



# ST AI Ecosystem

Ryk LIU

AI Competence Center

STMicroelectronics

# The explosion of AI-enabled devices is accelerating the inference shift from the cloud to the tiny edge

Inference

Cloud-centric



On device-centric

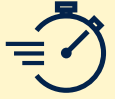


Tiny edge-centric



Enabled by a different class of hardware and software

# Inferring at the edge brings substantial benefits



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

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



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



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



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

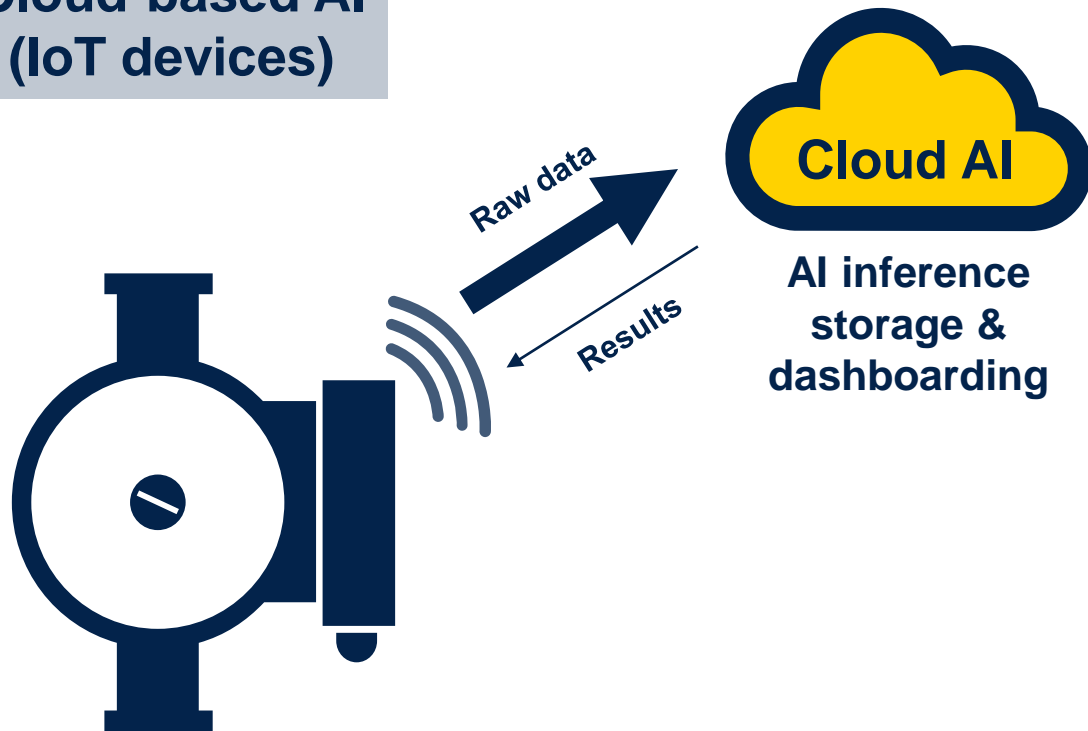
## Tiny edge-centric



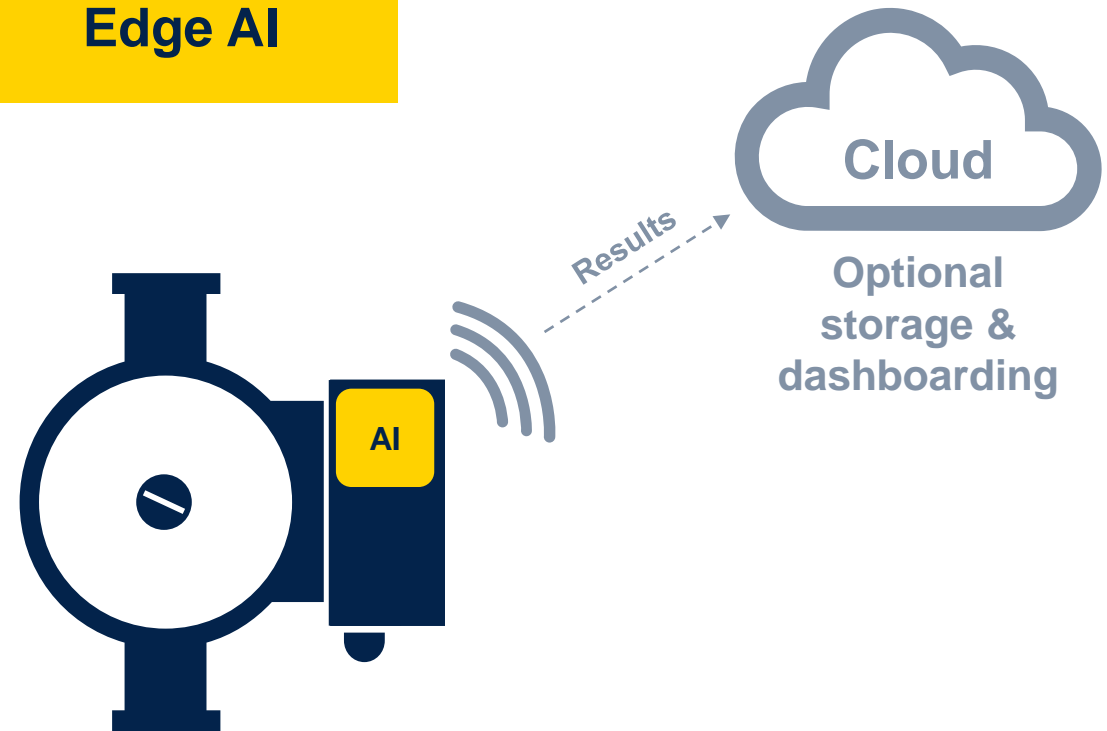
# The next leap for intelligent products

Technology is moving beyond simple connectivity to on-device autonomy, where products can infer locally

Cloud-based AI  
(IoT devices)



Edge AI



# Opening a new range of embedded AI applications

## Upgrade existing devices with AI-based services



Arc Fault  
Detection



Predictive  
Maintenance



Battery  
Management

## Reduce BoM by shifting from MPUs to MCUs



People  
Detection



Sound  
Analysis



Speech  
Recognition

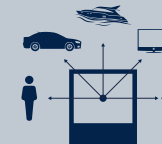
## Offer advanced services without cloud costs



Object  
Classification



Pose  
Estimation



Object  
Segmentation

A few KOPS

Tens of TOPS

# Adding tire pressure detection capability to e-bikes without adding any new hardware



## Virtual sensor

Tire pressure measured  
through the e-motor  
current consumption

Panasonic e-bike  
**Hundreds of thousands**  
of units annually,  
starting in 2024

