

# tinyML<sup>®</sup> Talks

*Enabling Ultra-low Power Machine Learning at the Edge*

## “NanoEdge AI Studio: All On-Device Anomaly Detection for Industry 4.0”

He Huang – Data scientist, STMicroelectronics

Pierrick Autret – Product Marketing Manager, STMicroelectronics

January 30, 2024



[www.tinyML.org](http://www.tinyML.org)



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



T I N Y



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](http://www.syntiant.com)

# Platinum Strategic Partners

T I N Y



TALKS  
*webcast*

# *embed* UR



**DEPLOY VISION AI  
AT THE EDGE AT SCALE**

**SONY**



# Gold Strategic Partners

Build the  
Future of tinyML

on **arm**



T I N Y



TALKS  
*webcast*



**EDGE IMPULSE**

# The Leading Development Platform for Edge ML

[edgeimpulse.com](https://edgeimpulse.com)



Decarbonization

Digitalization



Driving decarbonization and digitalization. Together.

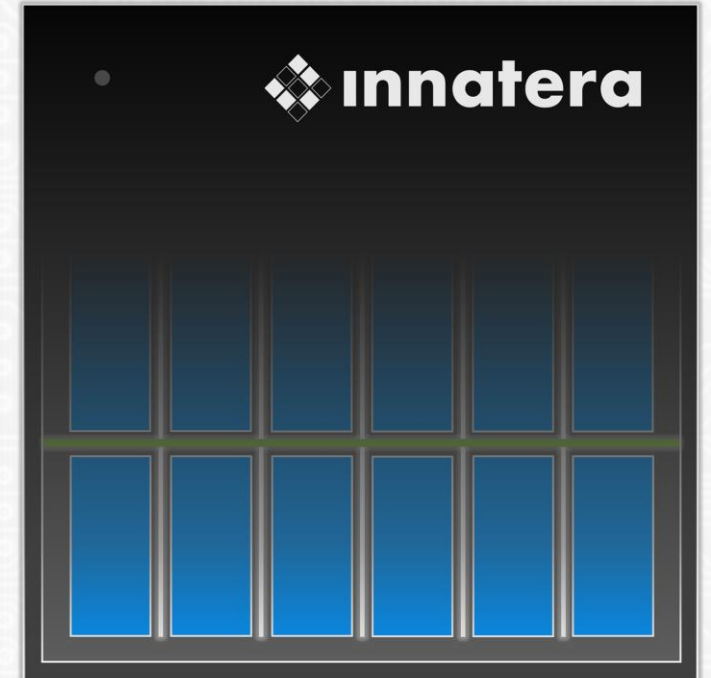
**Infineon serving all target markets as**  
**Leader in Power Systems and IoT**

[www.infineon.com](http://www.infineon.com)





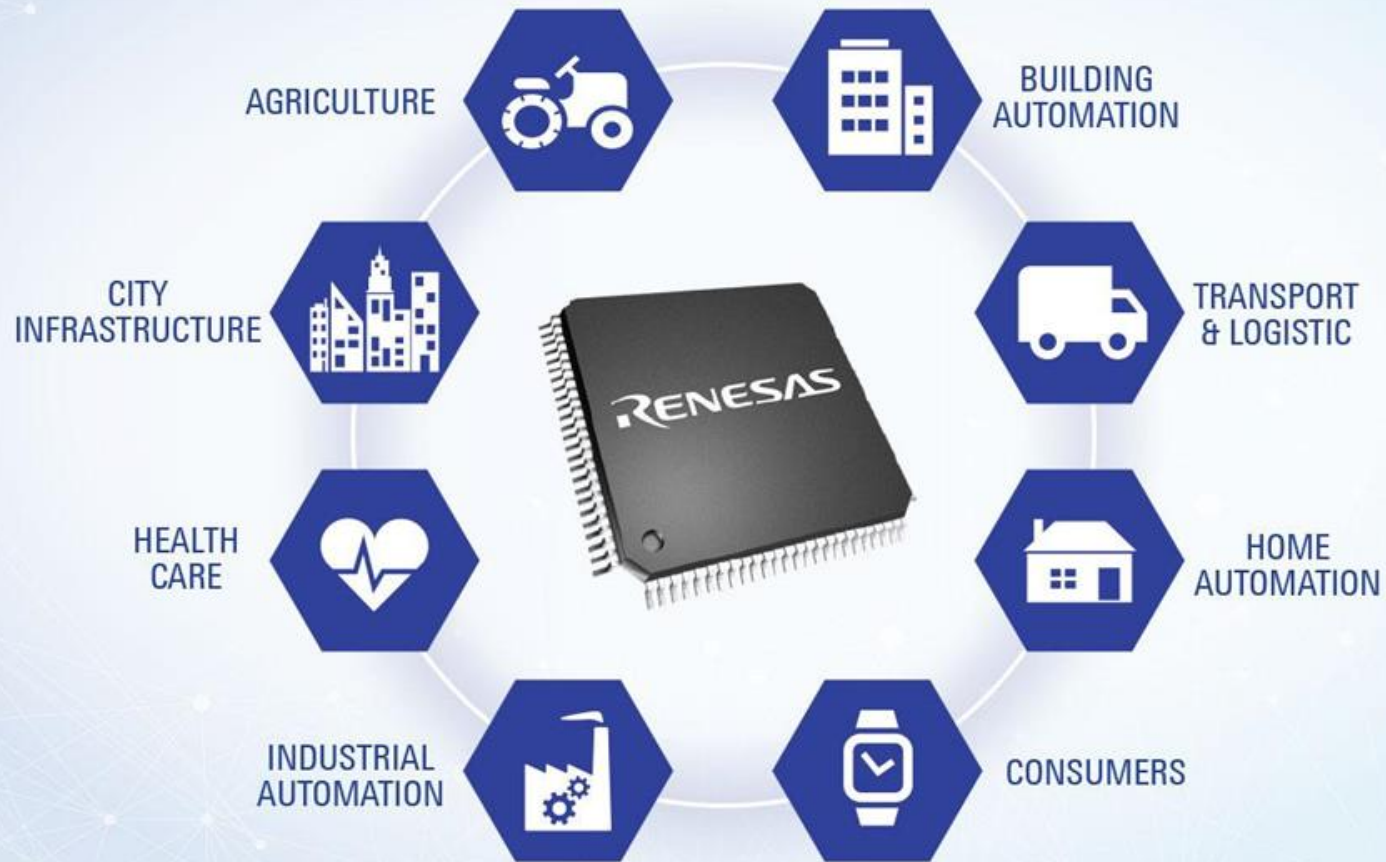
# NEUROMORPHIC INTELLIGENCE FOR THE SENSOR-EDGE



[www.innatera.com](http://www.innatera.com)



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



[renesas.com](https://www.renesas.com)



life.augmented

**STMicroelectronics provides extensive solutions to make tiny Machine Learning easy**





# ENGINEERING EXCEPTIONAL EXPERIENCES

We engineer exceptional experiences for consumers in the home, at work, in the car, or on the go.

[www.synaptics.com](http://www.synaptics.com)



T I N Y



# Silver Strategic Partners



brainchip



GREENWAVES  
TECHNOLOGIES



£Grovety Inc.



NotaAI





# Join Growing tinyML Communities:



18.9k members in  
49 Groups in 41 Countries

**tinyML - Enabling ultra-low Power ML at the Edge**

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



4.2k members  
&  
14.5k followers

**The tinyML Community**

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







Subscribe to  
**tinyML YouTube Channel**  
 for updates and notifications  
*(including this video)*

[www.youtube.com/tinyML](http://www.youtube.com/tinyML)



**tinyML**  
4.33K subscribers

**11.4k subscribers, 659 videos with 420k views**

HOME VIDEOS PLAYLISTS COMMUNITY CHANNELS ABOUT

106 views · 4 days ago	138 views · 4 days ago	54 views · 4 days ago	47 views · 4 days ago	132 views · 4 days ago	137 views · 4 days ago
122 views · 4 days ago	262 views · 2 weeks ago	511 views · 3 weeks ago	229 views · 3 weeks ago	265 views · 3 weeks ago	286 views · 1 month ago
351 views · 1 month ago	462 views · 2 months ago	374 views · 2 months ago	133 views · 2 months ago	287 views · 2 months ago	336 views · 2 months ago
378 views · 2 months ago	214 views · 2 months ago	448 views · 2 months ago	159 views · 2 months ago	190 views · 2 months ago	545 views · 2 months ago

# tinyML Research Symposium

## April 22, 2024

### Call for Papers



**Research Symposium - April 22, 2024**

The tinyML research symposium serves as a flagship venue for related research at the intersection of machine learning applications, algorithms, software, and hardware in deeply embedded machine learning systems.

[Call for Papers](#)



# 2023 Edge AI Technology Report

The guide to understanding the state of the art in hardware & software in Edge AI.





