tinyML® Talks

Enabling Ultra-low Power Machine Learning at the Edge

"Embedded Edge Intelligence with Infineon New Products and Imagimob Studio"

Moenes Iskarous – CTO IoT AI/ML, Infineon Technologies
Sam Al-Attiyah – Head of Customer Success, Imagimob

May 16, 2024

www.tinyML.org

TINY ML TALKS webcast

Thank you, **tinyML Strategic Partners**,
for committing to take tinyML to the next Level, together

TINY ML TALKS webcast

Executive Strategic Partners

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

Accelerate Your Edge Compute

SYNTIANT

Making Edge AI A Reality

www.syntiant.com

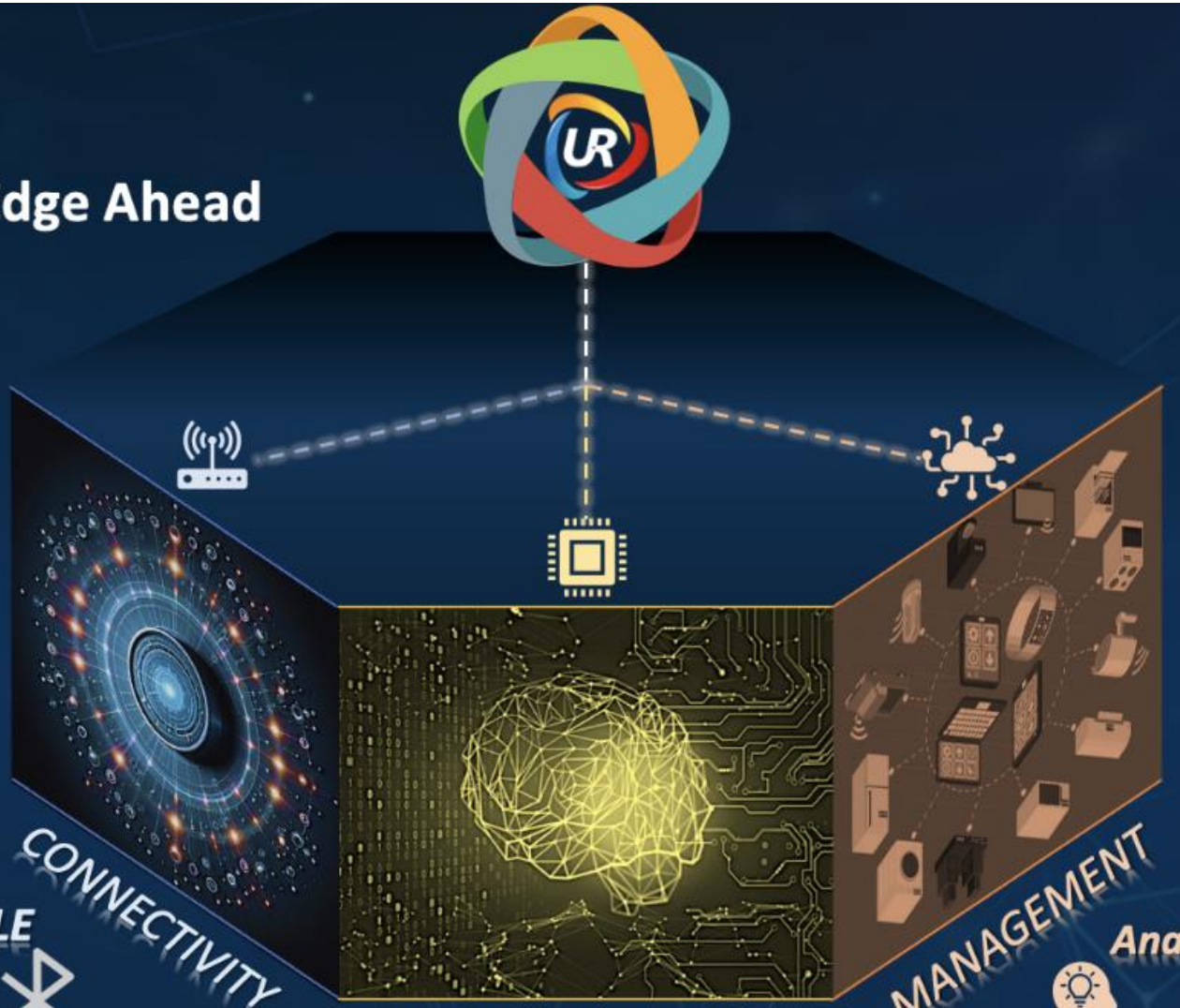
TINY ML TALKS webcast

Platinum Strategic Partners

embedUR

Software Services to Edge Ahead

TINY ML



BLE

WiFi

5G

Security

Edge AI

System

Analytics

Network Control

AI/ML

IMAGINE IT. DONE.

AITRIOS

DEPLOY VISION AI AT THE EDGE AT SCALE

SONY

TINY ML TALKS webcast

Gold Strategic Partners


Build the
Future of tinyML
on arm

TINY ML TALKS webcast

EDGE IMPULSE

The Leading Development Platform for Edge ML

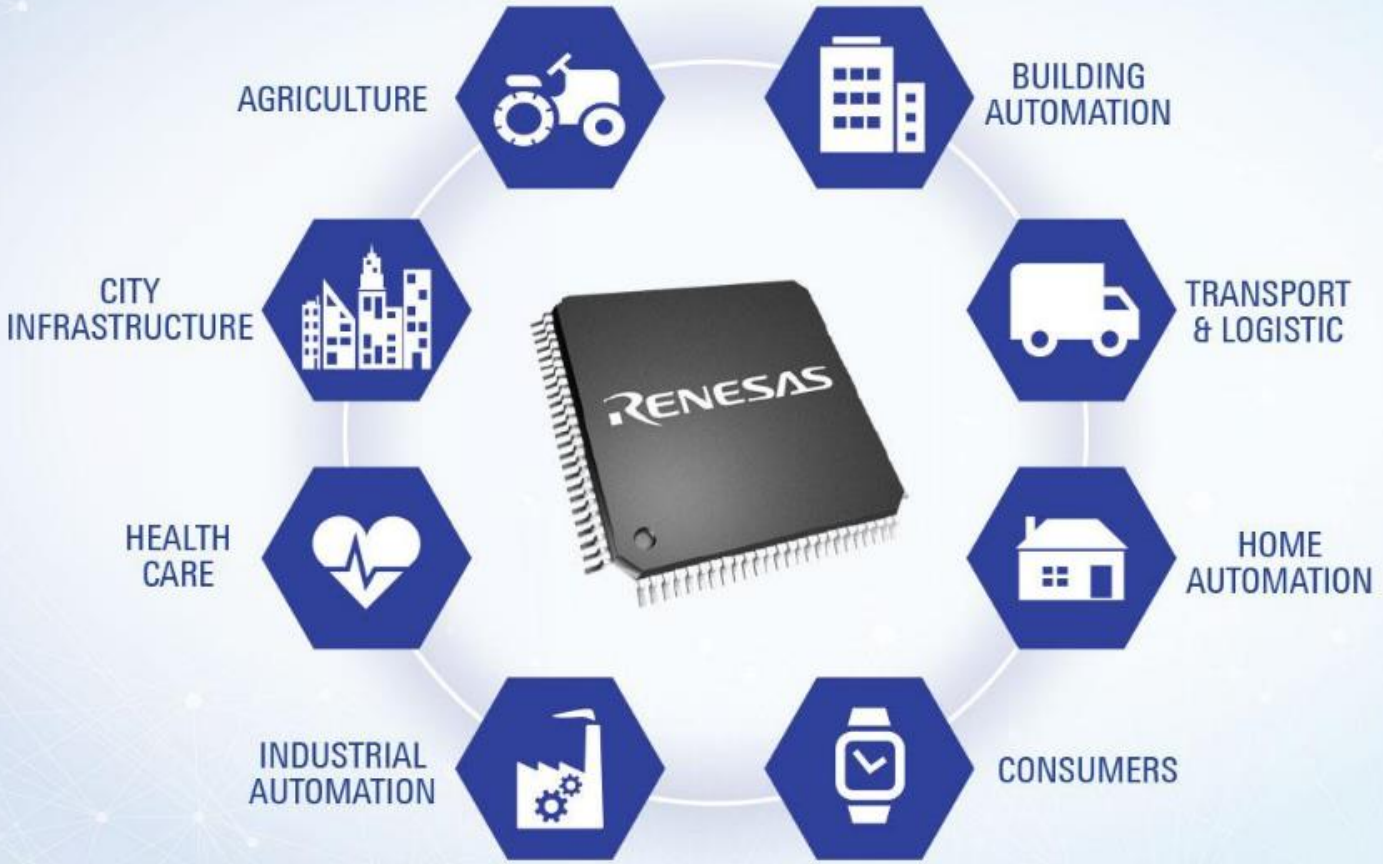
edgeimpulse.com

Driving decarbonization and digitalization. Together.