A 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  
Production starting in 2024  
for **millions** units



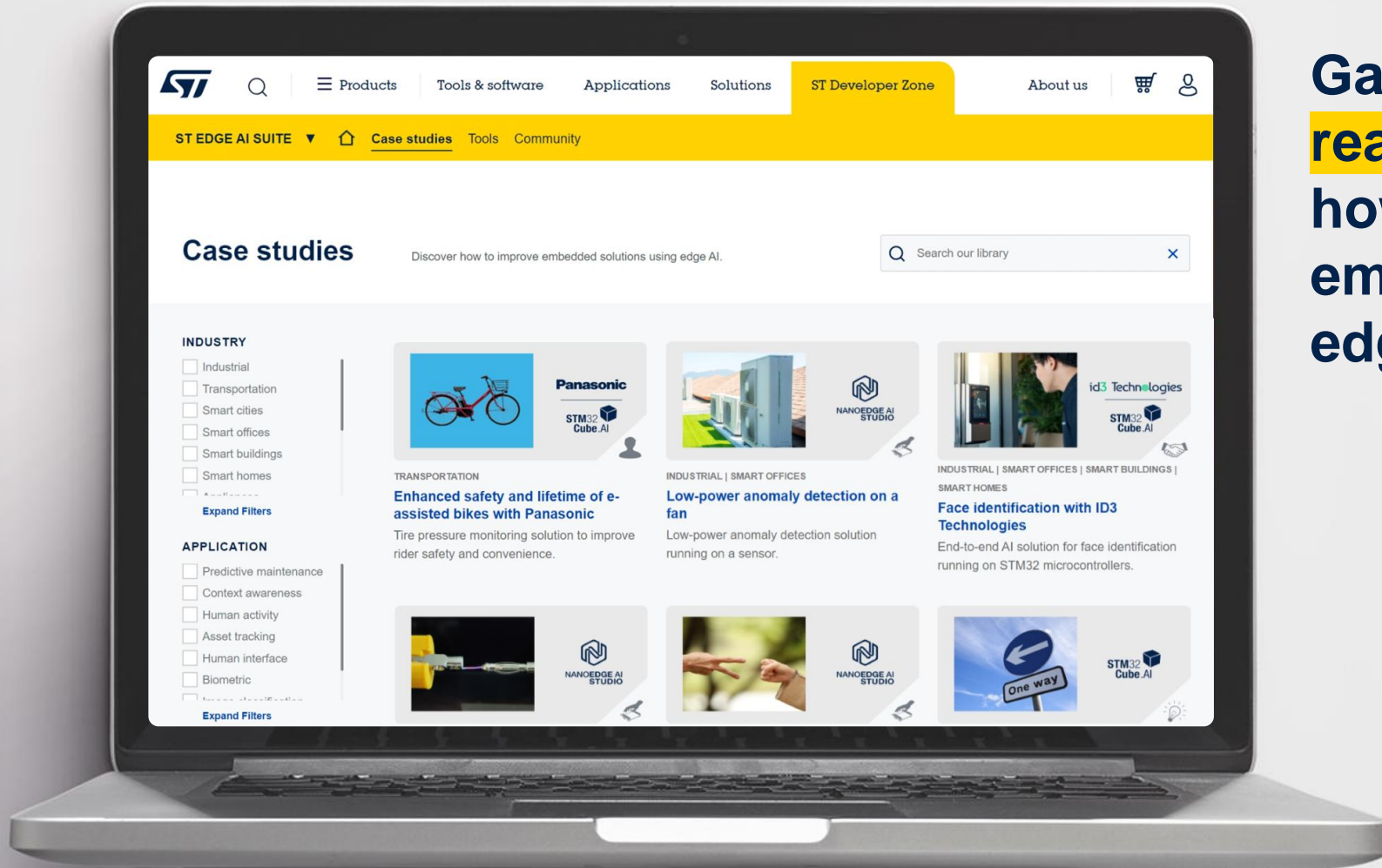
# Optimize laptop power policies by avoiding overheating and battery drain