# Reminders

Slides & Videos will be posted tomorrow



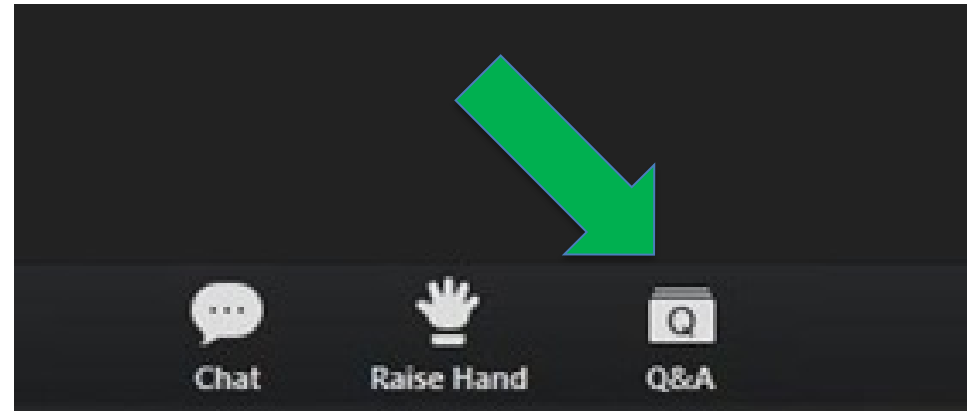
[tinyml.org/forums](https://tinyml.org/forums)



[youtube.com/tinyml](https://youtube.com/tinyml)



Please use the Q&A window for your questions





## He Huang



He Huang is a data scientist and researcher with expertise in embedded Machine Learning. He holds a Ph.D. in electronics from INSA Toulouse in France, earned in 2015 and completed a post-doctoral research fellowship at Politecnico di Torino in Italy in 2016. As the former chief data scientist at French start-up Cartesiam from 2017, He played a pivotal role in the development of NanoEdge AI Studio, an innovative platform for on-device machine learning at the edge. Currently, he serves as senior staff data scientist at STMicroelectronics, where he continues to advance the field of machine learning and develop cutting-edge solutions for Industry 4.0 applications.





## Pierrick Autret



Pierrick Autret has been Artificial Intelligence Solutions Manager in STMicroelectronics' microcontroller division since 2019. With a master's degree in engineering specializing in Embedded Systems from the Ecole Polytechnique and a master's degree in Marketing, Economics and Management from the IAE d'Aix-en-Provence, he began his career as Hardware Product Owner on the STM32 series before developing ST's software offering.



life.augmented

# NanoEdge AI Studio: all on-device anomaly detection for Industry 4.0

**He HUANG, Pierrick AUTRET**  
AI solutions group, STMicroelectronics

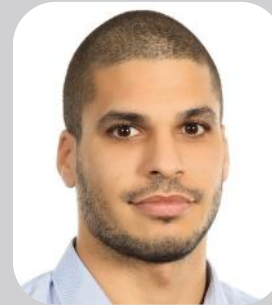
✉ [he.huang1@st.com](mailto:he.huang1@st.com), [pierrick.autret@st.com](mailto:pierrick.autret@st.com)

# Who we are



He HUANG (Henri)

Data scientist  
AI Solutions group, STMicroelectronics  
Paris, France



Pierrick AUTRET

Product Marketing Manager  
AI solutions group, STMicroelectronics  
Rousset, France



# What edge AI can bring to embedded systems



**Ultra-low latency**  
Real-time applications

**01** **Reduced data transmission**  
**10** Generate meaningful information



**Enhanced privacy and security**  
No data sharing in the cloud



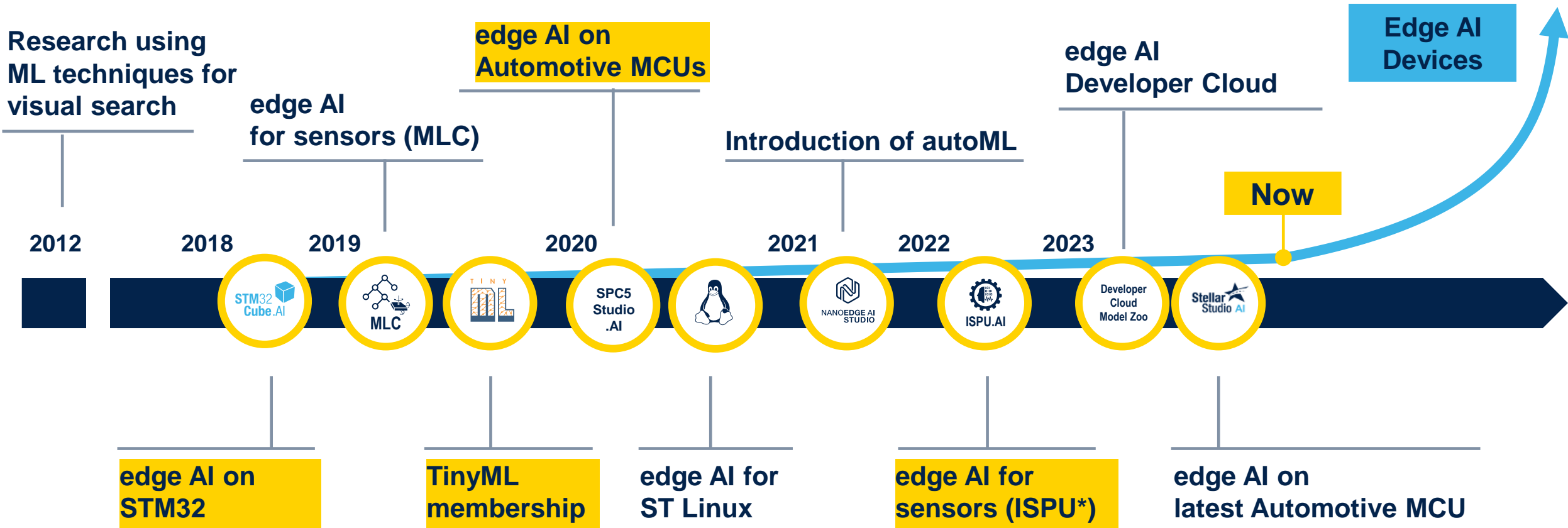
**Lower cost of inference to**  
enable a new class of operations



**Sustainable on energy**  
Low data, Low power



# More than a decade of research, development, and deployment



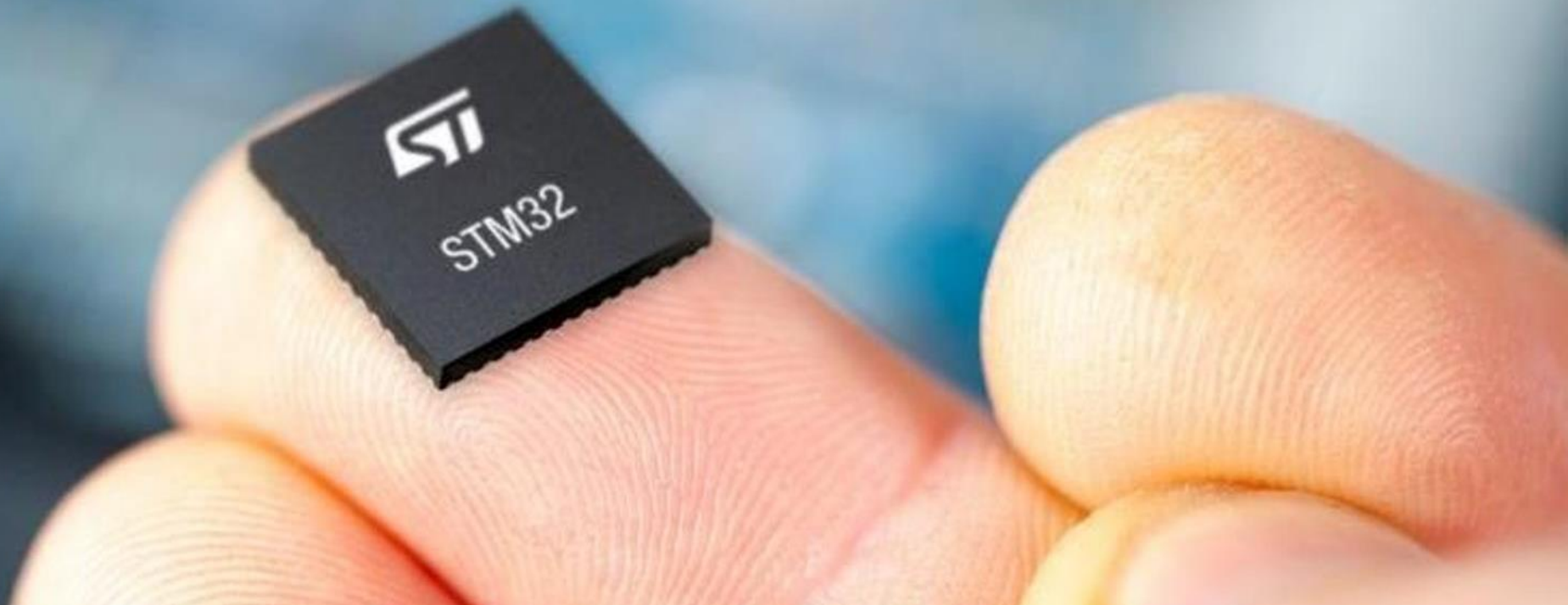


Industrial pumps **learn** their own optimal mode of operation and **detect anomalies** by themselves