Infineon serving all target markets as
Leader in Power Systems and IoT

www.infineon.com

Renesas is enabling the next generation of AI-powered solutions that will revolutionize every industry sector.



renesas.com

RENESAS
BIG IDEAS FOR EVERY SPACE

STMicroelectronics provides extensive solutions to make tiny Machine Learning easy

life.augmented

www.st.com/ai

TINY ML TALKS webcast

Silver Strategic Partners

Join Growing tinyML Communities:

20k members in
50 Groups in 42 Countries

tinyML - Enabling ultra-low Power ML at the Edge

https://www.meetup.com/tinyML-Enabling-ultra-low-Power-ML-at-the-Edge/

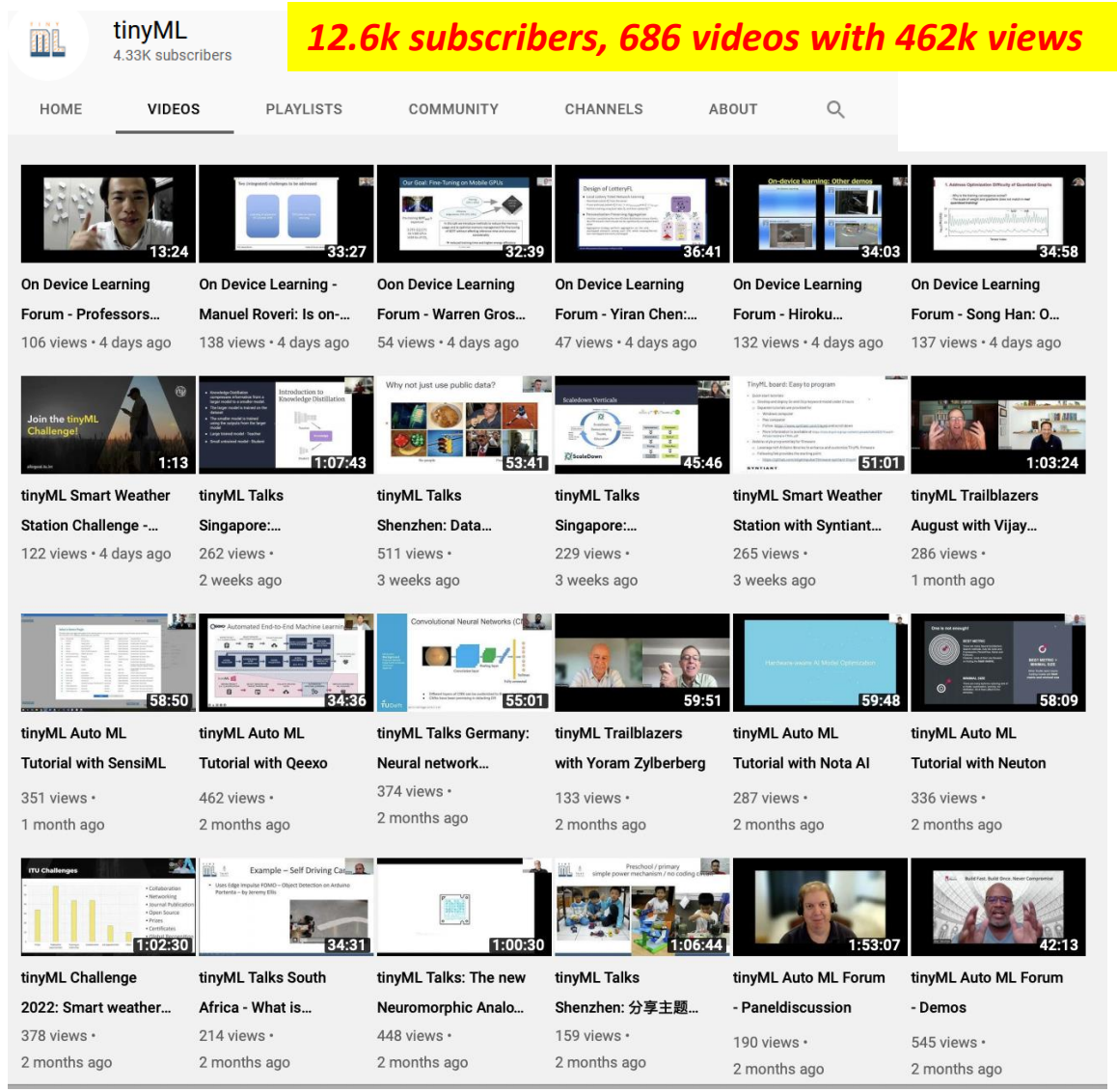
4k members
&
16k followers

The tinyML Community

https://www.linkedin.com/groups/13694488/

TINY ML TALKS webcast

YouTube



Subscribe to tinyML YouTube Channel for updates and notifications

(including this video)

www.youtube.com/tinyML

tinyML EMEA 2024

TALKS webcast

Amplifying Impact – Unleashing the Potential of TinyML

REGISTER NOW

tinyML EMEA
June 24 -26, 2024 in Milan, Italy

Reminders

Slides & Videos will be posted asap

tinyml.org/forums

youtube.com/tinyml

Please use the Q&A window for your questions

TINY ML TALKS webcast

Moenes Iskarous

Moenes is passionate about embedded development and believes that edge intelligence is a major growth area for AI. He has more than 32 years of industrial experience in the computer and semiconductor industry with focusing on ML technology development. Moenes received his PhD from Vanderbilt University in Neural Networks Architectures and application in Robotics then he worked at C-Cube Microsystems, Intel, SiMa.ai and SK Hynix before joining Infineon as the CTO for IoT AI/M leading the Machine Learning Center of Excellence.

TINY ML TALKS webcast

Sam Al-Attiyah

Sam has been working at Imagimob for more than 7 years. He has a Master in Electrical Engineering having studied both in Australia and in Sweden. Over this time he has had experience working with the end-to-end machine learning applications on the edge and led multiple projects to production. Sam is now heading the Customer Success team at Imagimob which focuses on ensuring customers get to production with their machine learning applications. Sam is also the product manager for the Imagimob Ready Models which are fully trained models ready deployment and commercialization.

TINY ML TALKS webcast

Ashutosh Pandey

Ashutosh Pandey is currently a Lead Sr. Principal Systems Engineer at Infineon Technologies where he is responsible for Machine learning solutions, architecture, and tooling. He holds a PhD from the University of Utah and has over 50 papers and patents on speech/audio/machine learning systems and algorithms.

Agenda

1. IoT @ Infineon
2. Embedded Edge Intelligence with Infineon New Products
3. Imagimob Studio
4. Q&A

IoT @ Infineon

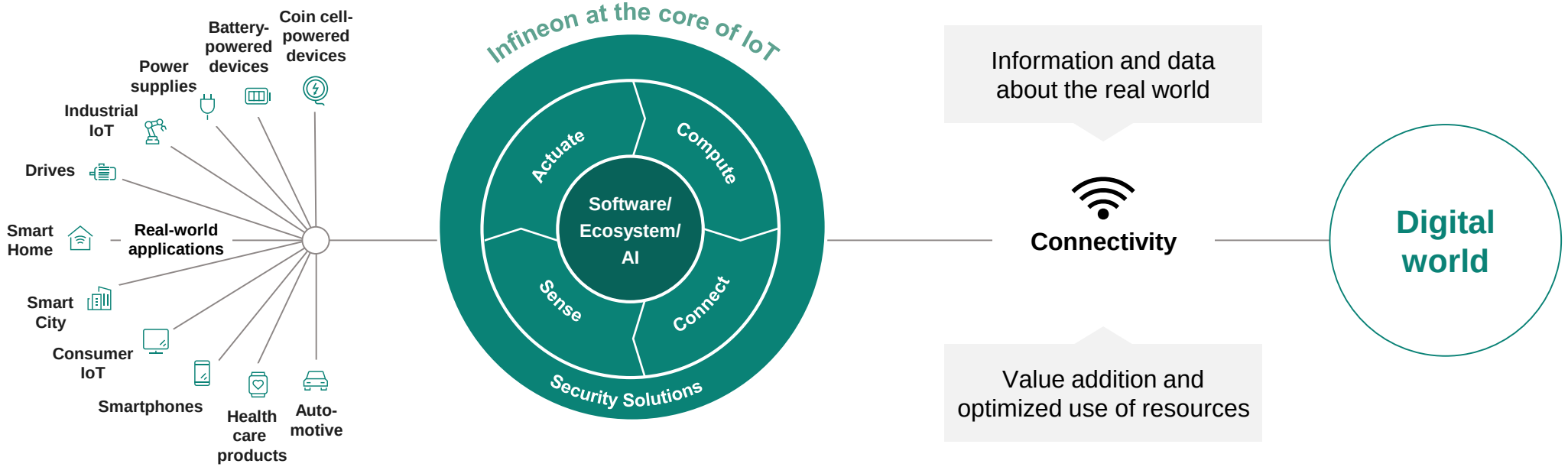
Infineon leader in IoT – driving digitalization by serving strongly growing multi-application markets