**Smart context  
detection**

**In/out bag - on table**



# Explore inspiring case studies

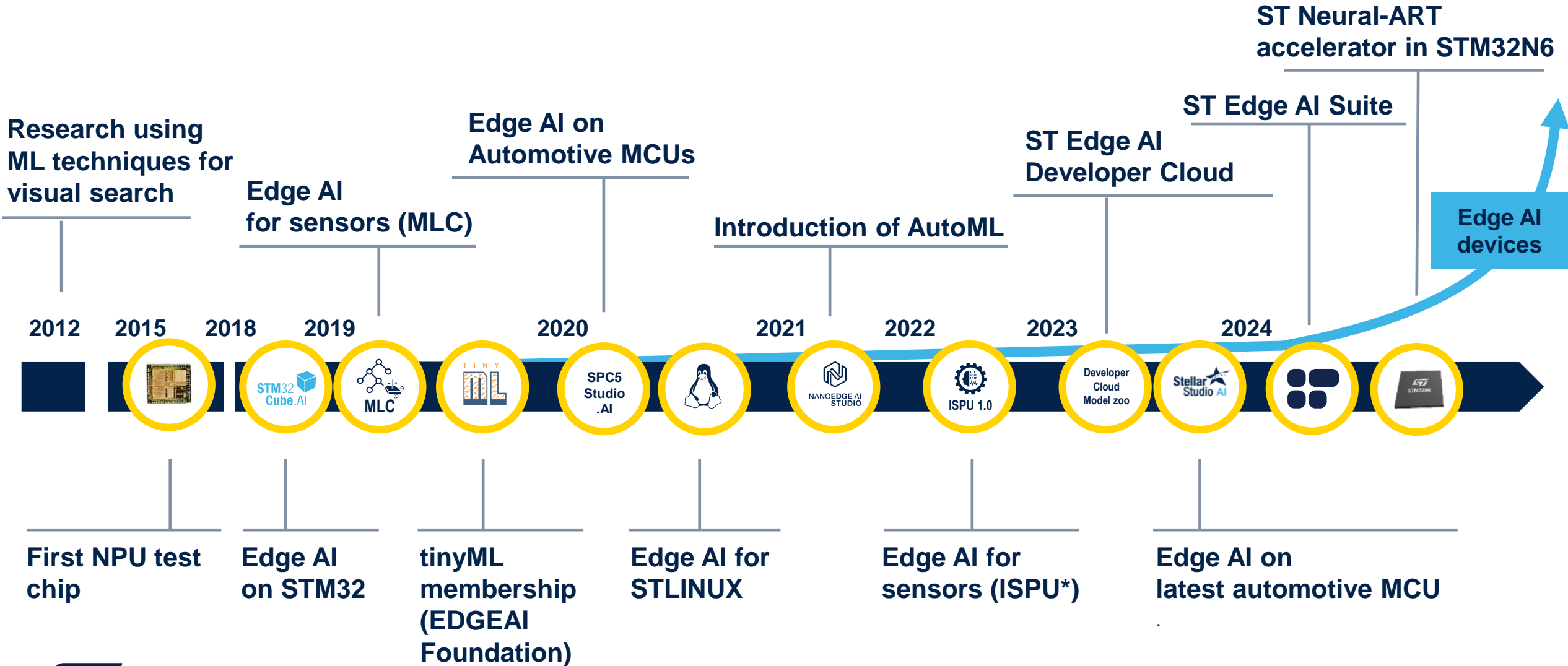


Gain **valuable insights from real-world case studies** on how to improve your embedded solutions using edge AI.



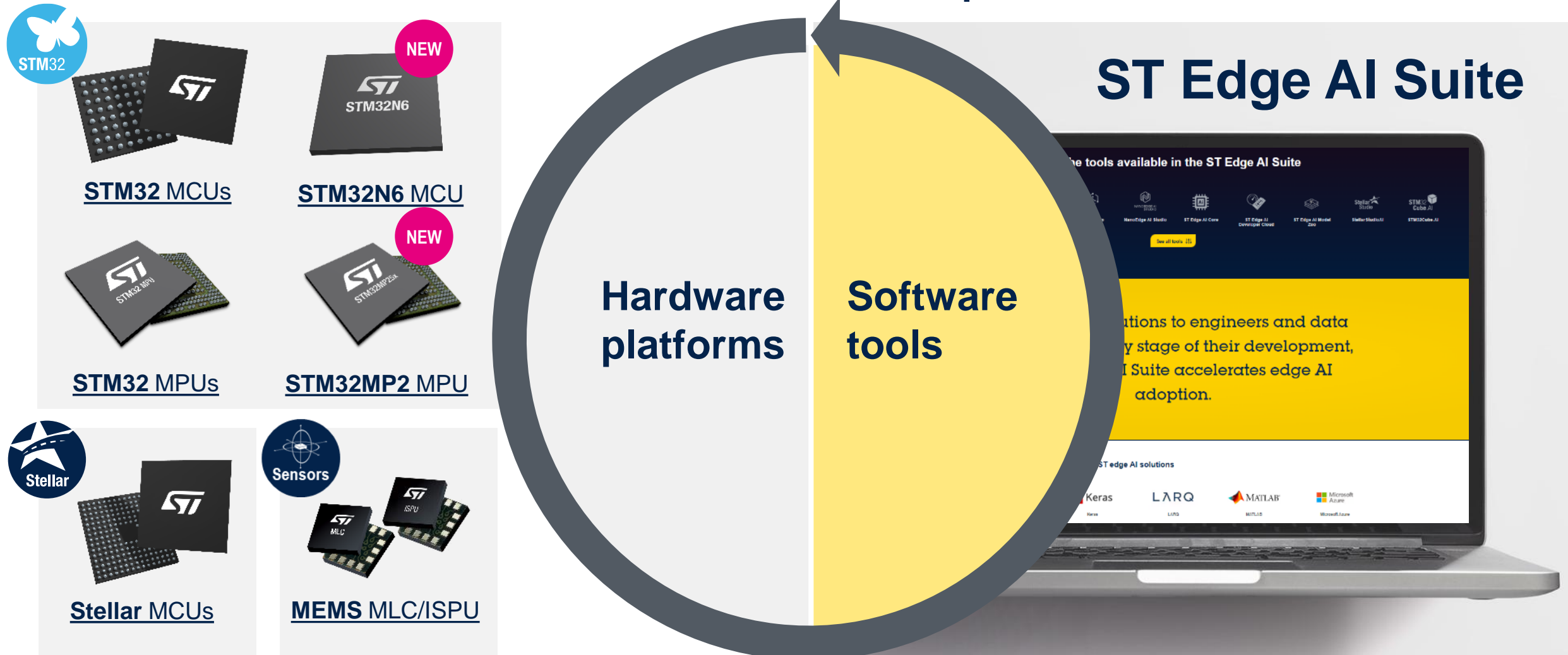
[Edge AI case studies](#)

# 10+ years of research, development, and deployment



\*Intelligent sensor processing unit

# A comprehensive approach to help developers accelerate their product transformation



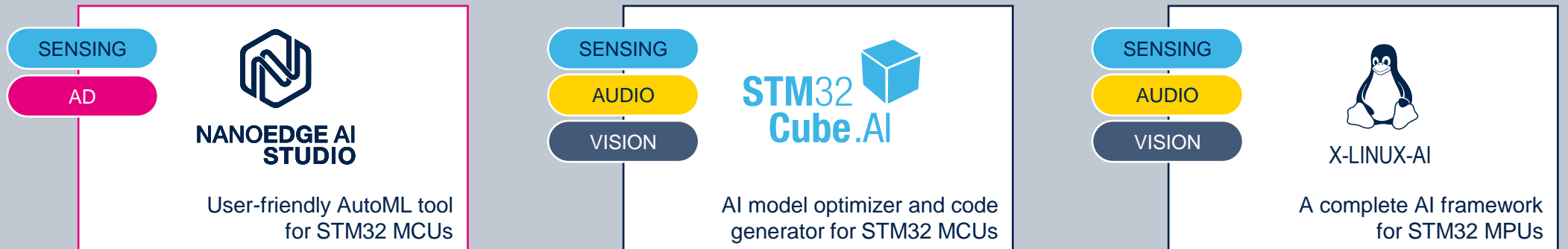
 Click on the products to find out more

# STM32 product offering

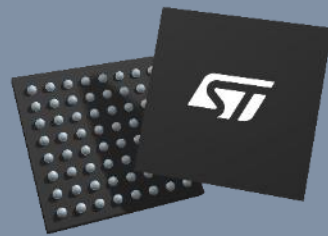
Enabling major edge AI technologies



Software tools for any user profile



Large choice of general purpose & accelerated hardware



STM32 MCUs



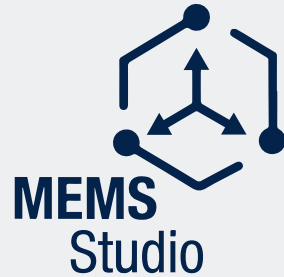
STM32N6 MCU with AI acceleration



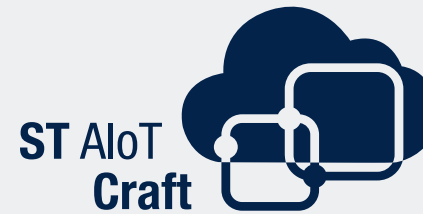
STM32MP1 & STM32MP2 MPUs

# Sensor for running edge AI

**Software** tools to develop **in-sensor AI** features



- Profile and optimize NN and ML models
- Configure the MEMS machine learning core and ISPU