**220%**  
reduced  
downtime

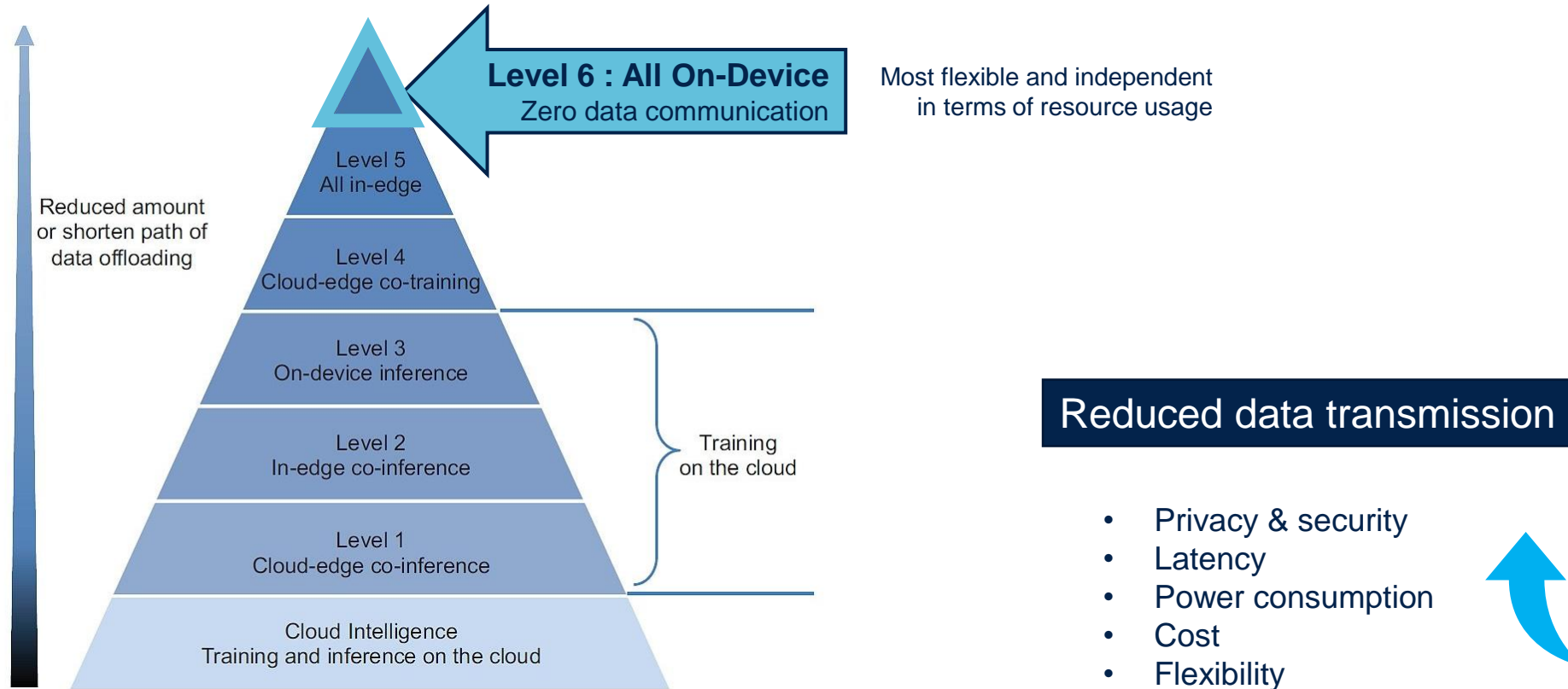


# All on-device AI for anomaly detection





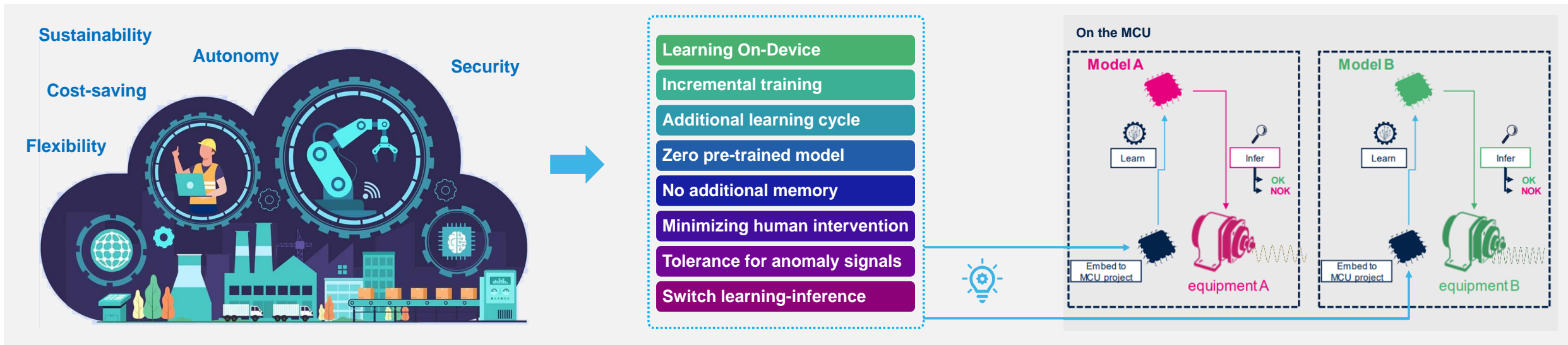
# Why do we need all on-device AI?



Six-level rating for EI (Edge Intelligence)

Z. Zhou, X. Chen, E. Li, L. Zeng, K. Luo and J. Zhang, "Edge Intelligence: Paving the Last Mile of Artificial Intelligence With Edge Computing," in Proceedings of the IEEE, vol. 107, no. 8, pp. 1738-1762, Aug. 2019

# How do we design our solution?



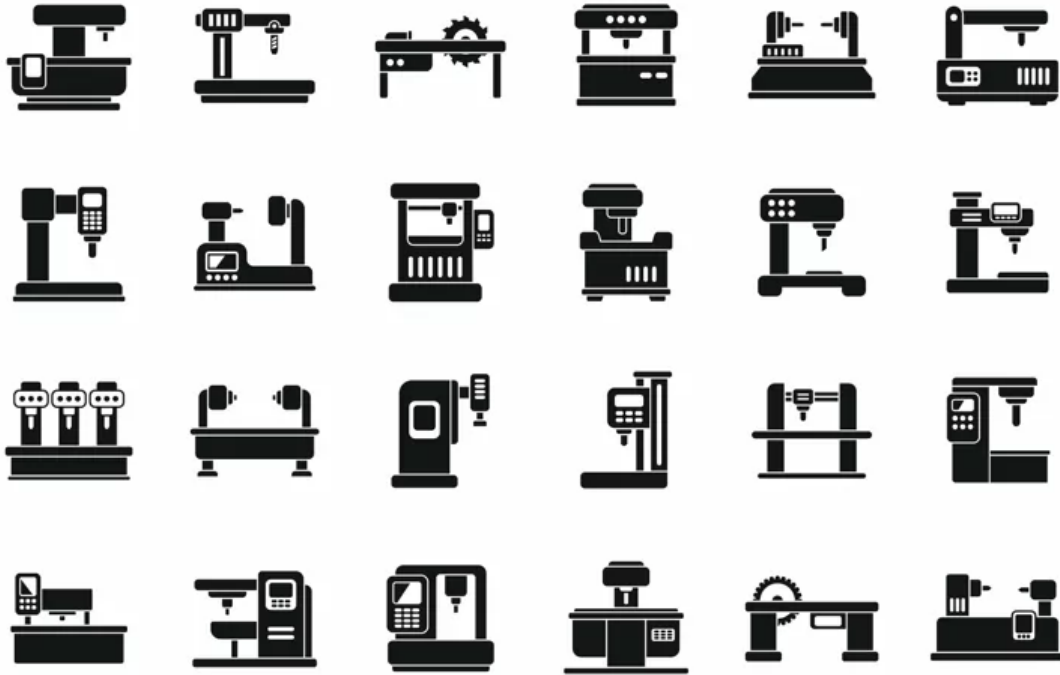
Step 1: analyze the requirements of industrial predictive maintenance

Step 2: define the algorithm criteria

Step 3: implement on-device learning: individual adaptive AI model

[Predictive Maintenance: solutions for new business models \(device-insight.com\)](https://device-insight.com)

# The needs of industrial predictive maintenance



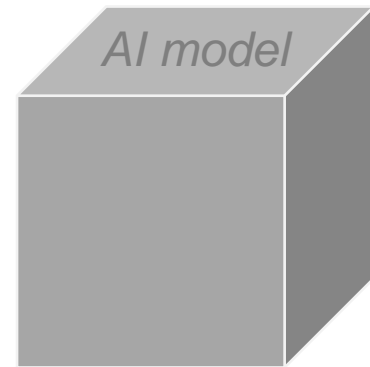
Billions of machines, functions, environments, updates, and maintenance requirements.

