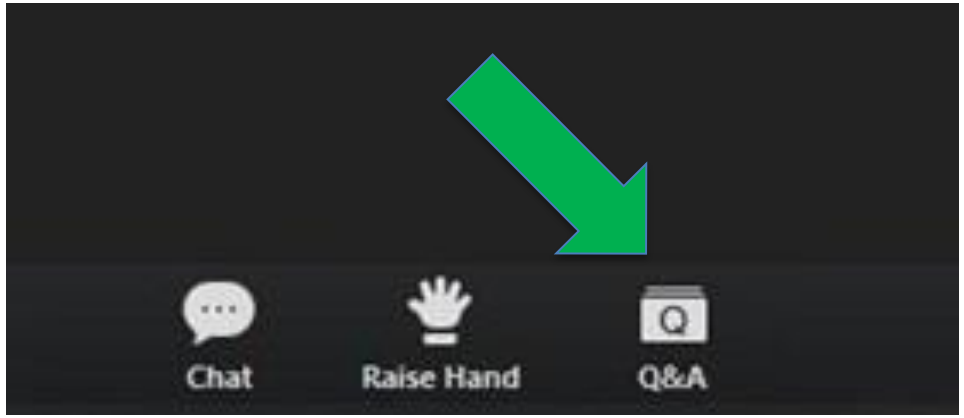
Reminders

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

Daniele Gamba

CEO and Co-Founder of AISent Srl. Active in the field of tinyml since he was working at the MS dissertation at the department of Mechatronics Engineering at the University of Bergamo. In the last years he led growth and innovation at AISent delivering customized AI applications in the field of computer vision and signal processing.

TinyML in industrial equipment

Daniele Gamba

AISENT

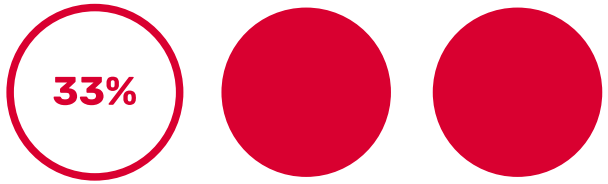
Artificial Intelligence Sentinels

We are a R&D startup focused on AI Algorithms and in building solutions around AI.

We started in 2018 as four former researchers from an IT laboratory of University of Bergamo using the algorithm we were studying in the industry.

