



life.augmented

Enhance your products using edge AI solutions on STM32



A short introduction to AI concepts



What is artificial intelligence

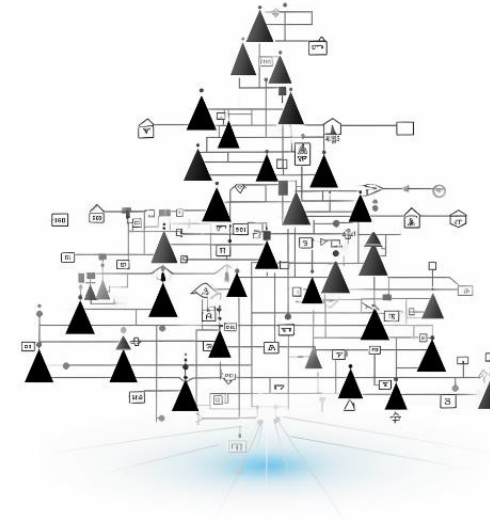
Embedded artificial intelligence

The ability of **machines** to perform tasks that typically require **human intelligence**

From a technical perspective, AI is another way to program a machine.



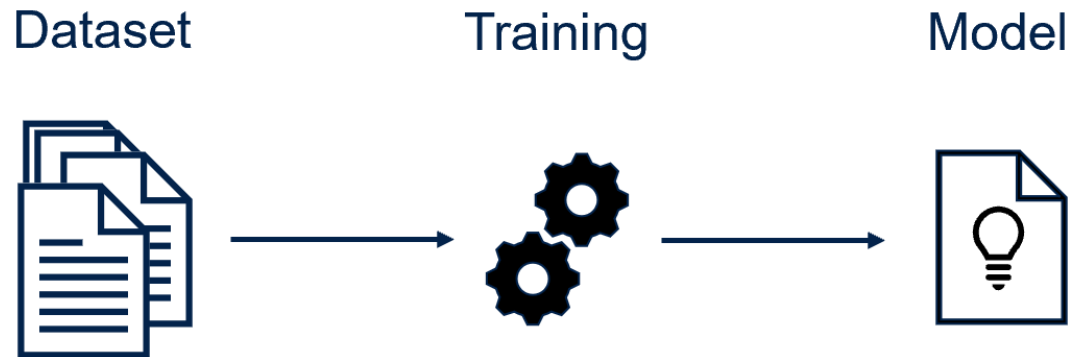
Perception
Reasoning
Learning
Decision-making