Create end-to-end sensor-to-cloud solutions



A broad range of **evaluation** & **prototyping** tools



**20+** smart sensors released



6-axis IMUs with ISPU



6-axis IMUs with high-g axl



6-axis IMUs for industrial



6-axis IMUs for automotive



3-axis smart accelerometers



Industrial accelerometers



Industrial inclinometers

# Software ecosystem

Jump-start your project

## with an STM32 MCU

Follow these few simple steps to get started.

Start a project

Microcontrollers, boards and hardware tools, software tools and embedded software packages, and so much

MCU.

**3,300+**  
part numbers.

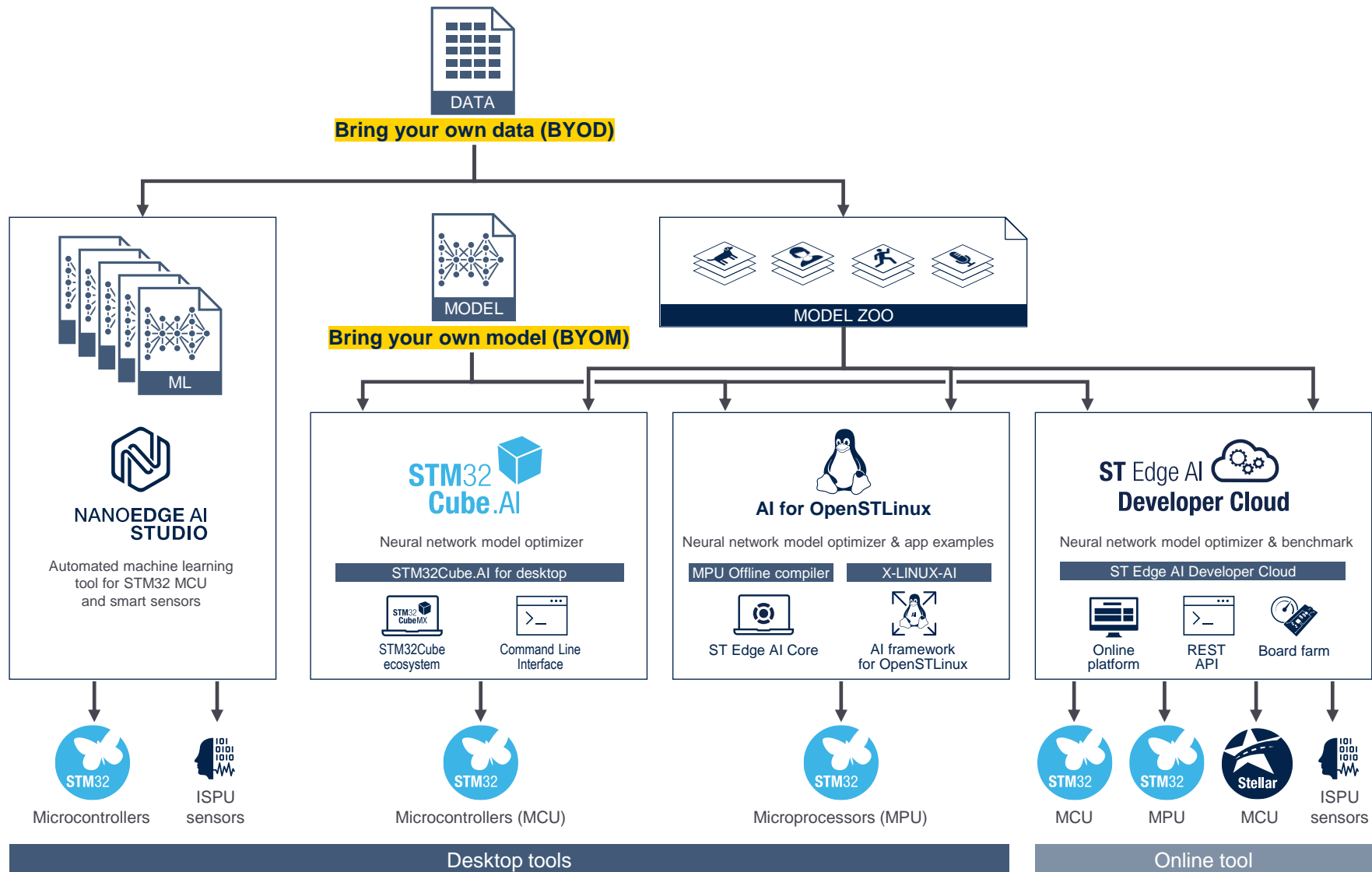
Tools.

**500+**  
software & boards.

Community.