**Edge AI models need to be universal and adaptive**

Every machine can benefit from its own edge AI model



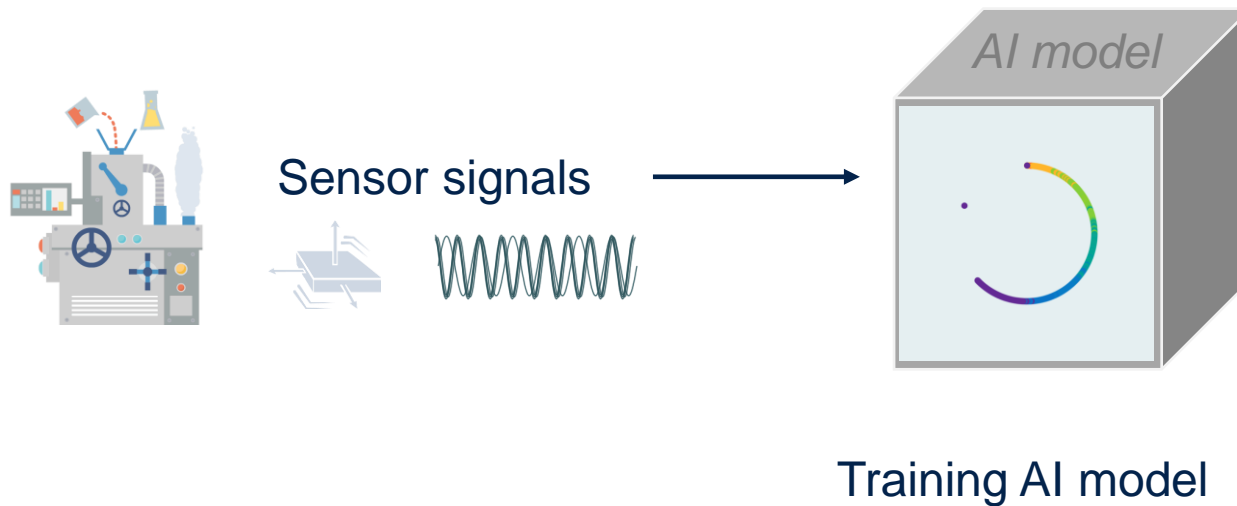
# On-device learning in NanoEdge AI Studio



Empty model

Without pre-trained knowledge

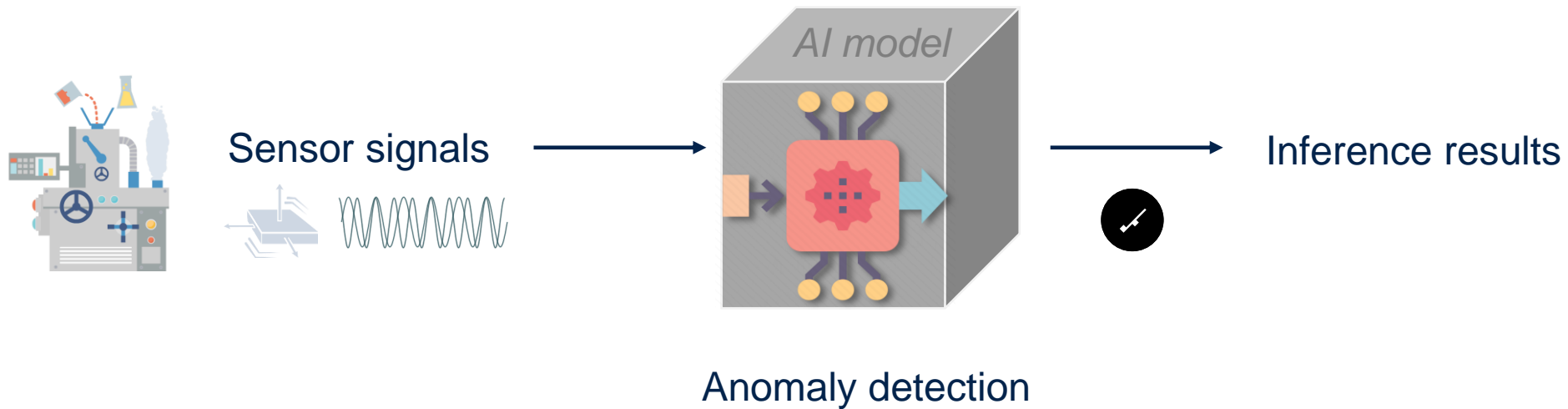
# On-device learning in NanoEdge AI Studio



No signal history stocked

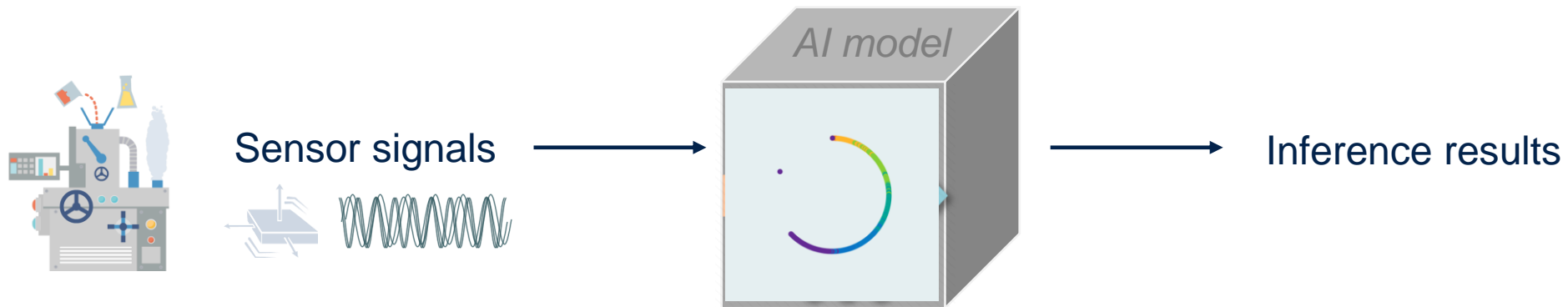
Only AI model updated

# On-device learning in NanoEdge AI Studio





# On-device learning in NanoEdge AI Studio



Switch learning – inference

# On-device learning in NanoEdge AI Studio

## 8 criteria

- Zero pretrained model
- Incremental training
- Additional learning cycle
- Switch learning – inference
- No additional memory
- Adaptive on-device learning
- Minimizing human intervention
- Tolerance for anomaly signals

A simple example: the average value

$$m_n = \frac{1}{n} \sum_{i=1}^n a_i$$

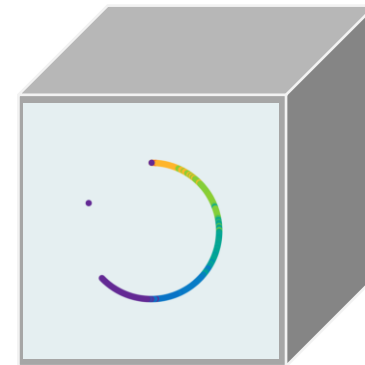


$$m_n = \frac{(n-1)m_{n-1} + a_n}{n}$$

# On-device learning in NanoEdge AI Studio

## 8 criteria

- Zero pretrained model
- Incremental training
- Additional learning cycle
- Switch learning – inference
- No additional memory
- Adaptive on-device learning
- Minimizing human intervention
- Tolerance for anomaly signals



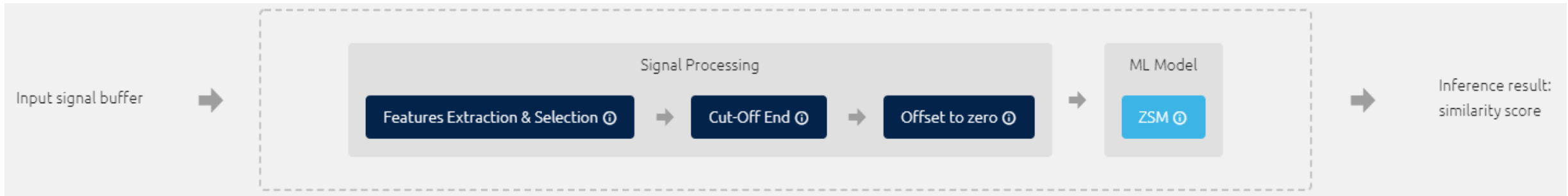
NANOEDGE AI  
STUDIO 

Incremental on-device machine learning



# On-device learning in NanoEdge AI Studio

Algorithm flowchart example 1:



Algorithm flowchart example 2:

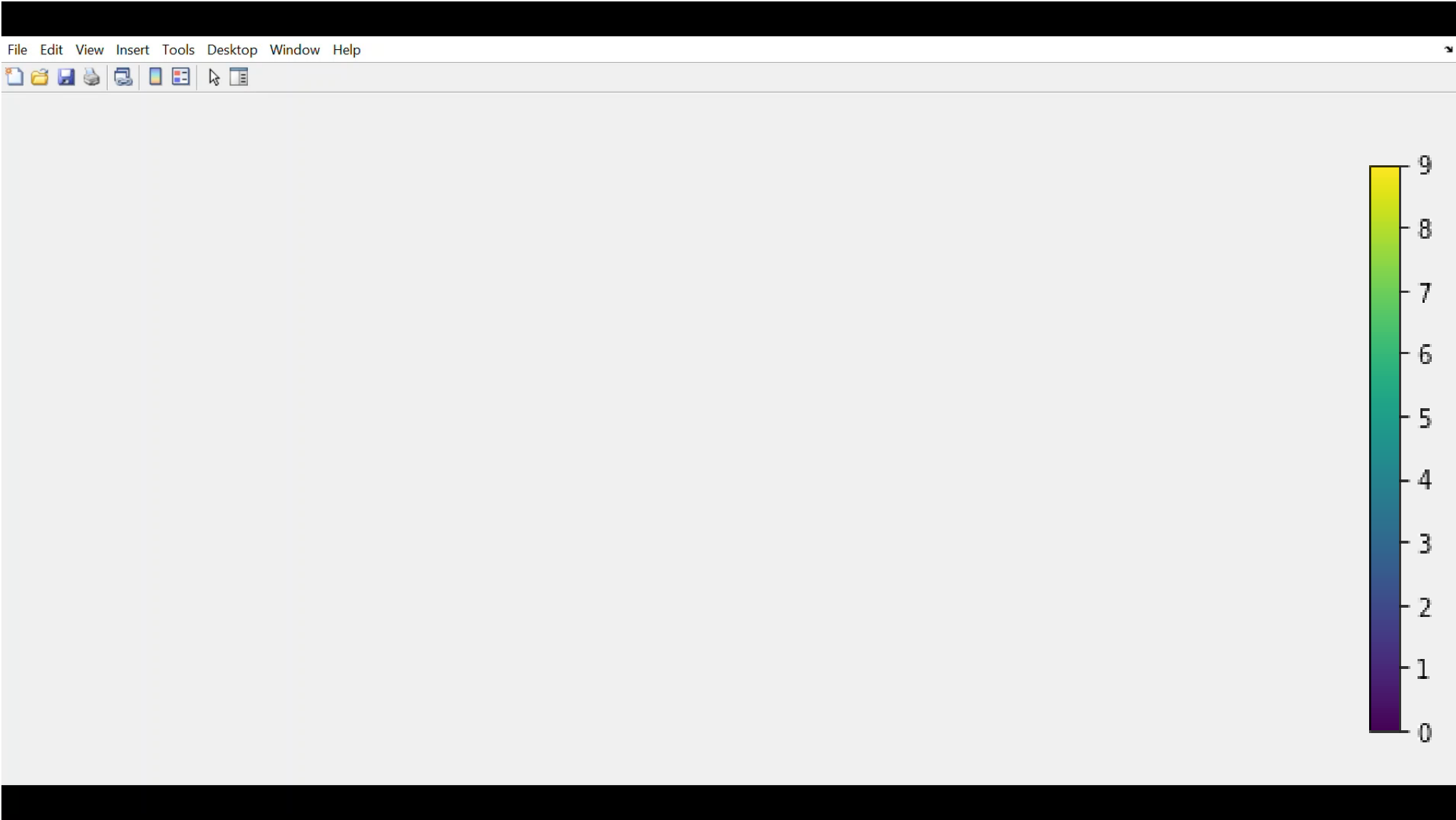


Input

Algorithms

Output

# Our on-device learning algorithms

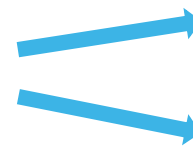


# On-device learning vs classification



Machines used for training

Dataset regular / abnormal



On-device learning

Adaptive model

Classification

Static model

# On-device learning vs Classification

Dataset for test



Machines used for training

Dataset regular / abnormal (new)



Machines not used for training

Dataset regular / abnormal

Adaptive model

Static model

On-device learning

Classification

?	?
?	?



# On-device learning vs Classification

Machines used for training

Machines not used for training

	On-device learning	Classification
fan_id_02_6dB		
fan_id_04_6dB		
fan_id_06_6dB		

Benchmarked with 

# On-device learning vs Classification

Machines used for training

Machines not used for training

	On-device learning	Classification
fan_id_02_6dB		
fan_id_06_6dB		
fan_id_04_6dB		

Benchmarked with 

# On-device learning vs Classification

Machines used for training

Machines not used for training

On-device learning	Classification
	More suitable for known individuals
More adaptive for new individuals	

# On-device learning vs Classification

If the dataset is sufficiently representative for all dispersions



Classification  
static model

If the model is used for a new machine, a new individual  
an updating behaviors  
a new measurement position  
a new installation environment

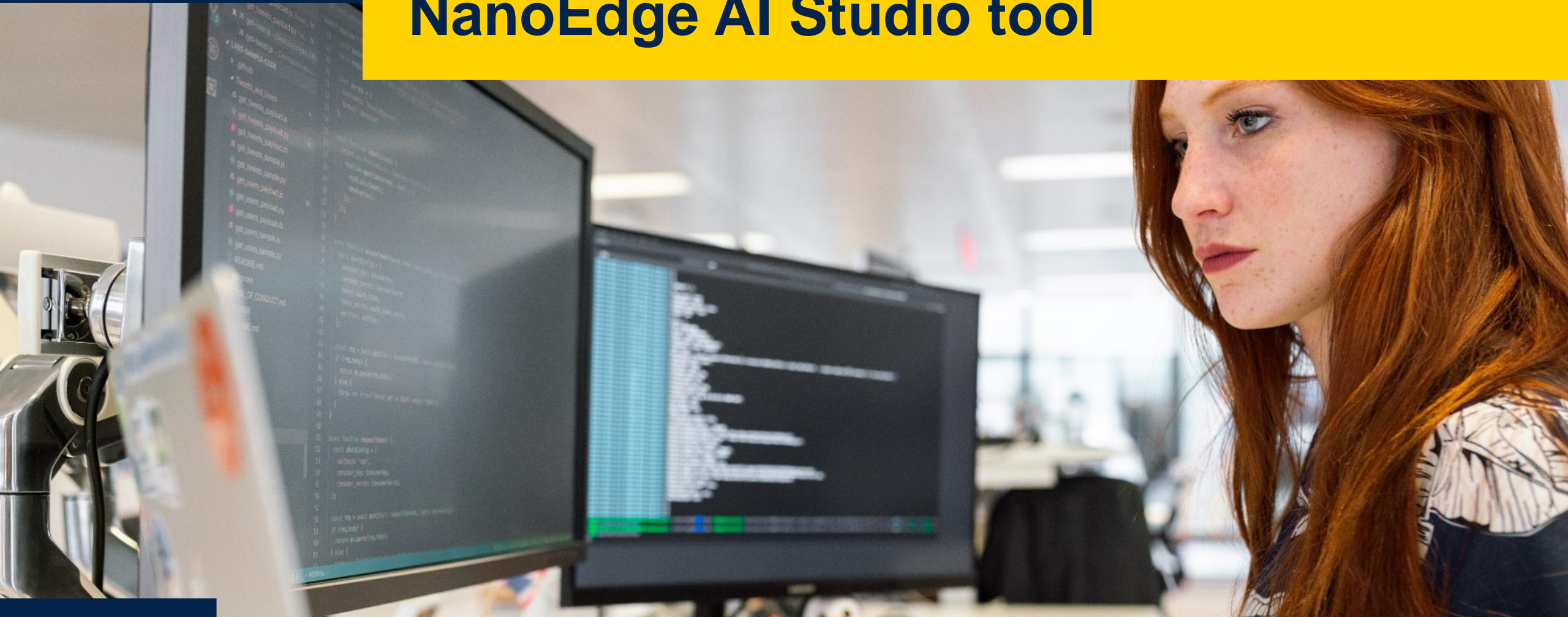


On-device learning  
adaptive model

...

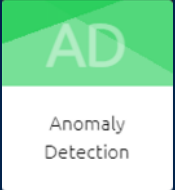


# Create your application with the free NanoEdge AI Studio tool



life.augmented

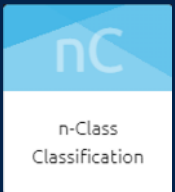
# State-of-the-art machine learning for smarter products



“

**I want to anticipate product failures**

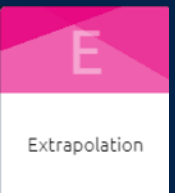
Industrial pumps **learn** their own optimal mode of operation and **detect** anomalies **by themselves**



“

**I want to identify the activity, the environment, the usage**

Smart watches classify human activity recognition without using connectivity features