Using AI algorithms implies two phases

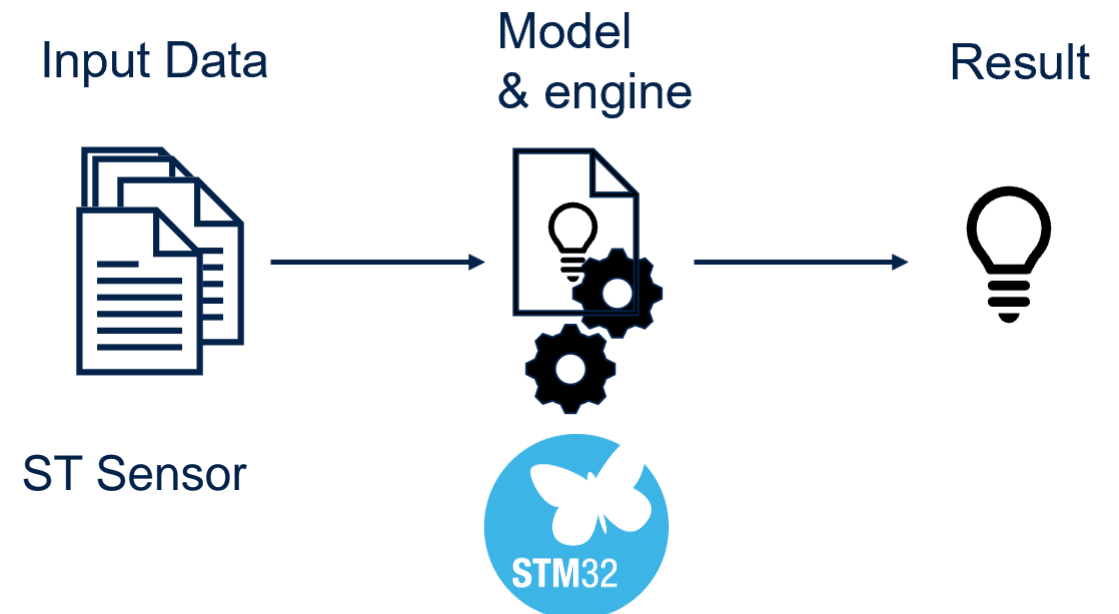
Training

Using **datasets** (like examples) and a training engine/framework we are producing a **model**.

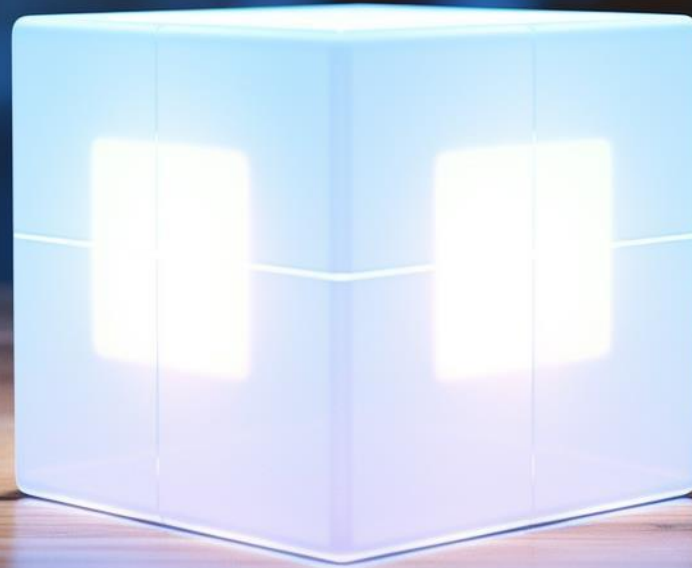


Inference

Then during the second phase, we can use this **model**. This is the inference phase.



Embedding intelligence at the edge



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



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

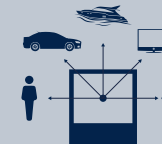
Offer advanced services without cloud costs



Object
Classification



Pose
Estimation



Object
Segmentation

A few KOPS

Tens of TOPS



life.augmented

Edge AI application examples



AI addresses a wide variety of projects

Smart city

Energy optimization
Traffic management
Public safety
Public transportation

Smart buildings

Energy optimization
Access control
Safety
Predictive maintenance

Smart home

Outdoor cameras
Babycams
Smart doorbells
Home appliances
Energy management
Lawn mowers
Vacuum robots

Energy

Solar panels
Breaker's control
EV chargers
Smart meters
Fraud detection

Industry 4.0

Predictive maintenance
Tools safety
Environment monitoring

Healthcare

Wearables
Patient monitoring
Fall detection
Predictive maintenance

Battery management

Arc fault detection

Face / object recognition

Anomaly detection

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

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

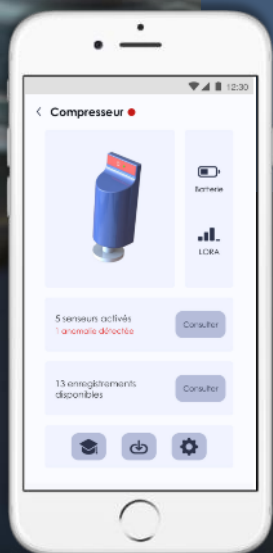


Virtual sensor

Tire pressure measured
through the e-motor
current consumption

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

After-market predictive maintenance intelligent sensor with wireless connectivity