**135,000+**  
members.

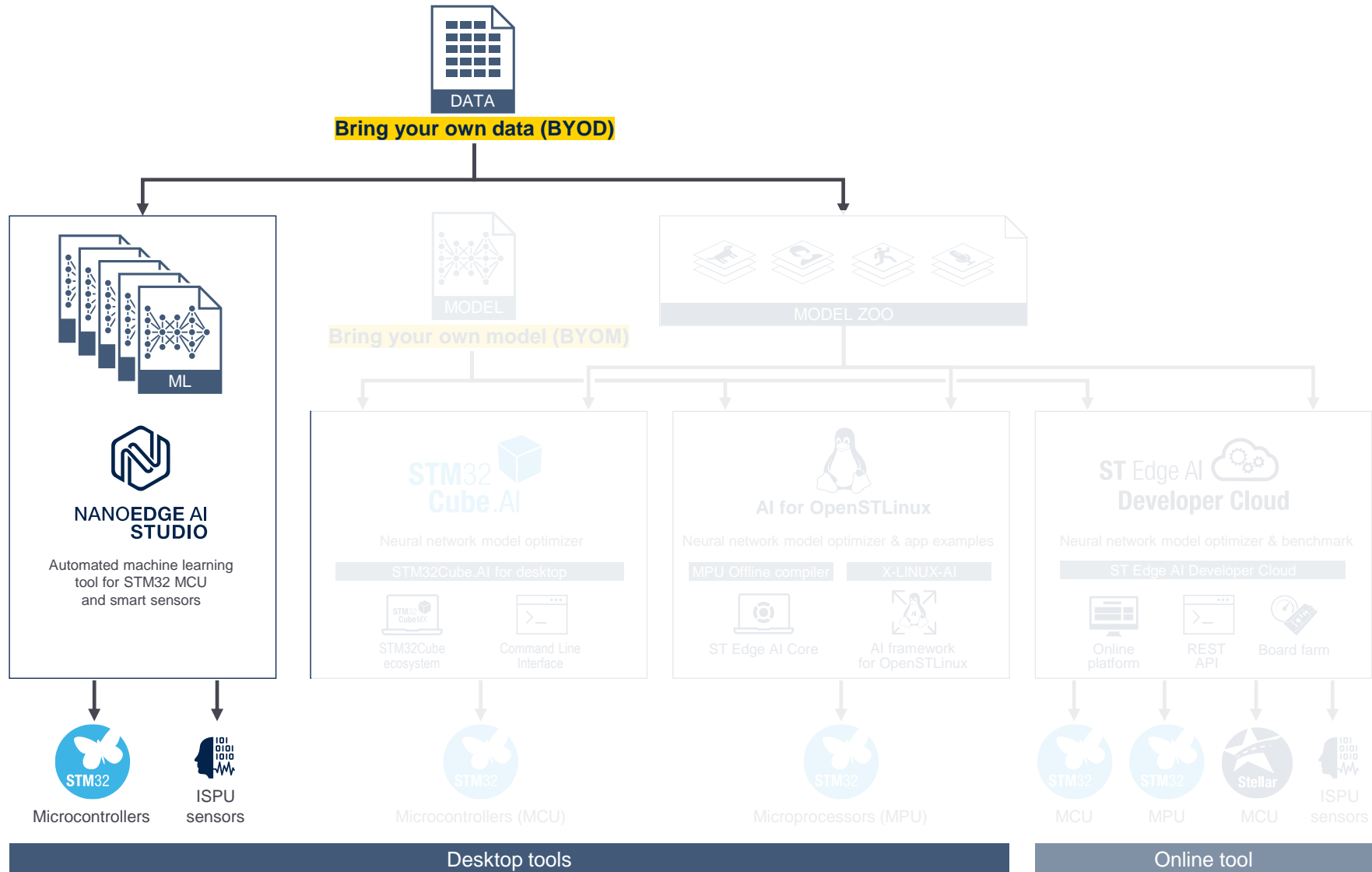
# A broad collection of free tools



# NanoEdge AI Studio, the AutoML tool



# Software tool: NanoEdge AI Studio



# Three machine learning algorithm families to make your product smarter

NanoEdge AI Studio AutoML projects include three types of projects

Anomaly detection →



**I want to anticipate product failures**

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

Classification →



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

Smart watches **classify** human activity recognition without using connectivity features