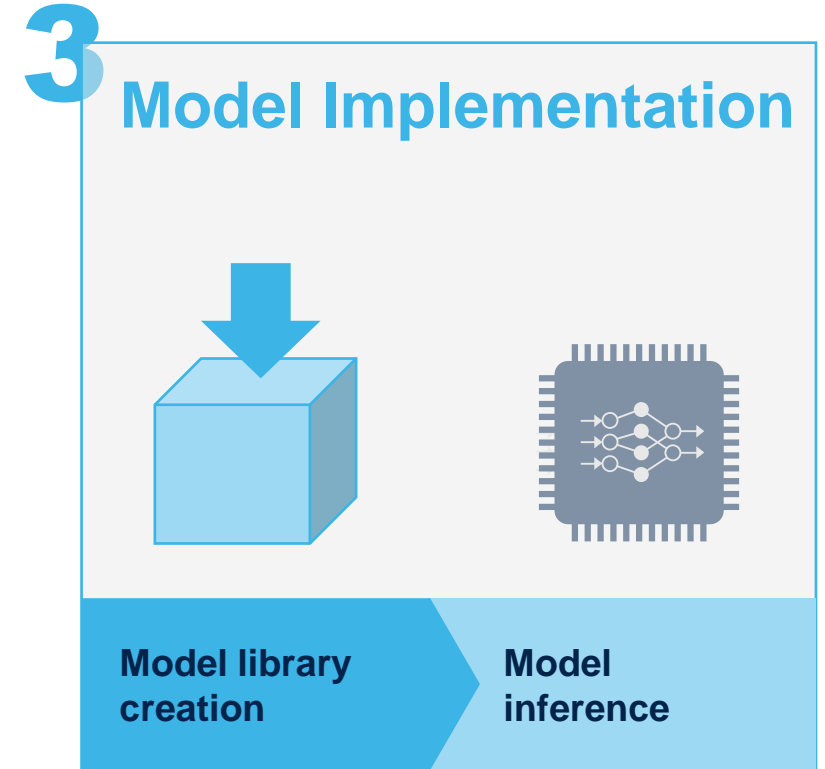
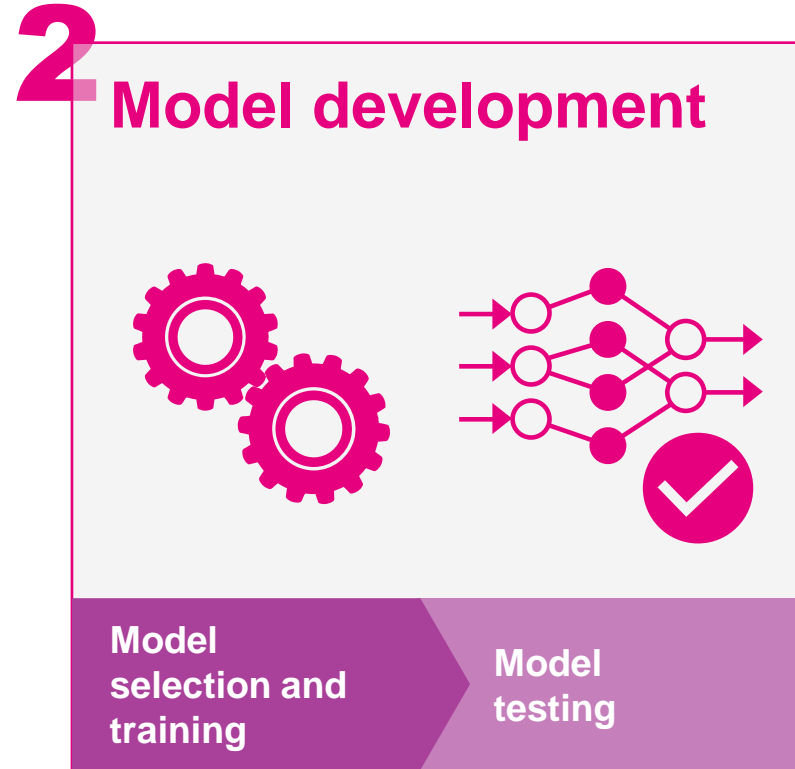
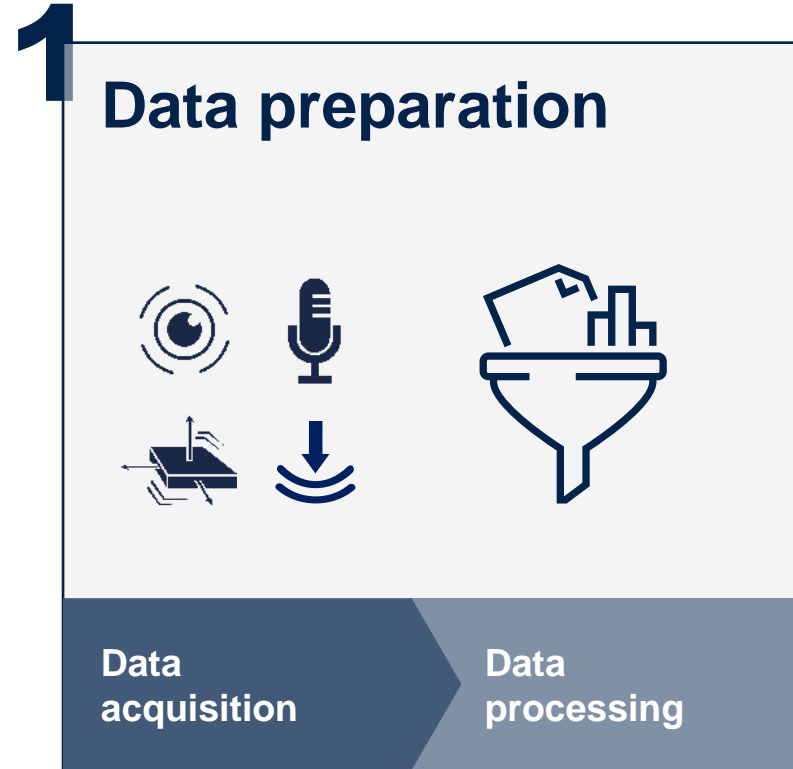
“

**I need to predict future states**

Washing machine uses motor control algorithms to **weigh clothes** and optimize water, detergent, and energy used **without additional sensors**



# AI development workflow – NanoEdge AI



# AI development workflow – NanoEdge AI

1

## Data preparation

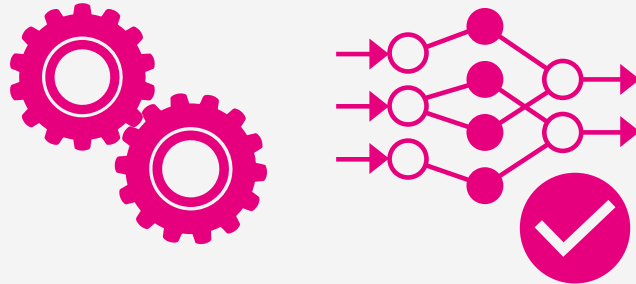


Data acquisition

Data processing

2

## Model development

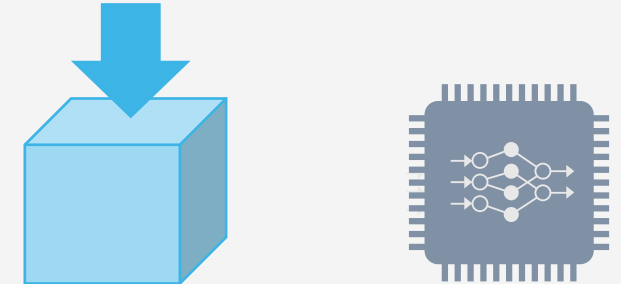


Model selection and training

Model testing

3

## Model Implementation



Model library creation

Model inference

Automated edge AI software

NANOEDGE AI STUDIO 





# All the tools needed for data preparation



“ I want to optimize my startup parameters ”

**Sampling Finder**

Step 1: Select files. Each file should include one class, at least 2 classes are needed. Continuous data with one column by axis. For example: 1 axis = 1 column, 2 axes = 2 columns.

Files input: 0 file(s) (0 B)

Step 2: Set number of axes used. Number of axes: 3

Step 3: Set sampling frequency used. Max frequency: 1000 Hz

Step 4: Set minimal frequency to test. Min frequency: fs/32

Step 5: START SAMPLING FINDER

**Welcome to Sampling Finder**

This tool is designed to help you find the appropriate sampling rate and sampling size to use for your project : In artificial intelligence applications involving time series signals with sensors, choosing the optimum sampling rate and sampling size is often a difficult task. The "Sampling Finder" enables you to automatically determine the smallest possible sampling size and sampling duration, while maintaining the highest distinction percentage.

Sampling size (by axis)	666.7Hz (fs/1)	333.35Hz (fs/2)	166.675Hz (fs/4)	83.3375Hz (fs/8)	41.66875Hz (fs/16)	20.834375Hz (fs/32)
4096	100%	100%	100%	100%	100%	100%
2048	100%	100%	100%	100%	100%	100%
1024	98%	100%	100%	100%	100%	100%
512	94%	95%	100%	99%	100%	99%
256	93%	91%	94%	99%	100%	100%
128	79%	94%	92%	94%	98%	97%
64	72%	77%	85%	90%	94%	93%
32	71%	72%	79%	88%	84%	90%
16	66%	70%	71%	77%	79%	73%

**Best recommendation**

- Sampling size (by axis): 1024
- Sampling rate: 208.34375Hz (fs/32)
- Sampling duration: 4.915s

# All the tools needed for data preparation



DL  
Data Logger

“

I want the tools to easily create my dataset”



# All the tools needed for data preparation



“

I want to quickly and efficiently clean my data”

**File** **Action** **Result**

**Log\_ISPU\_Spacer\_2022-06-13-163838.log**

Ignore first header line

182 Lines 768 Columns

Delimiter: Space

File preview

1	2	3	4	5	6	7	8
-453	-1523	15386	517	-154	16106	1068	729
659	95	16368	1006	388	16779	313	-164
722	1012	15909	552	1007	17633	-461	-510
272	269	17697	-851	-911	17448	-836	-1172
45	-463	17629	-727	-1399	16834	-349	-1244

Rows per page: 5 1-5 of 100

Add file(s)

Drop files or click to import

**Action**

Extract lines

1 
1
182

Extract 182 lines

**RUN**

1,234 1,234 1,234  
 1,234 1,234 1,234  
 1,234 1,234 1,234  
 1,234 1,234 1,234  
 1,234 1,234 1,234

Remove column(s)