Consumer IoT

Industrial IoT

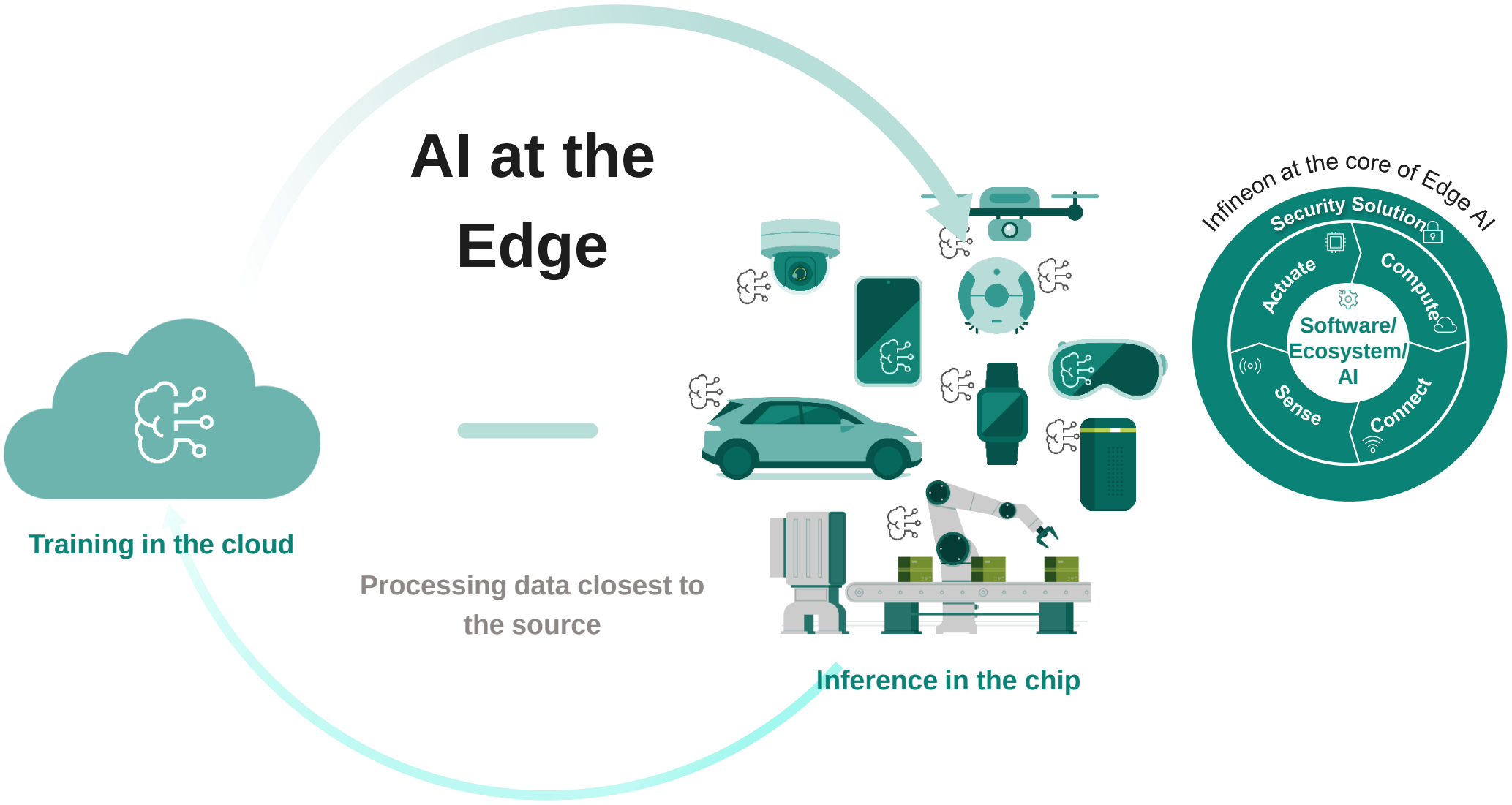
Automotive IoT

Products: MCU – Connectivity (Wi-Fi, BLE, NFC) – Sensors – Security – Power supply & switches



Embedded Edge Intelligence with PSOC™ Edge

Real-time requirements and the need for power-efficiency, security and privacy drives AI-processing at the edge



Key benefits of Edge AI

Low latency and real-time response

Higher power efficiency

Improved security and data privacy

Reduced cost

Infineon offers end-to-end technology solutions for your AI market entry.

Digital Services and AI solution platforms

INDUSTRIAL ANALYTICS

Hybrid AI-based services

In-field Power Analytics

e.g. RUL*

XENSIV™ sensor solutions

e.g. Sleep quality service

And others…

AI-models for a wide variety of applications

imagimob
An Infineon Technologies Company

Ready Models

- Baby cry detection
- Siren detection
- Surface detection
- Wearing detection (for headphones, helmets, etc.)
- Coughing detection
- Alarm detection
- Yelling & commotion detection
- ...
- Snoring detection
- Gesture detection
- Arc Fault Circuit Interrupters

End-to-end software solutions for easy training & deployment

TensorFlow

Keras

ONNX

imagimob
An Infineon Technologies Company

Imagimob Studio

ModusToolbox™

AI partner ecosystem:

SYNOPSYS®

DesignWare ARC
MetaWare Toolkit

EKHONO AI

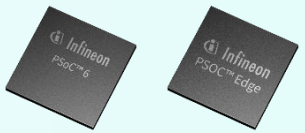
Cyberon
Leading Speech Solution Provider

SensiML™

Micro.ai

The right hardware for your Edge AI model

MCUs:



PSOC™

AURIX™

TRAVEO™

XMC™

XENSIV™ smart sensors portfolio for automotive, industrial and consumer

*Remaining useful life

Imagimob with ModusToolbox™ from data to edge model

Build a model

Bring your own model

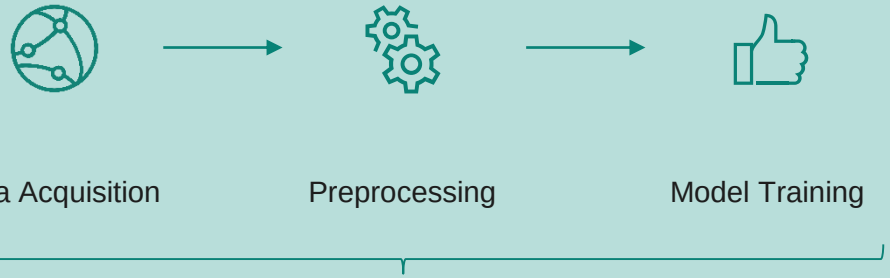
OR

Buy a Ready Model (no optimization needed)

Optimize & validate your model

Develop your embedded product with IFX MCUs

Imagimob Studio



Data Import
Data Capture
Data Generation
Data Augmentation
Data Labeling
Preprocessor / Feature extractor selection
Model Selection
Model Training
Model Validation

Imagimob Studio + ModusToolbox™

Hardware Optimization

ModusToolbox™

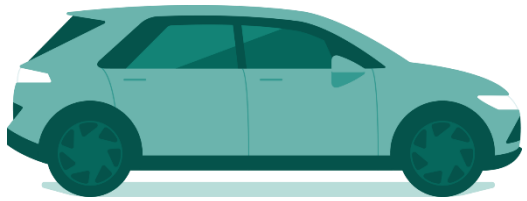
Code Generation
Target Device Test and Validation
Field Deployment

Our advanced technological solutions address a wide range of Edge AI applications

AI in IOT & Consumer

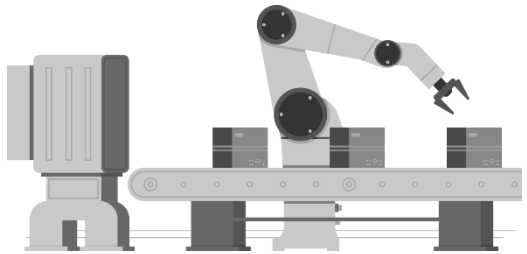
Democratizing AI by bringing the computational power of AI algorithms **closer to the source data** with **smarter** and **greener** devices for **intuitive real-time** interaction.

AI in Automotive



Ushering in a new era of connected and autonomous vehicles with **reliable**, **safe**, and **secure** systems for **real-time safety critical applications**.

Industrial AI



Creating self-learning systems for greater **productivity**, **quality**, and **efficiency** and supporting the adoption of sensor-based **predictive maintenance** models.

Infineon provides a comprehensive end-to-end embedded AI solution

In-house AI Software

imagimob
An Infineon Technologies Company

Development & AI Ecosystem

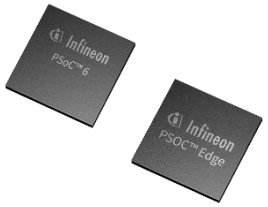
ModusToolbox™ Software

SensiML

Cyberon
Leading Speech Solution Provider

Micr.ai

PSoC™ 6 or PSOC™ Edge Microcontroller

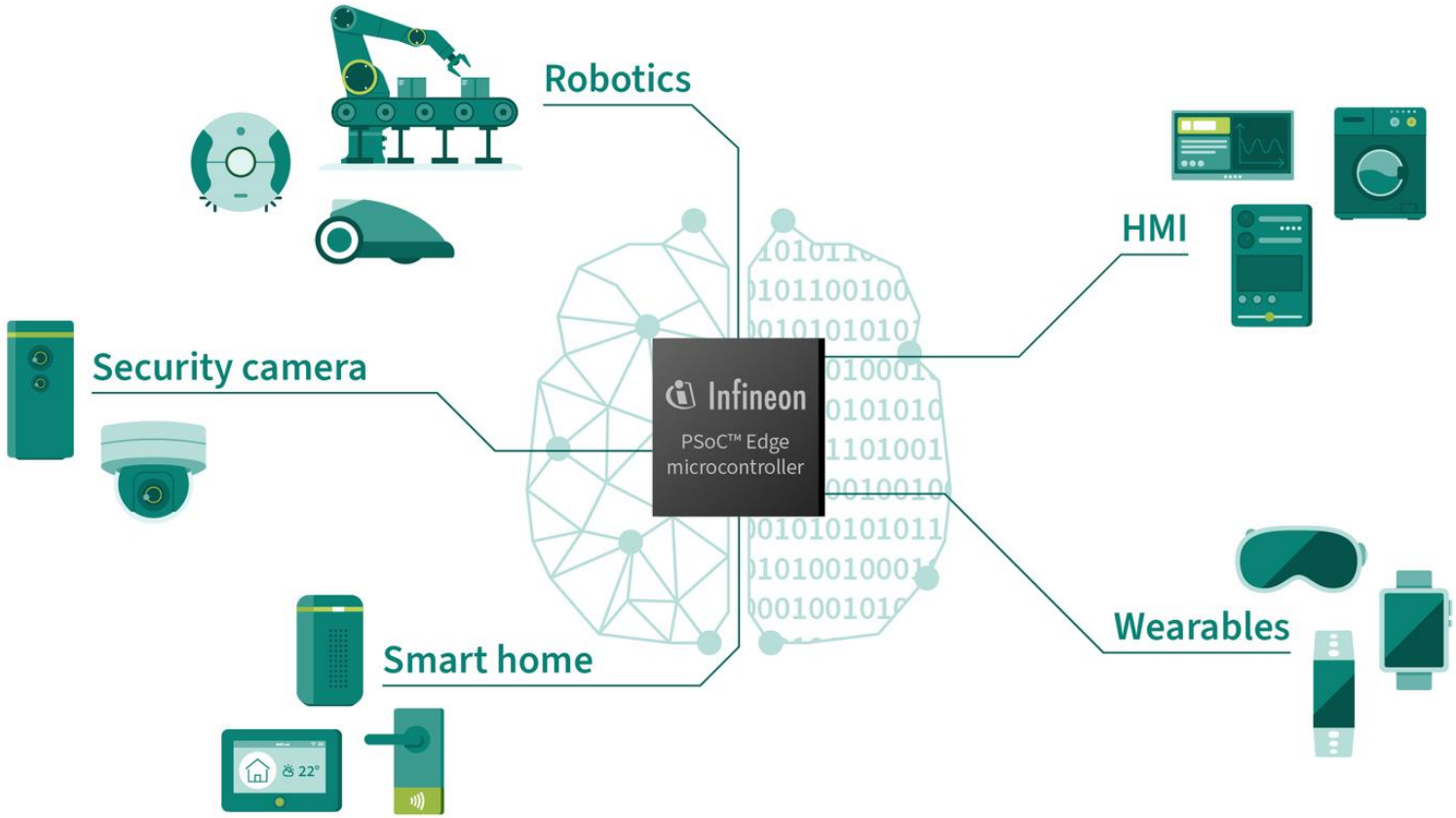


Infineon sensors

Connectivity
& security solutions

Next-generation PSOC™ Edge portfolio: Infineon PSOC™ Edge E81, E83 and E84 MCUs

PSOC™ Edge – Enables a new generation of responsive machine learning devices



Human Machine Interface (HMI)
- Appliances
- Industrial Device Usability
- Factory Automation