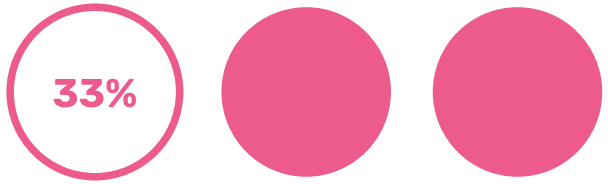
Team

PhD



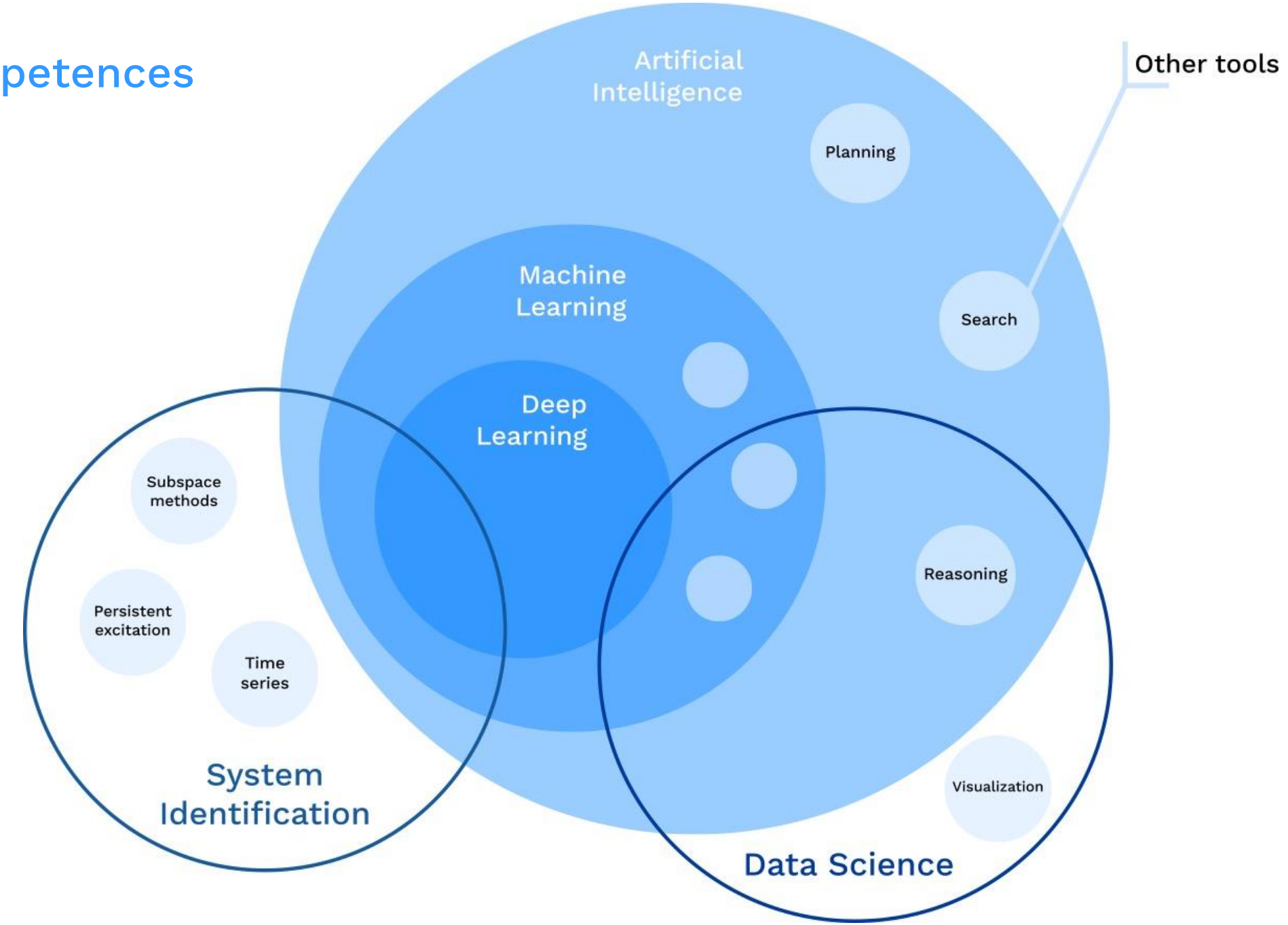
Average Age

Women



Staff - since 2018

Our core competences



Where we work

«It is difficult to think of a major industry that AI will not transform. AI is the new electricity» *Andrew NG*

Packaging

GDO & Retail

Transports

Machinery

Consumer Goods

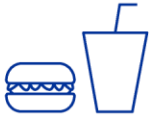
Pharma & Medicine

Chemical

Automotive

Cosmetics

Food & Beverage



Defence

Entertainment

Why ML

We are overwhelmed by data, and the number of data we could or do collect is growing every day.

We believe data is a tool to take **informed decisions**, ML allows us to extract information from these data.

It could be choosing to **open** an additional **security check** queue in an airport, or **improving the quality** of a process, **recommend** what **setting** should be best or performing a challenging operation automatically; whenever we can achieve even the smallest intelligent solution, we transform data into valuable information.

Why Tiny

Because often there is **no other option**.

Or when we do have other options, choosing the computation at the edge ease everything else.

If our sensors become intelligent, we don’t design logics, we build processes.

This is a crucial change of approach in designing industries, where obtaining a simple information like “correctly produced products” is challenging.

Since we need to distribute intelligence in order to achieve flexibility, we need ML algorithms, libraries and solutions that allow us to deploy even on the lowest compute power.

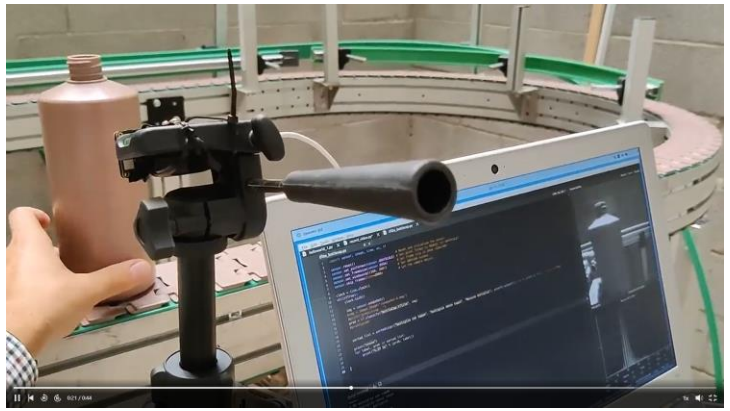
1 | A few examples

What if a **photocell** became a little camera, checking real time not only if something is going on the conveyor belt, but taking and elaborating **pictures**.

We could

- Have an exact counting of seen products
- Detect if the product is the correct one
- Perform some simple quality check
  (*Is the bottle correctly capped?*)

Then, all the other machinery could receive not only a simple counting, but events that need actions or decisions to be taken.



bit.ly/3pOBFAU

2 | A few examples

What if we want to build a sensor for blood in extracorporeal circulation that can be attached on the tube and able to read some blood parameter?