Change columns number

Shuffle

**(Log\_ISPU\_Spacer\_2022-06-13-163838.log)**

Log\_ISPU\_Spacer\_2022\_132ce\_0\_extract\_lines\_0

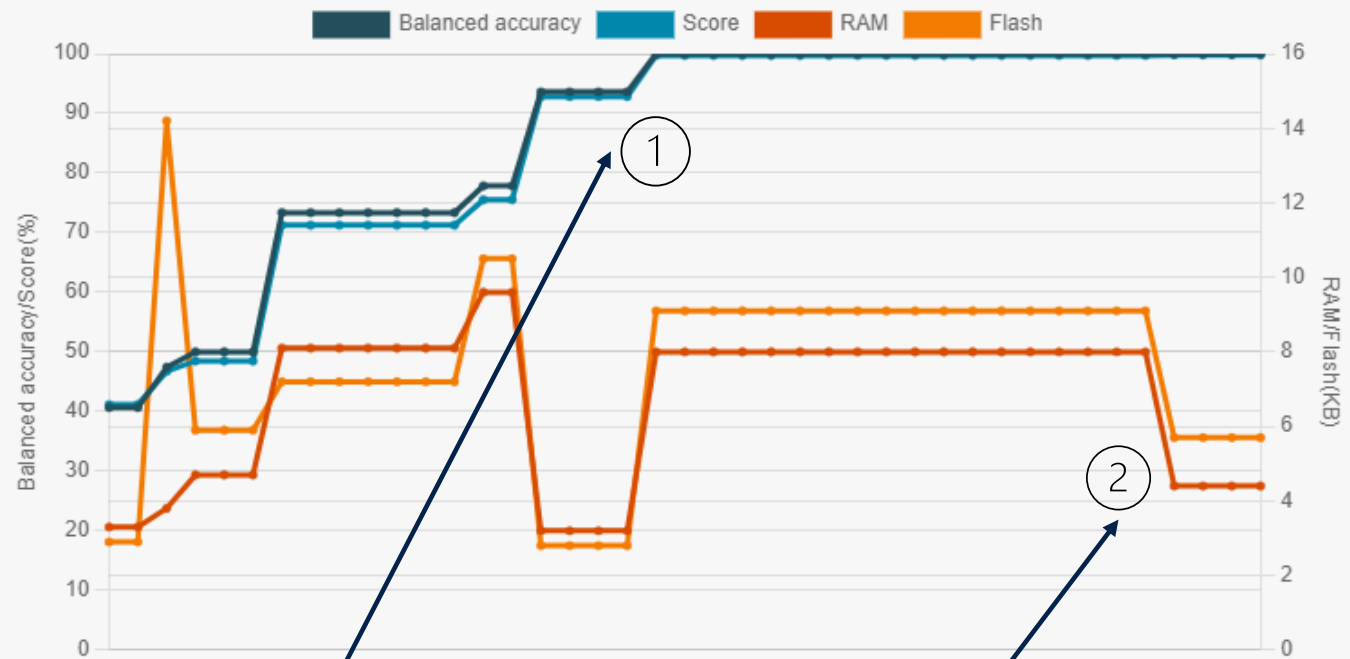
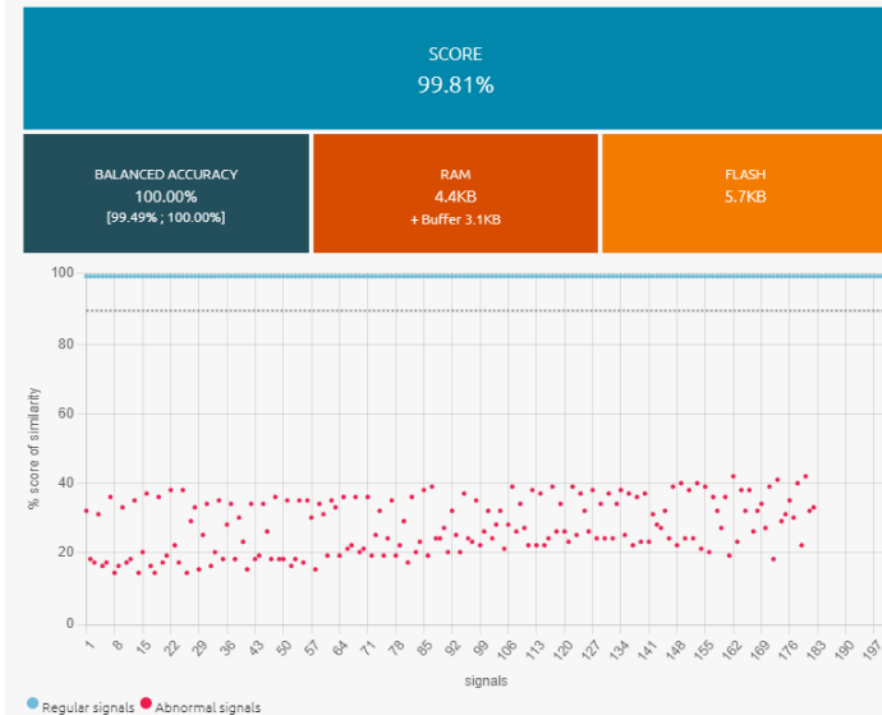
182 Lines 768 Columns

**RUN NEW ACTION** **SAVE AS**

DM  
Data Manipulation

# Get the best ML algorithm for your application

Internal benchmarking tool picks the best algorithm for your data

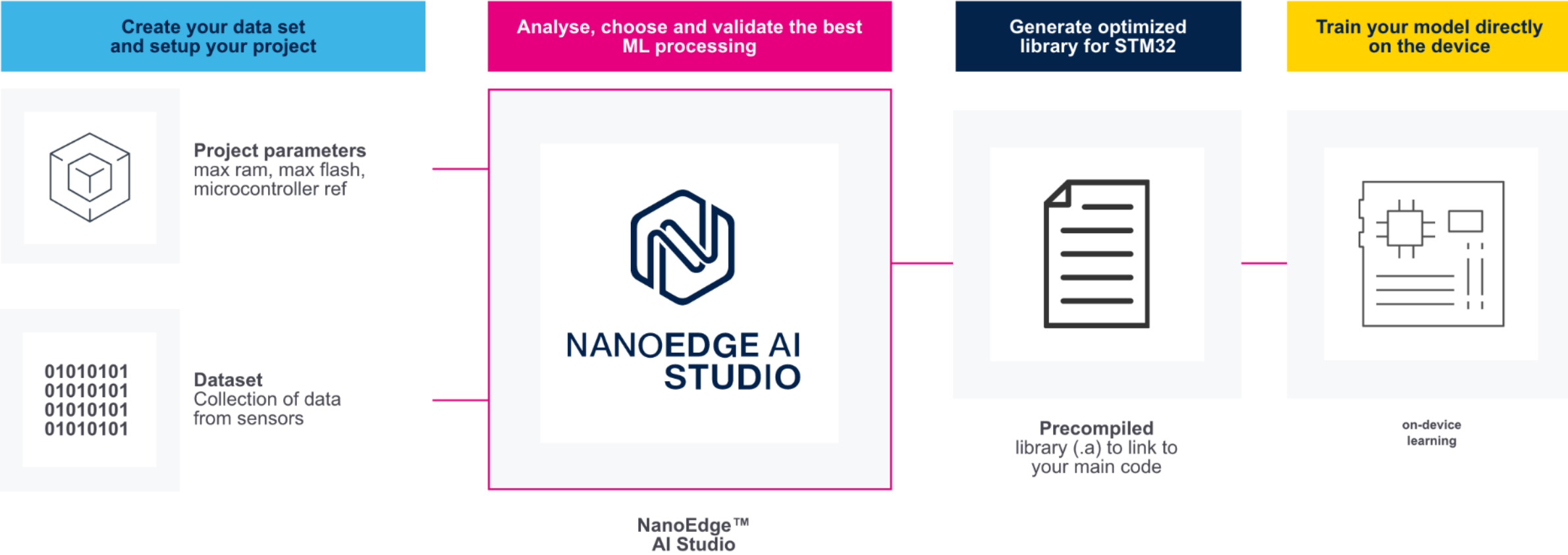


NanoEdge AI Studio improves the performance of the model..

.. And then optimizes it to reduce footprint and latency



# NanoEdge AI Studio workflow



# After-market predictive maintenance intelligent sensor with wireless connectivity



**Predictive  
Maintenance**

**Multi-sensors and  
learning on device**



Oxytronic  
Deployed at **Volvo Trucks**  
manufacturing plant

Washing machine uses **advanced motor control algorithms** to weigh clothes and optimize water, detergent, and energy used