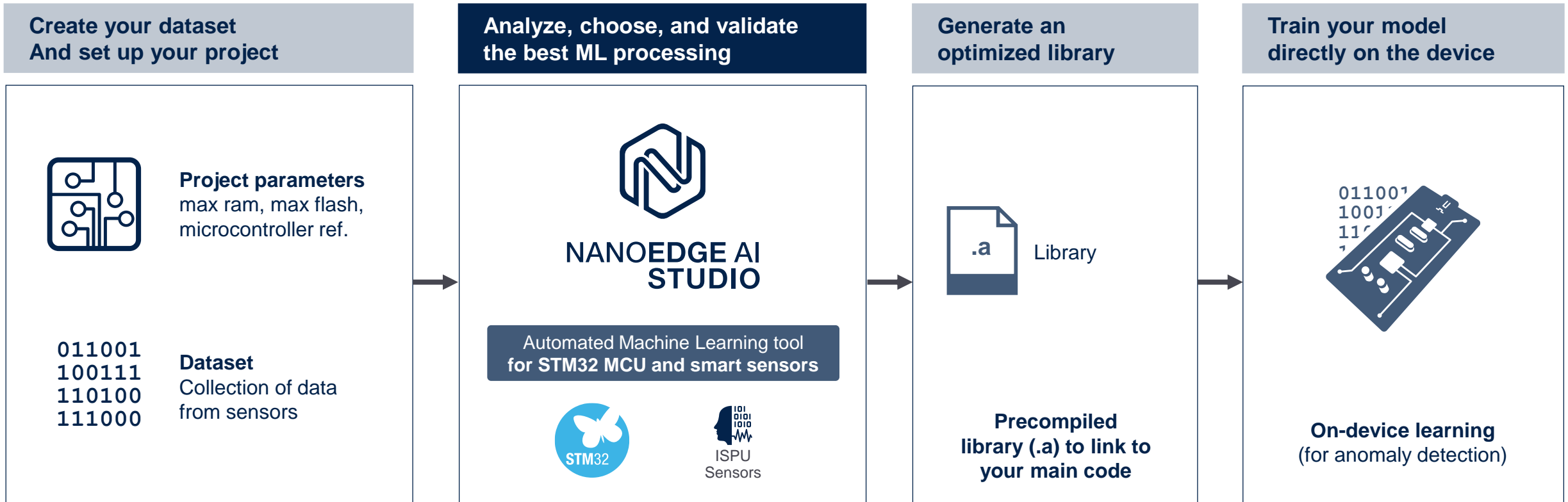
Extrapolation →



**I need to predict future states**

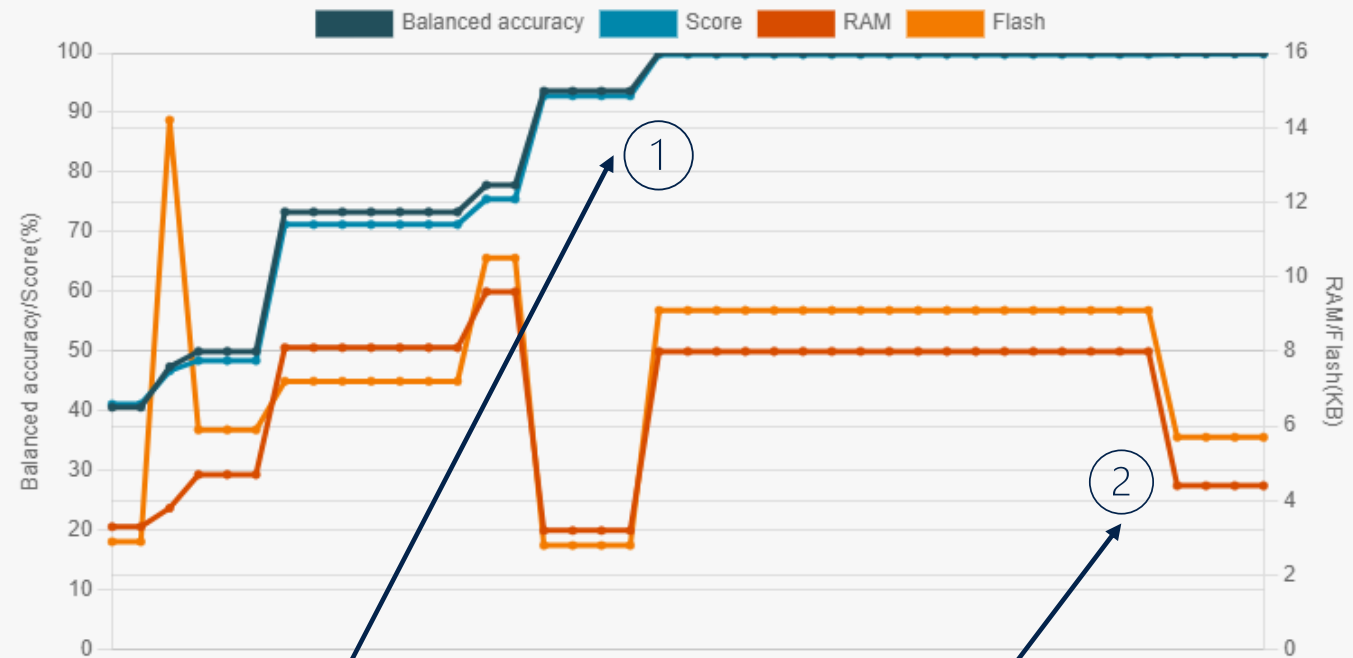
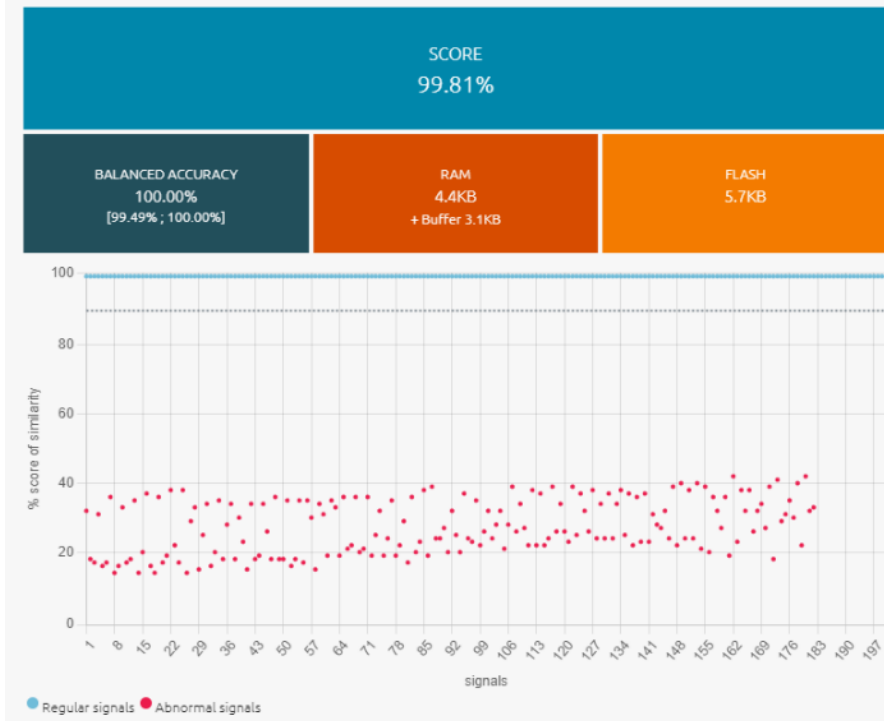
A **virtual sensor** for tire pressure in an eBike estimates pressure based on motor current measurements

# NanoEdge AI Studio workflow



# 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 optimizes it to reduce footprint and latency



# STM32Cube.AI & ST Edge AI Developer Cloud

The screenshot displays the ST Edge AI Developer Cloud interface. At the top, there are five tabs: Optimize, Quantize, Benchmark, Results, and Generate. The 'Optimize' tab is active. The interface shows the following details:

- Model currently selected:** MOBILENET\_V1\_025\_H5
- INPUT:** float [128x128x3]
- OUTPUT:** float [1x1x1x8]
- MODEL TYPE:** ResNet
- MACC:** 13871986