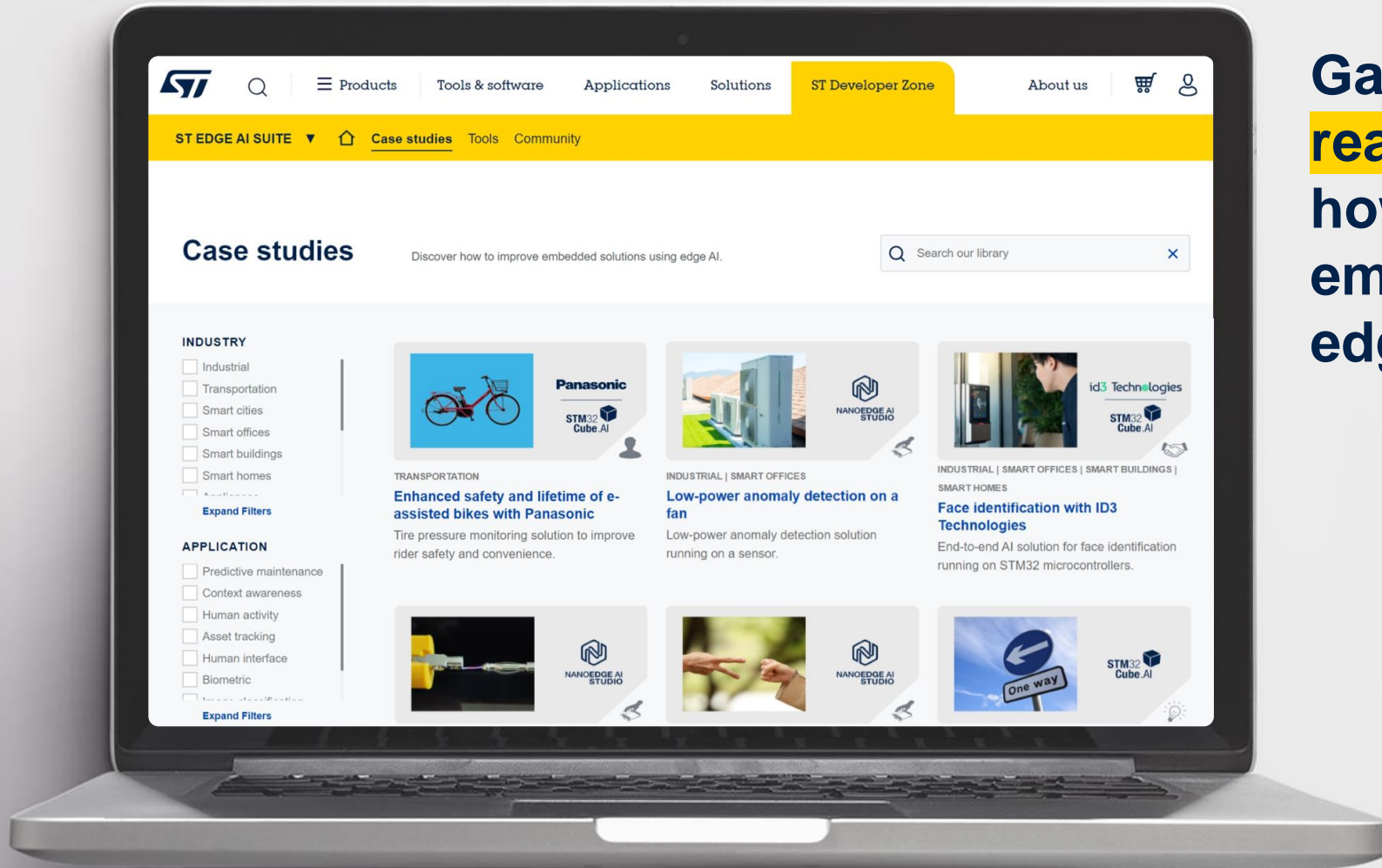
**Predictive
maintenance**

**Multisensors and
learning on device**



EMEA company
Deployed at **Volvo Trucks**
manufacturing plant

Explore inspiring case studies