Smart Home
- Thermostat
- Speaker
- Door lock

Robotics
- Vacuum Cleaner
- Vacuum Robots
- Service Robot
- Lawn & Garden Robotics
- Industrial Robotics

Wearables
- Fitness Watch
- AR/MR/VR Glasses & Accessories
- Audio Accessories

Security Camera
- IP Camera
- Doorbell
- Security Camera & Accessories

PSOC™ Edge – Next Gen MCU based Edge Device Platform

Fully Integrated MCU SoC Family with comprehensive Tools, Software & Enablement

… enables developers of tomorrow's applications to move faster, with a richer set of capabilities, at lower system power and cost

System/Software Compatibility

100% Pin Compatibility

High Performance & Low Power Compute

- Cortex®-M55 MCU core with Helium™ DSP
- Cortex®-M33 MCU core with NNLite for always-on operation
- Embedded ultra low-power RRAM technology

More Memory Resources & SoC Integration

- More and more memory available for next-gen apps
- Richer Peripheral set to reduce system cost

More Robust Security

- Infineon Edge Protect Technology (EPC2 or EPC4)

Enablement

- ModusToolbox™ : Software, tools, middleware & more
- Imagimob AI/ML Solution
- Hardware evaluation kit

E81

Machine Learning

- Ethos-U55 for high-performance AI/ML hardware NN compute

E83

Graphics

- Graphics fully synchronized with voice and smart apps

E84

Driving tomorrow's products with:

More Features
Lower Power
More Security
Motor Control
Machine Learning

- Presence/Gesture/Motion
- Predictive Maintenance
- Anomaly Detection
- Autonomous Operation

Advanced ML

- Voice & Natural Language
- Vision
- Access & Safety

Graphics

PSOC™ Edge Portfolio – Next Gen MCU based Edge Device Platform

Scalable, compatible, future proof platform for your next-gen MCU based smart devices

Broad MCU Based SoC Family

PSOC™ Edge E81

Higher Performance
More Memory
More Peripheral Support
Higher Integration
Lower Power
Base Edge ML
Robust Security

PSOC™ Edge E83

Adds:

Advanced Edge ML
Audio/Voice
Vision
Presence/Gesture/Motion

PSOC™ Edge E84

Adds:

Graphics
Even More Memory

100% Pin Compatibility

System/Software Compatibility

PSOC™ Edge E8 MCU Family

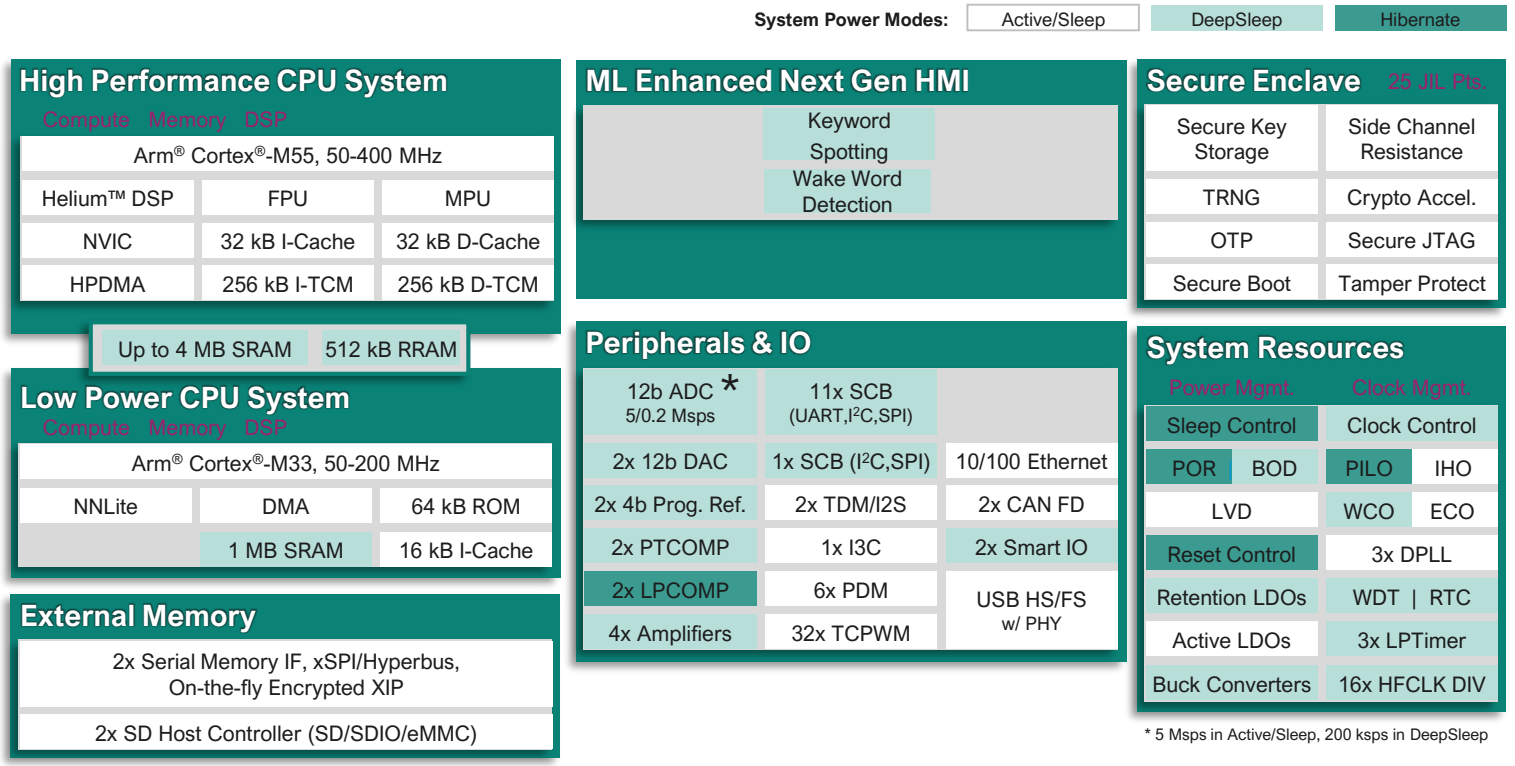
	PSOC™ Edge E81	PSOC™ Edge E83	PSOC™ Edge E84
Processor	Cortex-M55 + DSP (High Performance Domain) Cortex-M33 and DSP (Low-Power Domain)	Cortex-M55 + DSP (High Performance Domain) Cortex-M33 and DSP (Low-Power Domain)	Cortex-M55 + DSP (High Performance Domain) Cortex-M33 and DSP (Low-Power Domain)
Machine Learning	M55 w/ Helium DSP/NN accelerator NNLite	M55 w/ Helium DSP/NN Accelerator, NNLite Ethos-U55 - 128 MACs	
SRAM	Up to 4 MB (SoC SRAM) Up to 1 MB (Low-Power Domain)		Up to 5 MB (SoC SRAM) Up to 1 MB (Low-Power Domain)
RRAM	512 kB		
External Memory	2x SMIF, 2x SD Host Controller		
Peripherals & IO	USB, 10/100 Ethernet, CAN, SPI, UART, I2C, I3C, I2S		
Audio/Voice	ULP Always ON prog. analog for voice, audio, sensing 4x Analog Mic, 6x Digital Mic NNLite Wake Word & Acoustic Activity Detection	ULP Always ON prog. analog for voice, audio, sensing 4x Analog Mic, 6x Digital Mic U55 ML-based Wake Word & Acoustic Activity Detection Full Voice Inferencing	
Graphics	No	No	LP 2.5D GPU Up to 1024 x 768, MIPI-DSI/DBI formats
Vision	No	Position Detection/Face Recognition/Object Detection (VGA)	
Security	Secured Enclave, Edge Protect Category 2 and 4		

PSOC™ Edge E81: Next Gen, Low Power MCU

Features

- **High performance real-time compute domain**
  - Cortex®-M55 w FPU + Helium DSP
  - Up to 4 MB System SRAM, 256 KB I&D TCMs
  - 512 kB RRAM
- **Low power compute domain**
  - Cortex®-M33 and DSP + IFX NNLite for ML
  - 1 MB SRAM
- **HMI**
  - Traditional MCU HMI
- **ML**
  - Base ML leveraging M55, Helium DSP and NNLite
- **Peripherals & IO**
  - USB, 10/100 Ethernet, CAN, SPI, UART, I2C, I3C, I2S
  - Ultra-low-power always-on analog
- **Security**
  - Secured Enclave @ 25 JIL pts, fit for ARM PSA L4

Target Applications: Appliances, Thermostats, Home Security, Industrial HMI, etc.