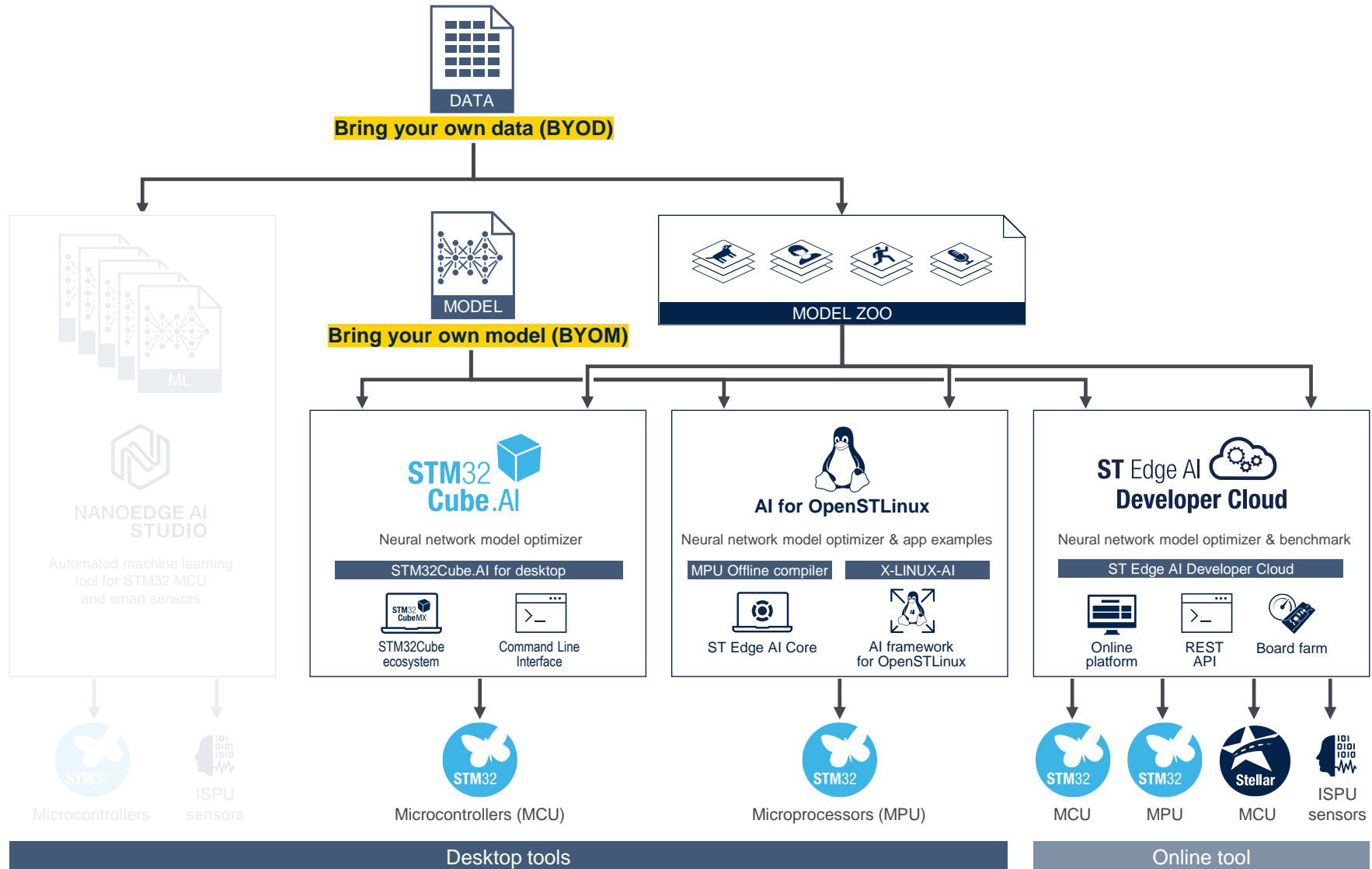
Below this, there is a section for selecting model optimization options:

- Balance between RAM size and inference time (-optimization balanced)
- Optimize for RAM size (-optimization ram)
- Optimize for inference time (-optimization time)
- Use activation buffer for input buffer (-allocate-inputs)
- Use activation buffer for output buffer (-allocate-outputs)

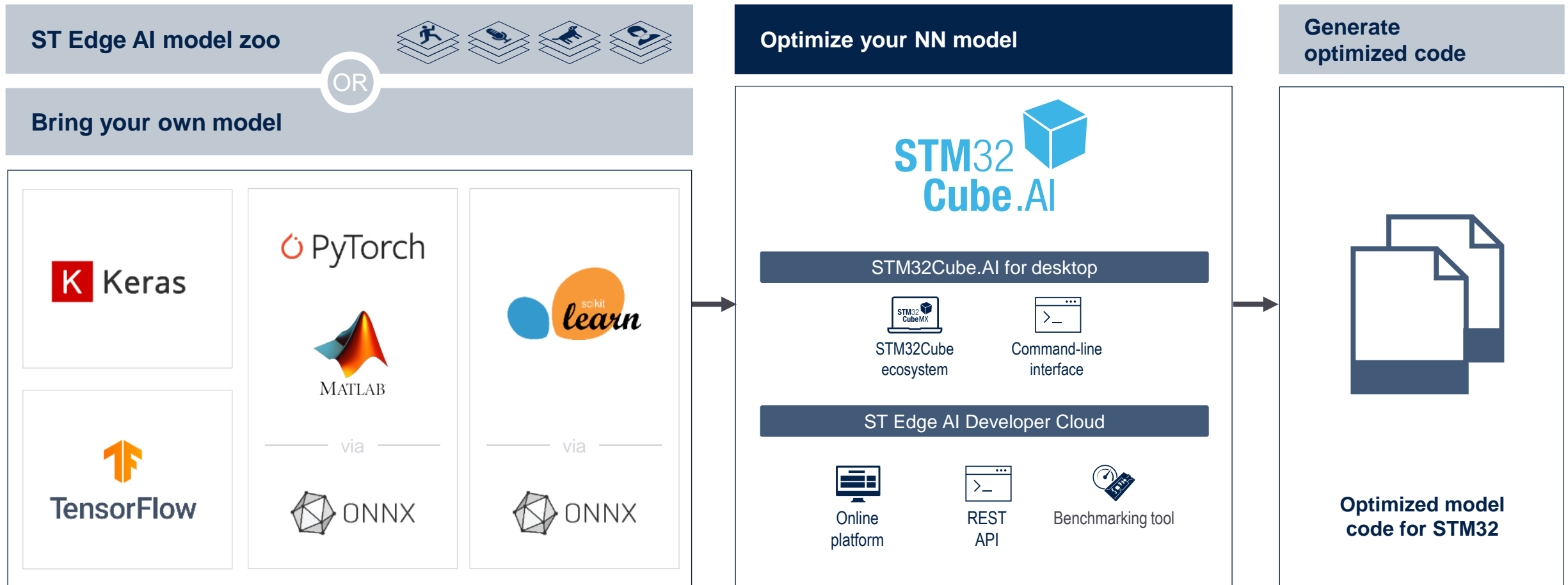
A 'Optimize' button is located to the right of these options. Below the optimization options is a 'History of optimization results' table. The table has columns for Date, Optimization, Allocate Input, Allocate Output, MACC, Peak size, and RAM size. The current row shows the following data:

Date	Optimization	Allocate Input	Allocate Output	MACC	Peak size	RAM size
1/19/23, 5:22 PM	Level Default	balanced	true	true	13871986	641943
					Total: 859 KiB (+0.00%)	617124
					Total: 949 KiB (+0.00%)	617124