It could ease the work of many doctors, performing a real time monitoring on the condition of the patient, without having to rely on big and expensive machines. Although not being “industrial”, quality of the result is strictly conditioned by its **production variability**.

What if an electro-valve is able to detect itself when having a quickly deterioration of its performance.
It could notify the machine it needs to be changed and schedule maintenance in advance.

What we learned – Robustness is everything

When we deploy machine learning models on devices with very strict constraints we focus extremely on performance and robustness.

Quoting the EU, we need a Trustworthy AI, and among the many aspects we have to be sure our model is robust.

How we achieve robustness

- Under-parametrized models
- Synthetic data
- Data augmentation
- Model testing and validation

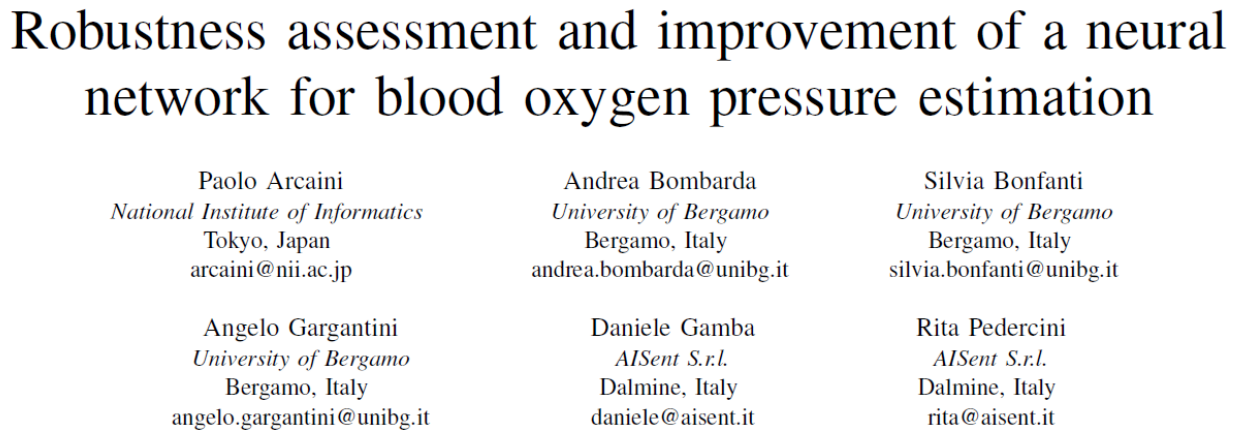
What we learned – Robustness is everything

We submitted a paper on how we built and tested a ML Model on STM32 micro at the International Conference on Software Testing, Verification and Validation 2022 (ICST 2022).

All the code will be open sourced in the ROBY package:

ROBustness analYzer for neural networks

https://github.com/fmselab/roby



Robustness assessment and improvement of a neural network for blood oxygen pressure estimation

Paolo Arcaini
National Institute of Informatics
Tokyo, Japan
arcaini@nii.ac.jp

Andrea Bombarda
University of Bergamo
Bergamo, Italy
andrea.bombarda@unibg.it

Silvia Bonfanti
University of Bergamo
Bergamo, Italy
silvia.bonfanti@unibg.it

Angelo Gargantini
University of Bergamo
Bergamo, Italy
angelo.gargantini@unibg.it

Daniele Gamba
AISent S.r.l.
Dalmine, Italy
daniele@aisent.it

Rita Pedercini
AISent S.r.l.
Dalmine, Italy
rita@aisent.it

Being robust to achieve flexibility

We live in a world that is evolving faster than ever.

Clients are better informed, with new priorities and changing needs in transparency, quality, sustainability and personalization.

Although this is exciting, our industries are not designed to be flexible. We need intelligence to achieve this resilence, even the smallest component should be intelligent enough to face the challenges of the future.

Thanks

aisent.io

info@aisent.io

AISENT

Copyright Notice

This presentation in this publication was presented as a tinyML® Talks webcast. The content reflects the opinion of the author(s) and their respective companies. The inclusion of presentations in this publication does not constitute an endorsement by tinyML Foundation or the sponsors.

There is no copyright protection claimed by this publication. However, each presentation is the work of the authors and their respective companies and may contain copyrighted material. As such, it is strongly encouraged that any use reflect proper acknowledgement to the appropriate source. Any questions regarding the use of any materials presented should be directed to the author(s) or their companies.

tinyML is a registered trademark of the tinyML Foundation.

www.tinyML.org