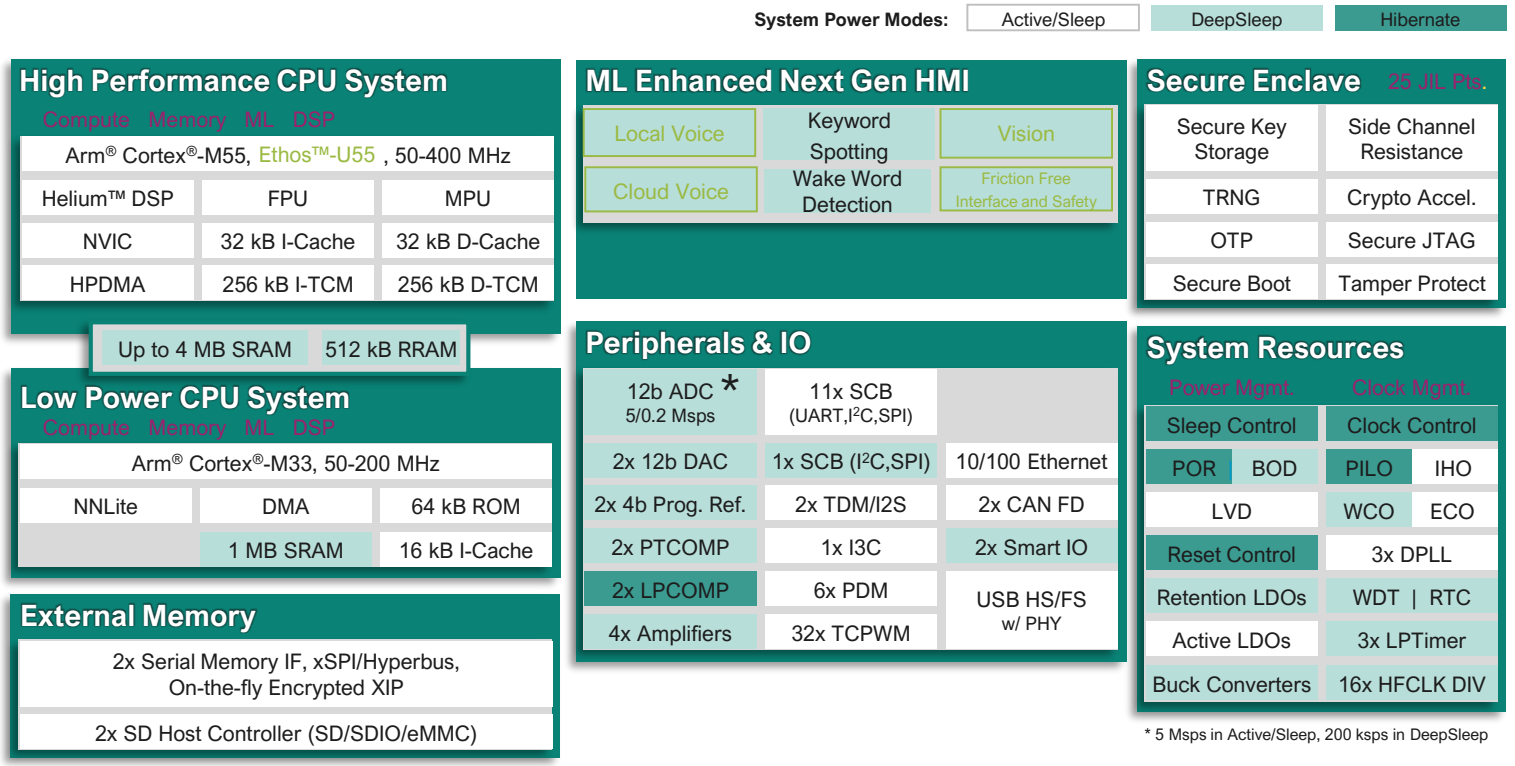
* 5 Msps in Active/Sleep, 200 ksps in DeepSleep

PSOC™ Edge E83: Next Gen, Low Power ML MCU with Voice/Vision

Features

- **High performance real-time compute domain**
  - Cortex®-M55 w FPU + Helium DSP + Ethos-U55 for ML
  - Up to 4 MB System SRAM, 256 KB I&D TCMs
  - 512 kB RRAM
- **Low power compute domain**
  - Cortex®-M33 and DSP + IFX NNLite for ML
  - 1 MB SRAM
- **HMI**
  - Traditional MCU HMI
  - Local voice, cloud voice
  - Vision for friction free interface & safety
- **ML**
  - NNLite
  - Advanced ML leveraging U55
- **Peripherals & IO**
  - USB, 10/100 Ethernet, CAN, SPI, UART, I2C, I3C, I2S
  - Ultra-low-power always-on analog
- **Security**
  - Secured Enclave @ 25 JIL pts, fit for ARM PSA L4

Target Applications: Appliances, Wearables, Thermostats, Residential AC, Speakers, Industrial HMI, etc.



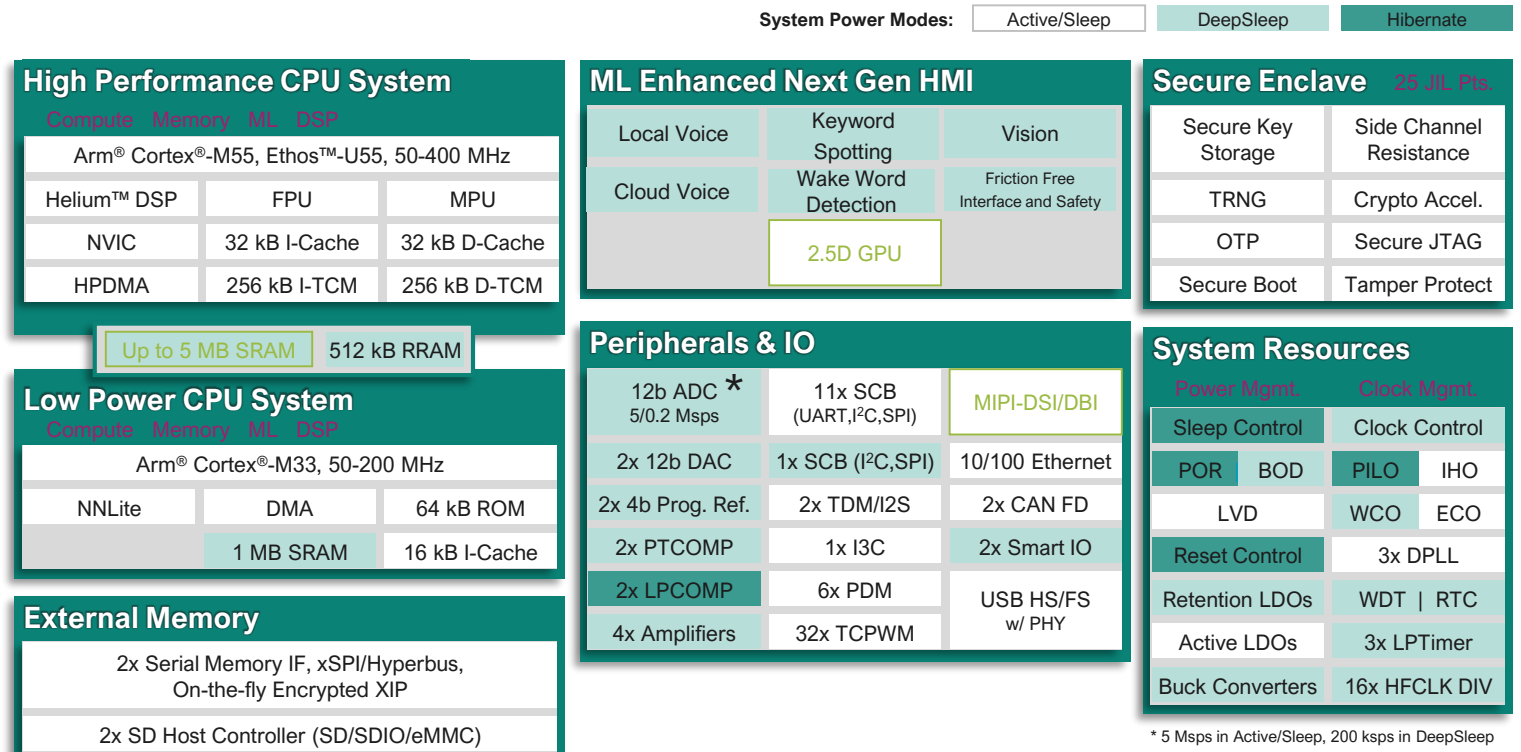
* 5 Msps in Active/Sleep, 200 ksps in DeepSleep

PSOC™ Edge E84: Next Gen, Low Power ML MCU adds GPU

Features

- **High performance, real-time compute domain**
  - Cortex®-M55 w FPU + Helium DSP + Ethos-U55 for ML
  - Up to 5 MB System SRAM, 256 KB I&D TCMs
- **Low power, real-time compute domain**
  - Cortex®-M33 and DSP + IFX NNLite for ML
  - 512 KB RRAM, 1 MB SRAM
- **HMI**
  - Traditional MCU HMI
  - Local voice, cloud voice
  - Vision for friction free interface & safety
  - Low power Graphics, up to 1024x768, MIPI-DSI/DBI
- **ML**
  - NNLite
  - Advanced ML leveraging U55 and NNLite
- **Peripherals & IO**
  - USB, 10/100 Ethernet, CAN, SPI, UART, I2C, I3C, I2S
  - Ultra-low-power always-on analog
- **Security**
  - Secured Enclave @ 25 JIL pts, fit for ARM PSA L2/L4

Target Applications: Wearables, Appliances, Thermostats, Residential AC, Speakers, Industrial HMI, etc.



Security is key in the context of AI and in our portfolio

Security is crucial for Edge AI + **Security is part of our DNA** › **New PSOC™ EDGE E8X product family**

- Multiple points of attack
- Critical IP
- Sensitive data

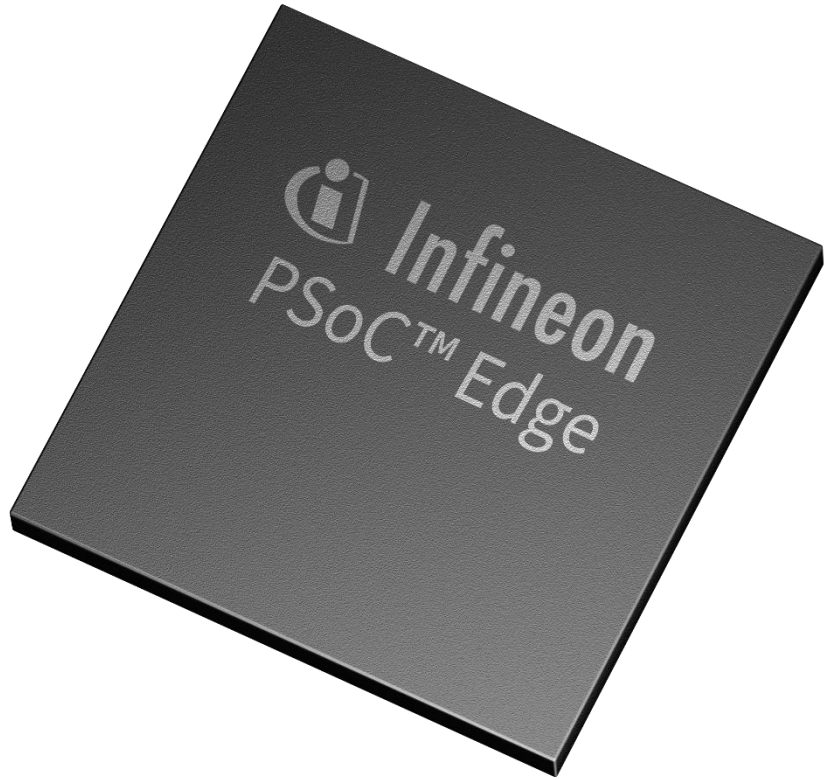
- Designed to meet highest certification level provided in the Platform Security Architecture (PSA) PSA L4 iSE
- Integrated secure enclave to support boot-time and run-time security services
- Isolation of security protection and AI acceleration computation

PSOC™ Edge – The Next Generation of ML Enabled MCUs

Available for Alpha Customers Now

For more information:
www.Infineon.com/PSOCEdge

Customer contact:
PSOCEdge@infineon.com



So, why Infineon for your AI solutions?