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

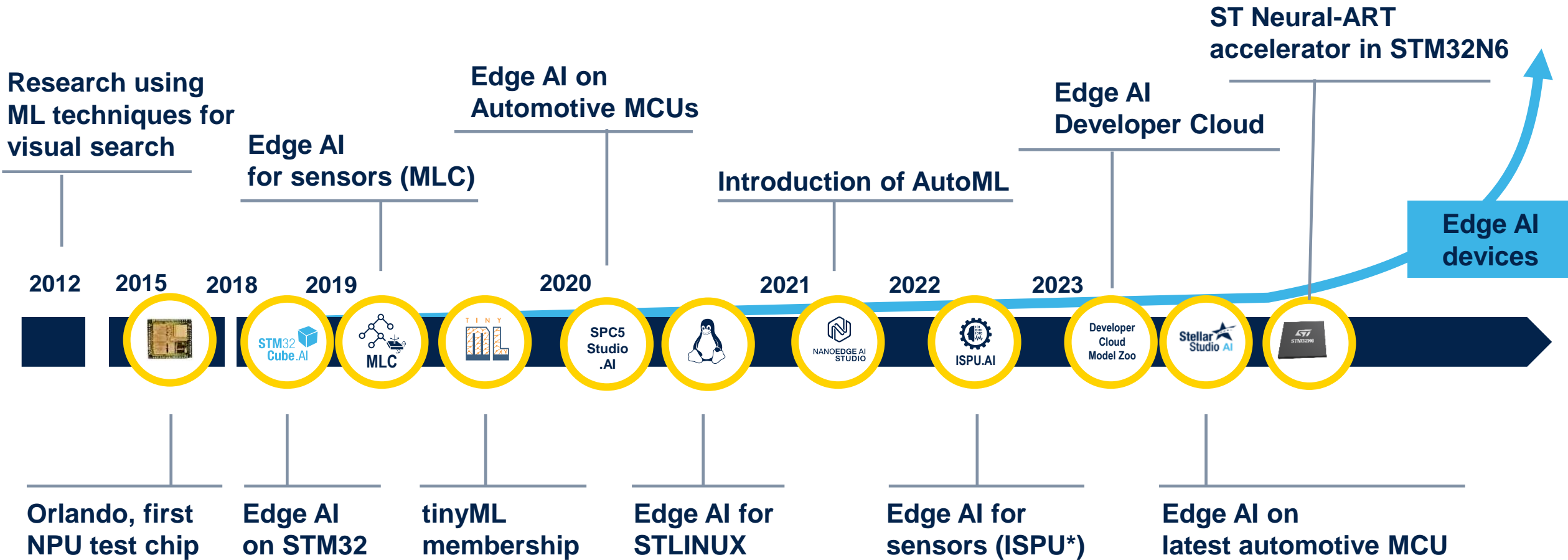


[Edge AI case studies](#)

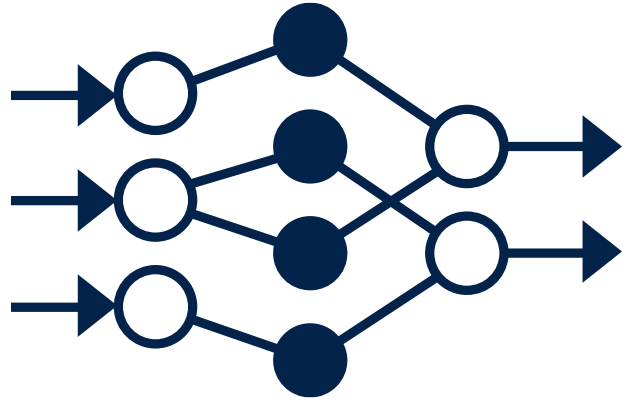
ST's investment in edge AI enabling broad adoption