# Software tool: STM32Cube.AI



# One tool – two versions to deploy AI on STM32



# Start with edge AI optimized models

## STM32 model zoo

A collection of application-oriented models optimized for STM32

Human activity



Motion Sensing

Image classification



Computer vision

Audio event detection



Audio classification

Object detection



Computer vision

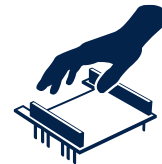


Hosted on GitHub



**Model training scripts**

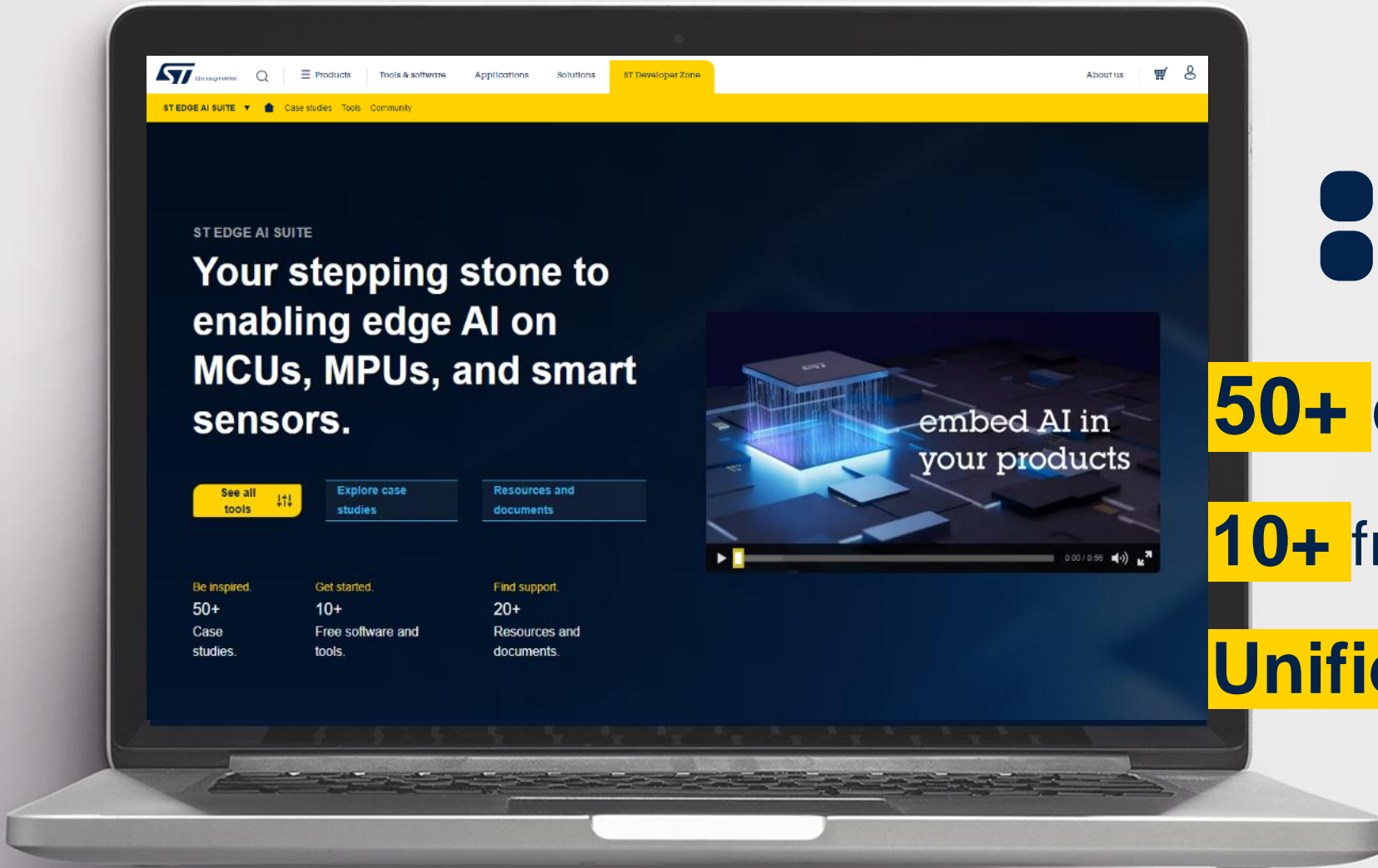
- Scripts to generate and validate



**Application code example**

- Designed to host optimized NN models
- Automatically generated from the trained models
- Easy to deploy for end-to-end evaluation

By bringing solutions to **engineers and data scientists** at **every stage of their development**, the ST Edge AI Suite accelerates edge AI adoption.



 **ST Edge AI Suite**

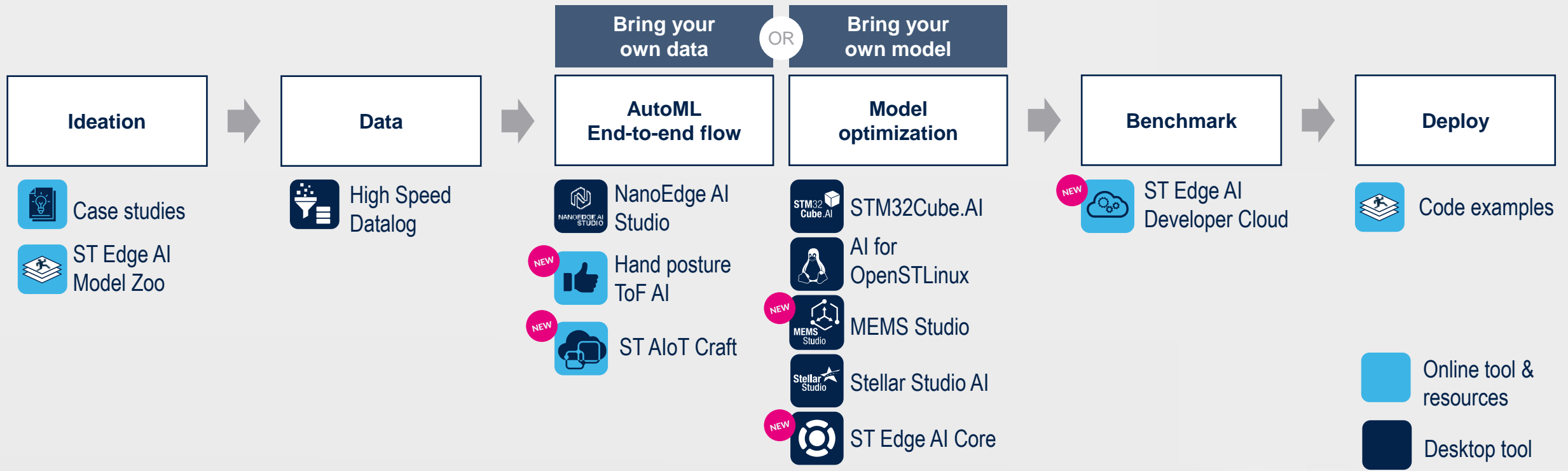
**50+** case studies

**10+** free software tools

**Unified** AI core technology

# Free tools to run edge AI on MCUs, MPUs, and smart sensors

Find the tools you need to optimize and deploy machine learning algorithms, from data collection to final deployment on hardware.



# Our technology starts with You



Find out more at [www.st.com](http://www.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](http://www.st.com/trademarks).

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