**~15-40%**  
Energy saving per  
washing cycle

Leader in white-goods  
**Millions** units annual  
starting in 2024



# Adding tire pressure capability to e-bike without any new hardware
































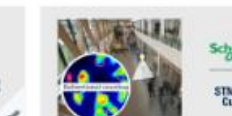




## Virtual sensor

Tire pressure measured  
through the e-motor  
current consumption

Asian e-bike maker  
**Hundreds thousands** units  
annual starting in 2024



# Explore multiple use cases on [stm32ai.st.com](http://stm32ai.st.com)

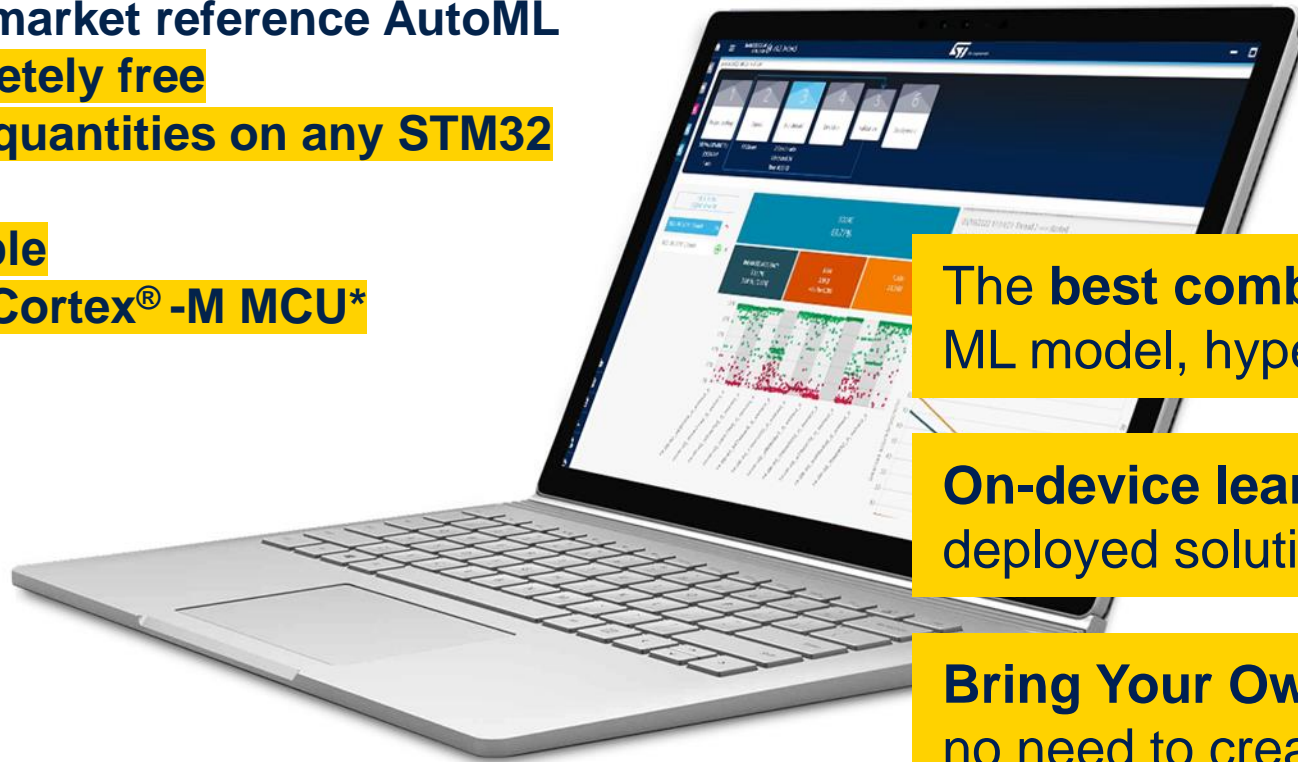
 <p><b>SMART CITY   IDEA</b> <b>Electrical fault detection and classification</b> Detect and classify electrical anomalies in a power system. <a href="#">More details</a></p>	 <p><b>WEARABLES   IDEA</b> <b>Human activity recognition using a smartphone</b> Using a smartphone to recognize human activities. <a href="#">More details</a></p>	 <p><b>INDUSTRIAL   IDEA</b> <b>Pump anomaly detection based on vibrations</b> Using AI to extrapolate torque and rotor temperature values to improve motor performance. <a href="#">More details</a></p>	 <p><b>INDUSTRIAL   DEMO</b> <b>Pump anomaly detection based on vibrations</b> Learn to detect abnormal behavior at the edge on a vibrating machine. <a href="#">More details</a></p>	 <p><b>SMART HOME   IDEA</b> <b>Floor type detection for vacuum cleaners</b> Advanced solution for material recognition of floor type (hard or soft) enabled by AI technology. <a href="#">More details</a></p>	 <p><b>INDUSTRIAL   DEMO</b> <b>Aftermarket wireless digit reader</b> Equip meters with aftermarket wireless &amp; low-power readers. <a href="#">More details</a></p>	 <p><b>INDUSTRIAL   CUSTOMER</b> <b>AI solution for industrial predictive maintenance with Oxytronic</b> Predictive maintenance solution for industrial equipment. <a href="#">More details</a></p>	 <p><b>INDUSTRIAL   IDEA</b> <b>Electric drive diagnosis (11 faults)</b> Classify data based on different types of faults in an electric drive. <a href="#">More details</a></p>	 <p><b>TRANSPORTATION   DEMO</b> <b>People counting embedded in a camera</b> People detection and counting on high-performance MCU. <a href="#">More details</a></p>
 <p><b>SMART CITY   IDEA</b> <b>Ensuring enhanced stability of an electrical grid</b> Using AI to determine if an electrical grid is stable. <a href="#">More details</a></p>	 <p><b>INDUSTRIAL   DEMO</b> <b>Accurately measure the weight of clothes inside a washing machine</b> Using AI to make your home appliances "smarter" and more energy efficient for a sustainable future. <a href="#">More details</a></p>	 <p><b>INDUSTRIAL   DEMO</b> <b>Motor fault detection and classification</b> Detection and classification of motor faults for predictive maintenance. <a href="#">More details</a></p>	 <p><b>INDUSTRIAL   DEMO</b> <b>Fan anomaly detection based on vibrations</b> Learn to detect abnormal behavior at the edge on a vibrating machine. <a href="#">More details</a></p>	 <p><b>SMART CITY   IDEA</b> <b>People detection and counting solution</b> Optimized computer vision using an MPU running at 8 FPS. <a href="#">More details</a></p>	 <p><b>INDUSTRIAL   CUSTOMER</b> <b>AI solution for industrial predictive maintenance with NKE Watteco</b> Predictive maintenance solution for industrial equipment. <a href="#">More details</a></p>	 <p><b>SMART BUILDING   IDEA</b> <b>People counting with a ranging sensor</b> Count the number of people passing through a door using a Time-of-Flight sensor. <a href="#">More details</a></p>	 <p><b>INDUSTRIAL   CUSTOMER</b> <b>AI solution for reflow oven monitoring with Lacroix Electronics</b> Predictive maintenance applied to industrial ovens. <a href="#">More details</a></p>	 <p><b>WEARABLES   DEMO</b> <b>Human Activity Recognition</b> Easily identify 5 different activities with a 3D accelerometer. <a href="#">More details</a></p>
 <p><b>INDUSTRIAL   IDEA</b> <b>Monitoring quality in a food production line</b> Creates an AI model that predicts the quality of processed food instead of measuring it. <a href="#">More details</a></p>	 <p><b>SMART HOME   DEMO</b> <b>Handwriting recognition</b> Handwriting recognition on ultra-low-power MCU. <a href="#">More details</a></p>	 <p><b>SMART BUILDING   DEMO</b> <b>Hand posture recognition without camera module</b> Hand posture recognition running on STM32F401 based on ST multi-zone Time-of-Flight ranging sensor. <a href="#">More details</a></p>	 <p><b>ENTERTAINMENT   IDEA</b> <b>Detecting the "let" in a table tennis game</b> Classification of the net vibrations with an accelerometer. <a href="#">More details</a></p>	 <p><b>INDUSTRIAL   DEMO</b> <b>Fan anomaly classification based on ultrasound analysis</b> Neural Network classification based on a high-frequency analog microphone pipeline. <a href="#">More details</a></p>	 <p><b>TRANSPORTATION   CUSTOMER</b> <b>AI solution for monitoring automatic doors with Crouzet</b> Predictive maintenance on motors for automatic door motors. <a href="#">More details</a></p>	 <p><b>TRANSPORTATION   IDEA</b> <b>Gearbox fault detection using vibrations</b> Vibration analysis to detect an abnormal behavior on a gearbox. <a href="#">More details</a></p>	 <p><b>INDUSTRIAL   DEMO</b> <b>People presence detection (visual wake word)</b> Human detection on high-performance MCU. <a href="#">More details</a></p>	 <p><b>INDUSTRIAL   DEMO</b> <b>Transfer learning applied to forest fire detection</b> Image classification on high-perf MCU. <a href="#">More details</a></p>
 <p><b>ENVIRONMENT   DEMO</b> <b>Plant leaf disease identification</b> Image classification on high-performance MCU. MobileNet v2_025 model from STM32 model zoo. <a href="#">More details</a></p>	 <p><b>ENVIRONMENT   DEMO</b> <b>Transfer learning applied to flower recognition</b> Image classification on high-performance MCU. MobileNet v2_alpha 0.35 model from STM32 model zoo. <a href="#">More details</a></p>	 <p><b>INDUSTRIAL   DEMO</b> <b>Anomaly detection in a motor running at different speeds</b> Smart sensor node over BLE connectivity to simplify the configuration and to be notified in case of detection via a mobile app. <a href="#">More details</a></p>	 <p><b>INDUSTRIAL   DEMO</b> <b>Anomaly detection in an electric motor</b> Current sensing to detect abnormal behaviors in motors. <a href="#">More details</a></p>	 <p><b>SMART OFFICE   CUSTOMER</b> <b>People flow counting Sensor with Schneider Electric</b> An innovative approach to measure people flows using an in-house thermal sensor. <a href="#">More details</a></p>	 <p><b>APPLIANCES   IDEA</b> <b>Shifumi gesture recognition</b> Trigger actions on a PC using a Time-of-Flight sensor to classify hand movements. Recognition of 3 different classes. <a href="#">More details</a></p>	 <p><b>SMART HOME   DEMO</b> <b>Food recognition</b> Image classification on high-performance MCU. <a href="#">More details</a></p>	 <p><b>TRANSPORTATION   CUSTOMER</b> <b>Railway monitoring with VapeRail</b> On-track predictive maintenance. <a href="#">More details</a></p>	 <p><b>ENTERTAINMENT   IDEA</b> <b>Ukulele chords classification</b> Analysis of the vibrations produced by the instrument to detect the chord played. <a href="#">More details</a></p>



# Simplify AI development workflow

Deployment of NanoEdge AI Studio libraries, the market reference AutoML tool, **is completely free for unlimited quantities on any STM32**

**... and available on any Arm® Cortex® -M MCU\***



NANOEDGE AI  
STUDIO 

The **best combination** for given data:  
ML model, hyperparameters and preprocessing

**On-device learning capability** to fine-tune  
deployed solution without retraining

**Bring Your Own Data** approach:  
no need to create edge AI models

*\*under a special license agreement*

Try NanoEdge AI Studio  
for free today



# Our technology starts with You



Find out more at [stm32ai.st.com](https://stm32ai.st.com)

© STMicroelectronics - All rights reserved.

ST logo is a trademark or a registered trademark of STMicroelectronics International NV or its affiliates in the EU and/or other countries.

For additional information about ST trademarks, please refer to [www.st.com/trademarks](https://www.st.com/trademarks).

All other product or service names are the property of their respective owners.



life.augmented



# Copyright Notice

This multimedia file is copyright © 2024 by tinyML Foundation. All rights reserved. It may not be duplicated or distributed in any form without prior written approval.

tinyML<sup>®</sup> is a registered trademark of the tinyML Foundation.

[www.tinyml.org](http://www.tinyml.org)



# 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](http://www.tinyml.org)**