Enabling our customers to develop their own AI applications by providing world-leading semiconductor products, software, tools, and services.

End to end ML software solution

End-to-end solutions from training to deployment.

AI simplified

Tools and ecosystem for a simplified NN training and deployment for all level of skills.

Application specific solutions

Infineon’s HW-/SW-/Services solutions and domain knowledge covering broad range of applications in IoT, Automotive and Industrial.

Low power and high performance at the Edge

Infineon offers application-specific optimization of inference stacks for lowest power-consumption at the edge.

Reliable, safe and secured AI solutions

Offering high-quality AI systems that provide highly reliable, safe and secured AI solutions for use in real-time critical applications.

The right option for your design

One stop shop, ranging from data, ML pipe-line and chips to high-performance, low-power AI-enabled MCUs, modern sensors and easy-to-integrate AI solutions.

Imagimob Studio

Story of Imagimob Studio

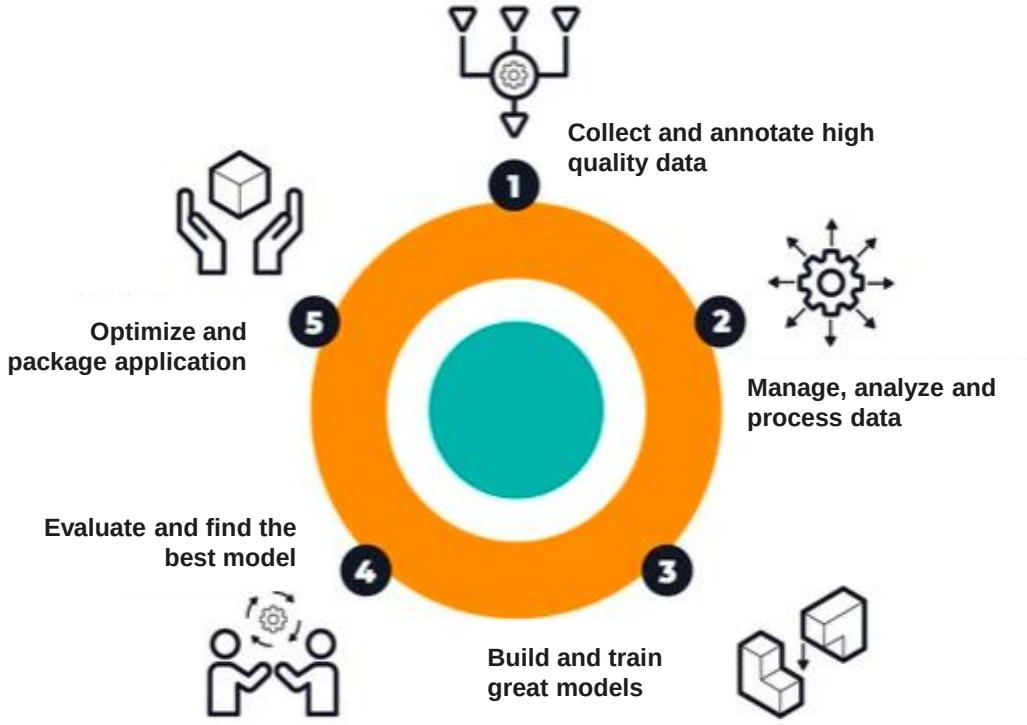
- Started early on in 2015 – there were no good tools to help us

- Developed our own platform that would help us to build solutions faster

The idea is the following;
- Make data collection easier
- Have the tools for analysing data
- Easily train models
- Get the models ready for deployment
- Have the flexibility to do this for different customer projects

Deep Dive into Imagimob Studio

Imagimob Studio helps you to take your edge AI ideas to production. The platform supports you from end to end in the ML development journey



End to end flow

- Data Collection
- Importing Data
- Labelling and annotation
- Data Management
- Pre-processing
- Model Generation
- Model Evaluation & Selection
- Packaging the Model
- Deployment

Collect & annotate high quality data | Manage data into different datasets | Build & train great models | Evaluate and find the best model | Optimize and package application | Deployment and Maintenance

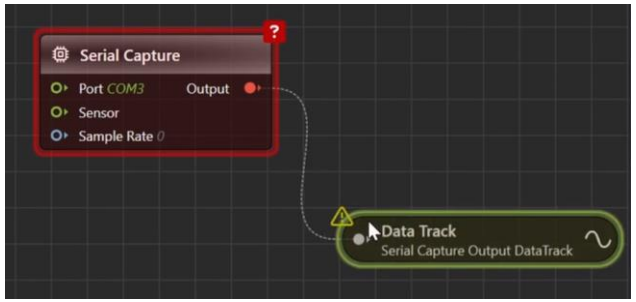
Data Collection & Labelling

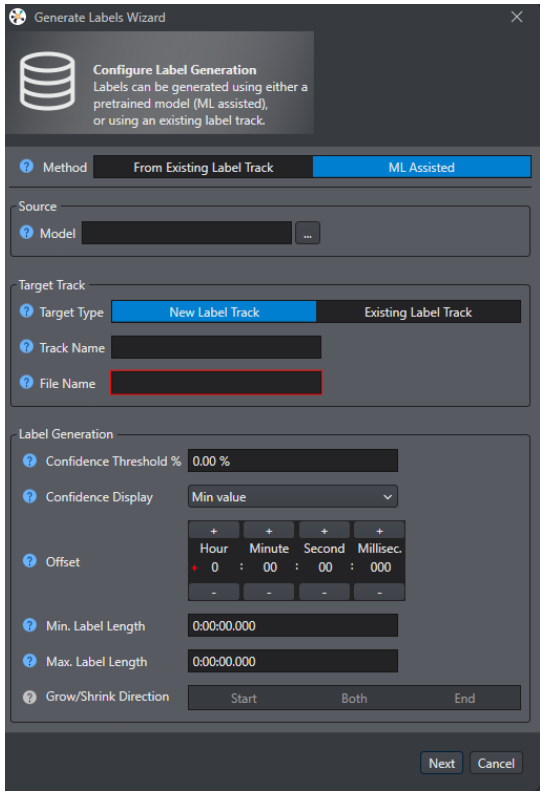
Problem:

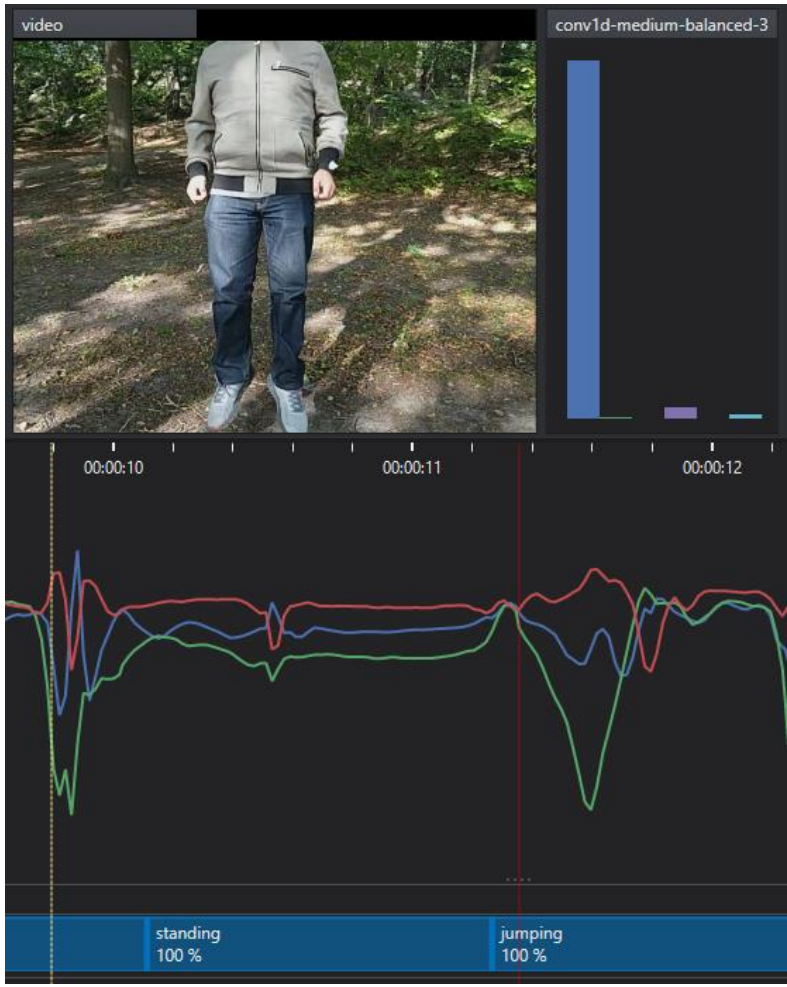
- How to get data from devices out in the field? And, from different sensor types?
- How to synchronise different signals?
- Establishing the ground truth & labeling data?

Solution:

- Standard protocol for devices to connect to
- Graphical UI for work with different components both hardware and software, on device or in the cloud
- Video used as meta data and as ground-truth
- Easy to use labelling; manual, ML-assisted and algorithm based







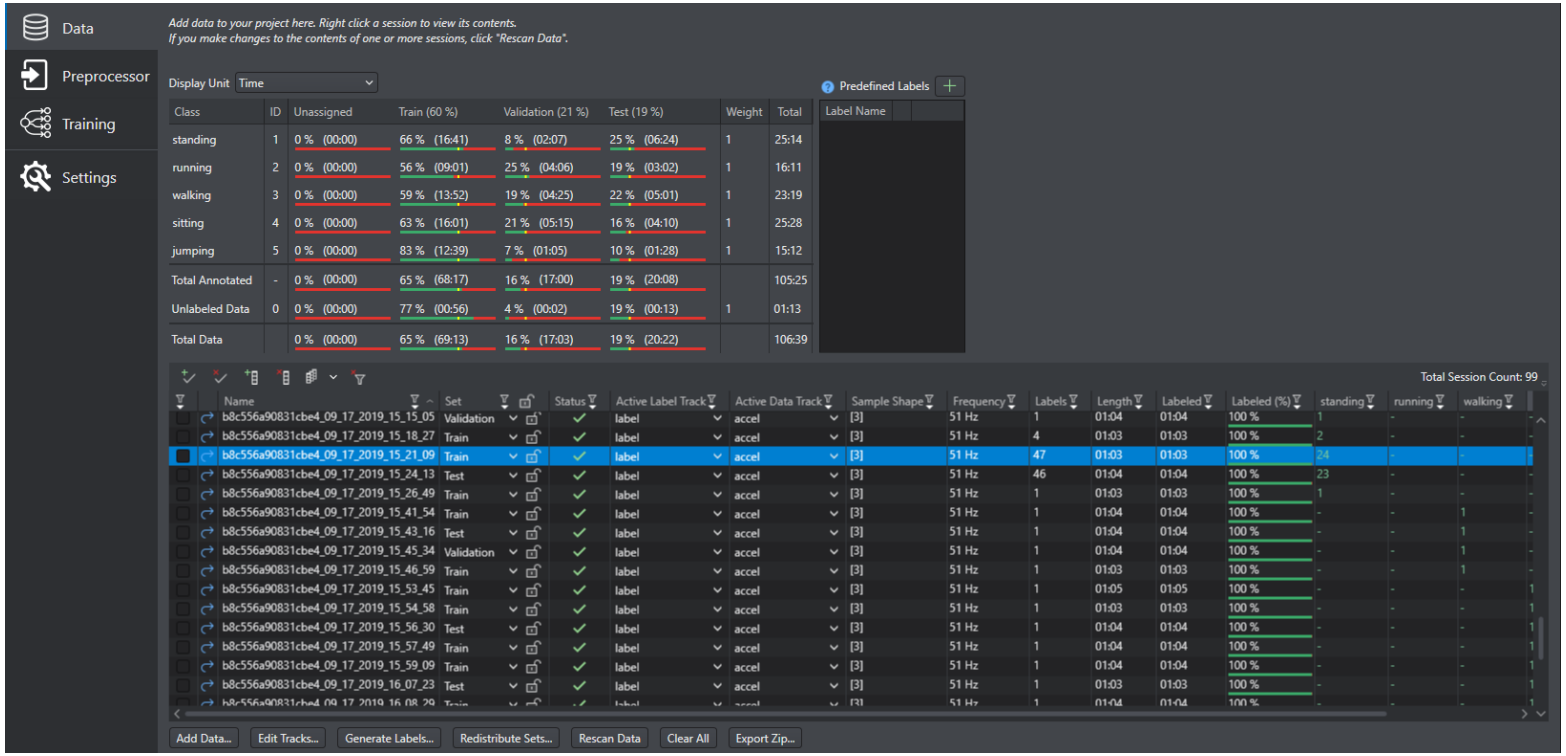
Data Management

Problem:

- How to manage large datasets?
- How do you categorise your data?
- How to identify issues in your data?

Solution:

- Dashboard view of dataset
- Displaying all properties and metrics
- With grouping and filtering functionality
- Assign to different sets and leave unassigned



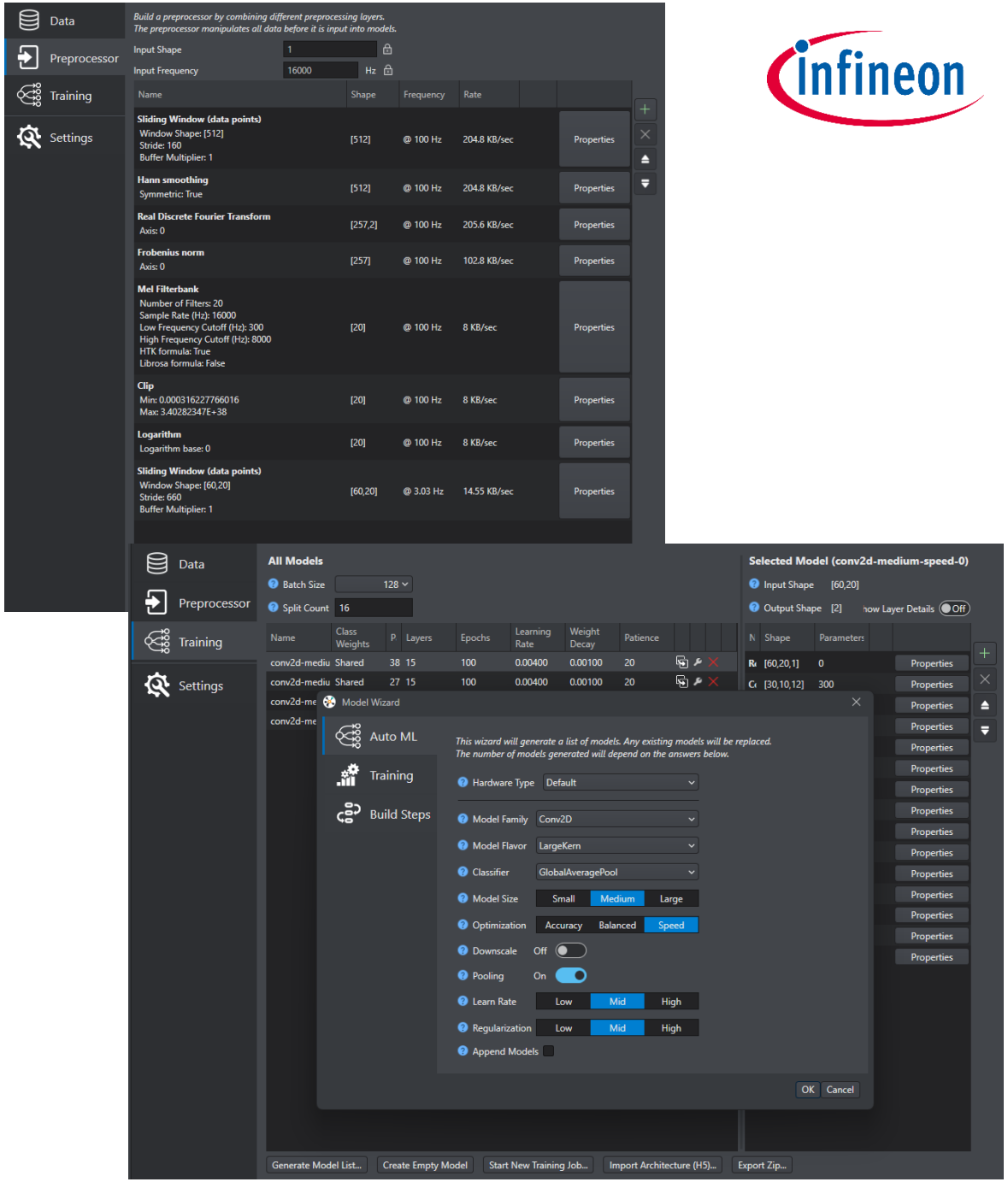
Pre-Processing & Model Generation

Problem:

- Creating a framework that ensures you don't have to recreate the wheel every-time whilst also being easy to use
- Keeping track of important valuable data processing functions
- Easy to use ML and maintaining the training loop

Solution:

- Layer library; anyone can create and add their own custom units
- Model wizard makes it easy to use models and architectures we've found to be successful



Visualisation is everything!

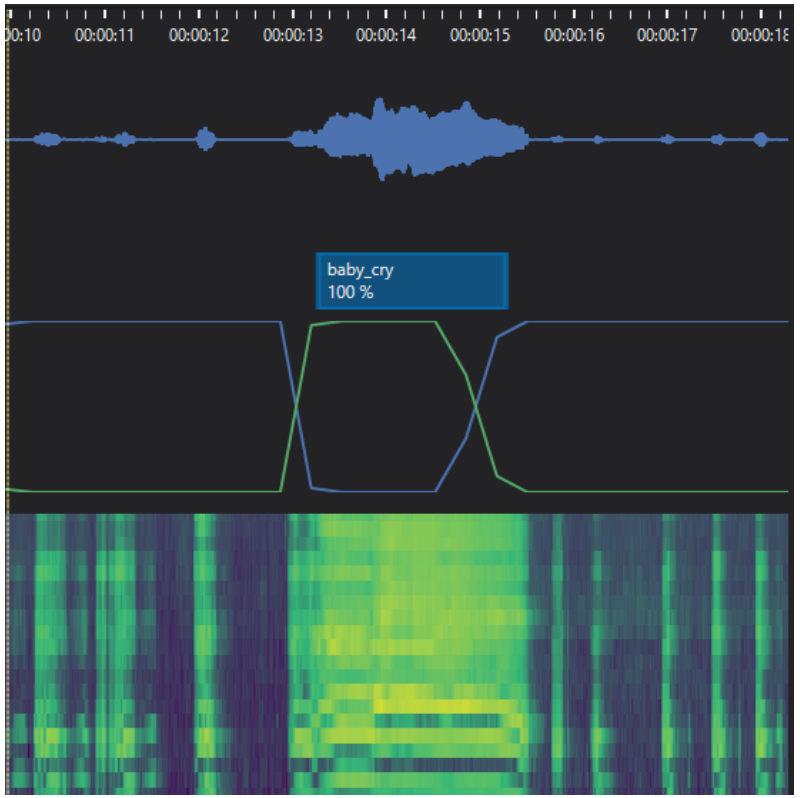
Problem:

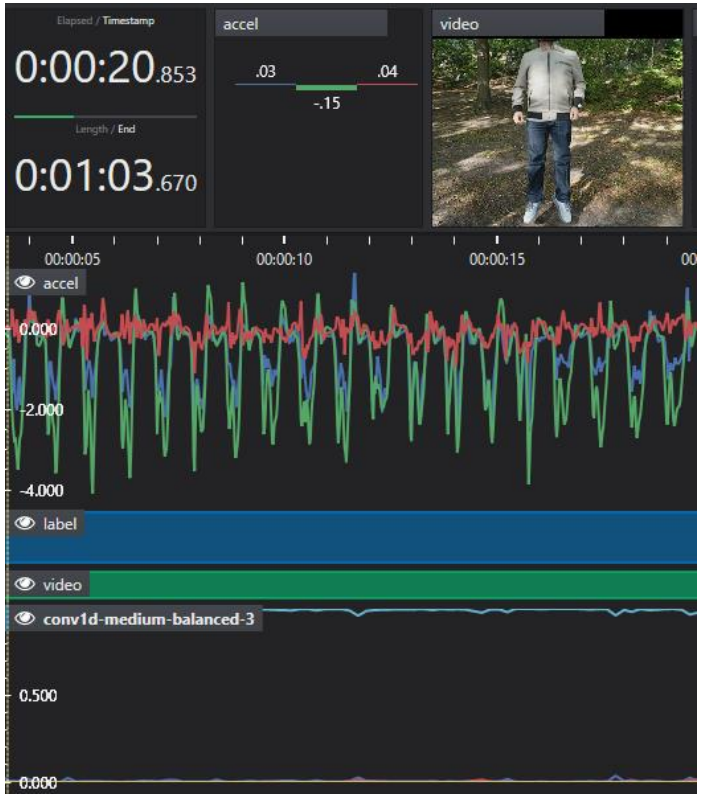
- Every problem or use-case can be quite complex
- Need to robust analysis tool

Solution:

- Ability to play through files and compare the video of the event against the sensor data for deeper understanding

- Visualising different parts of the system; from time-series data to pre-processing, labels, model output etc.





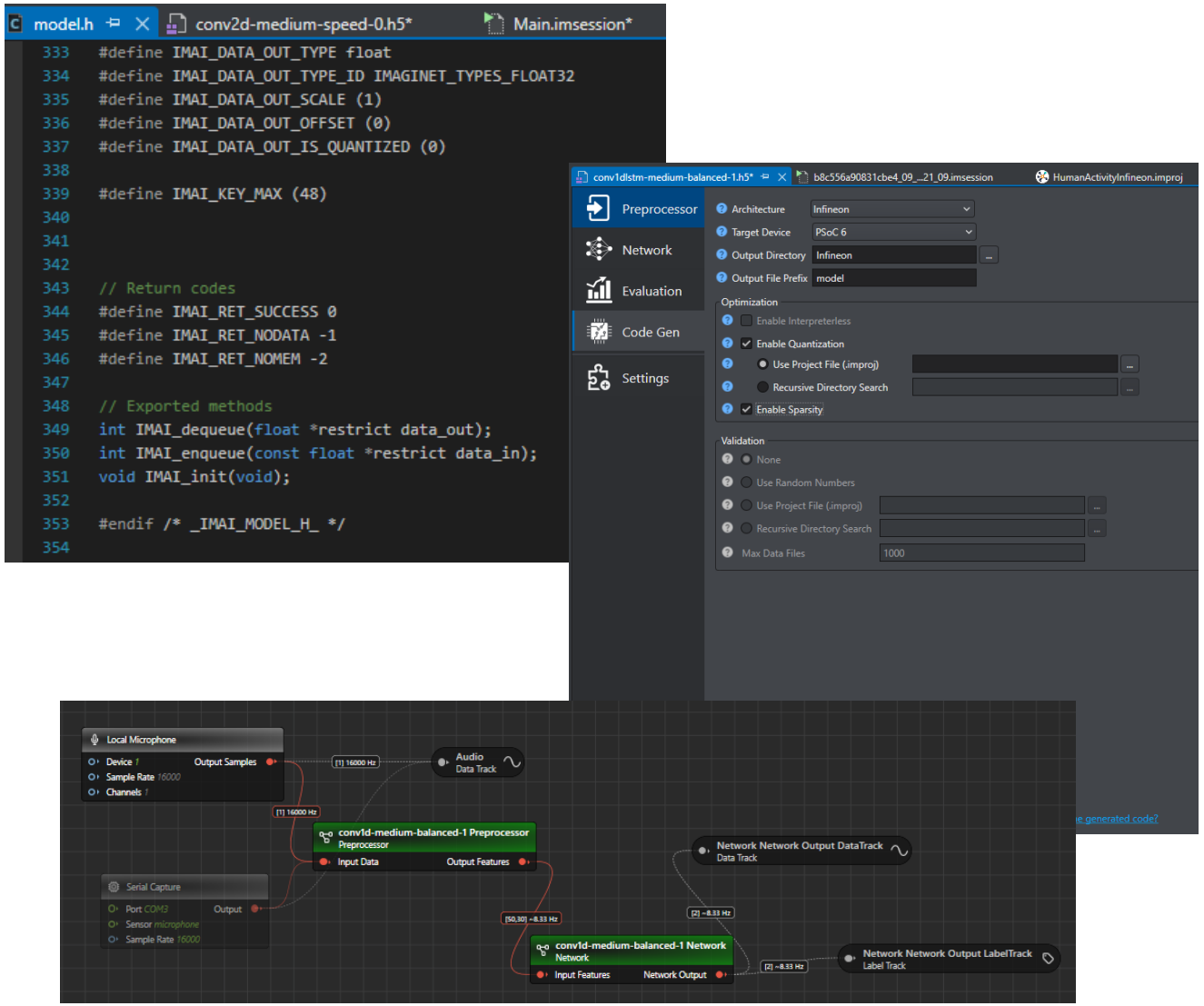
Package for the edge

Problem:

- A way to deploy our models with minimal effort
- Certainty that what you see is what you get

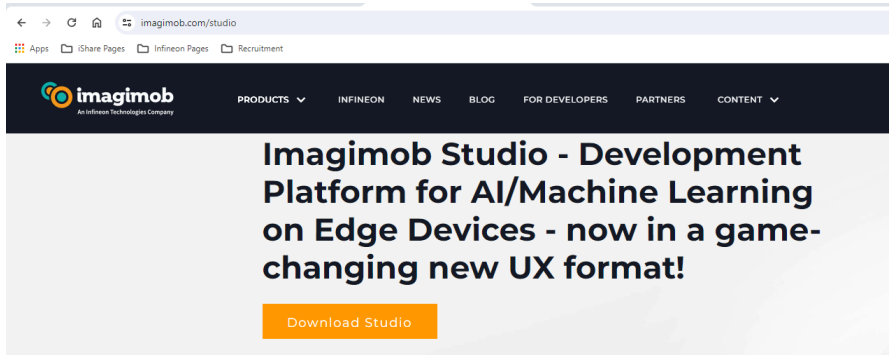
Solution:

- Model translation for all common layers
- Easy to use UI
- Easy to use API
- Ability to stream data and analyse the model in real-time on live data

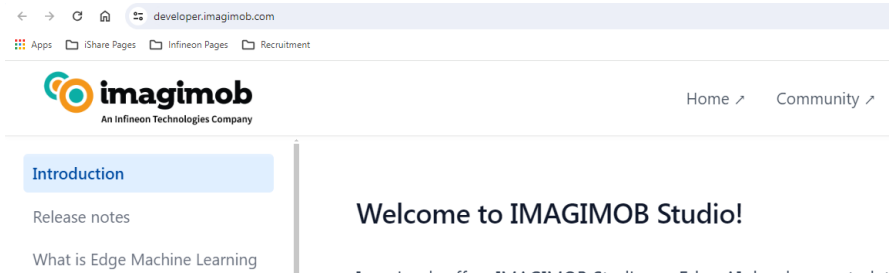


Resources

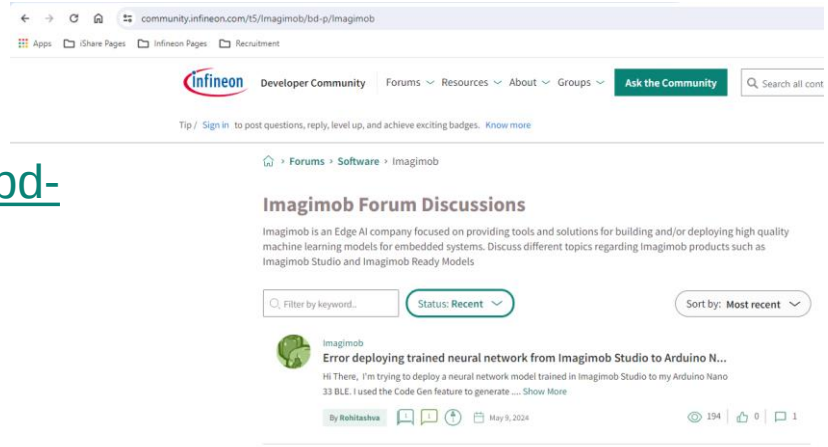
– Download Imagimob Studio at https://www.imagimob.com/studio



– Find the documentation at https://developer.imagimob.com/



– Ask questions in the community at https://community.infineon.com/t5/Imagimob/bd-p/Imagimob



It's **free** for prototyping and development


Infineon

Q&A

Thank you for attending

Infineon and you - driving the AI revolution
https://www.infineon.com/cms/en/product/promopages/artificial-intelligence

PSOC™ Edge
https://www.infineon.com/cms/en/product/microcontroller/32-bit-psoc-arm-cortex-microcontroller/32-bit-psoc-edge-arm/

Imagimob
https://www.imagimob.com/

Copyright Notice

www.tinyml.org

Copyright Notice

www.tinyml.org