More than a decade of research, development, and deployment



The challenge of deploying embedded AI



AI expertise

Data

Memory footprint

Inference time

Power consumption

eSW development

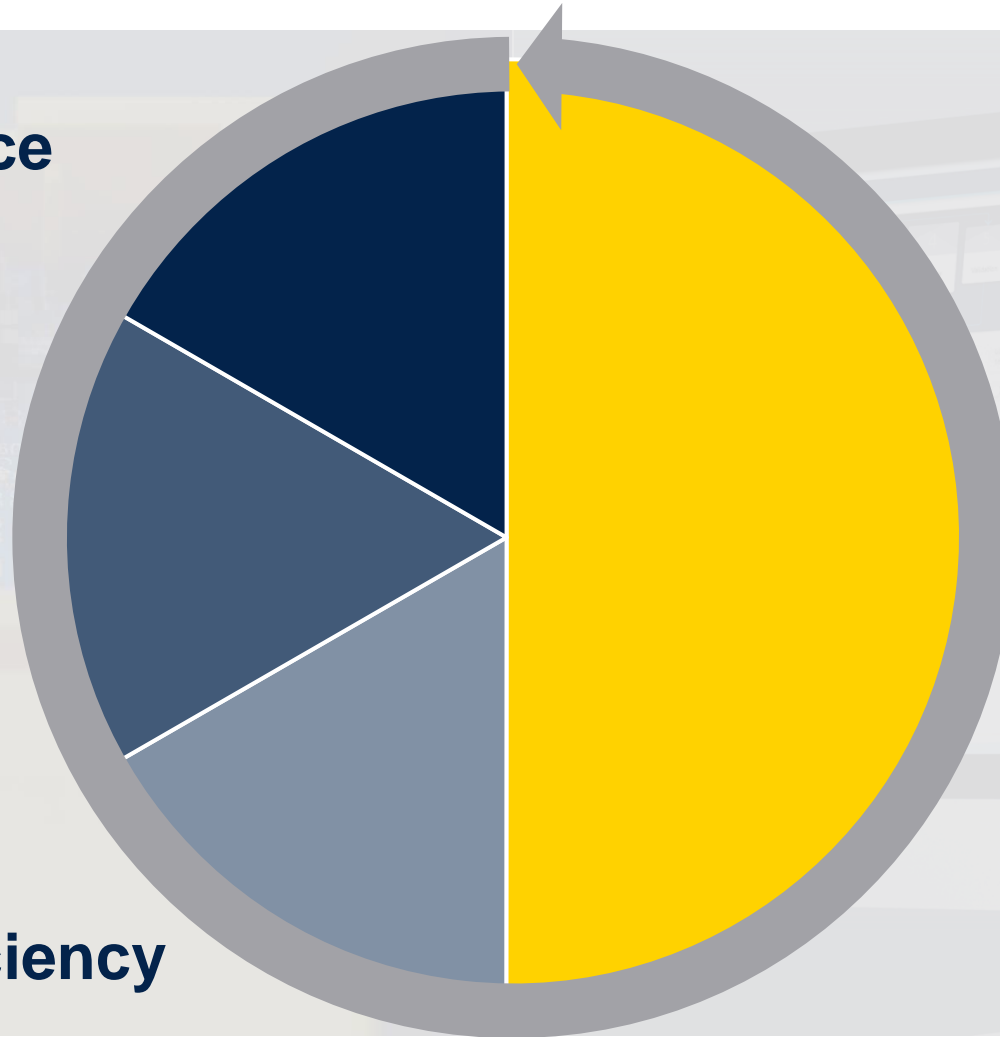
**Deploying embedding AI on microcontrollers
presents significant challenges**

Taking up the edge AI developer's challenge

Performance

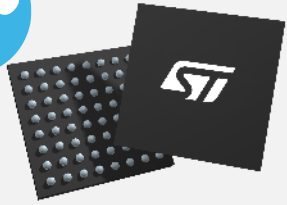
Security

Power efficiency



**Create, deploy
& maintain ML
models**

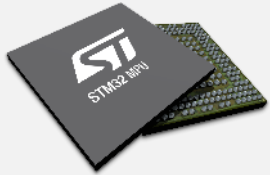
A comprehensive approach to help developers accelerate their product transformation



STM32 MCUs



STM32N6 MCU



STM32 MPUs



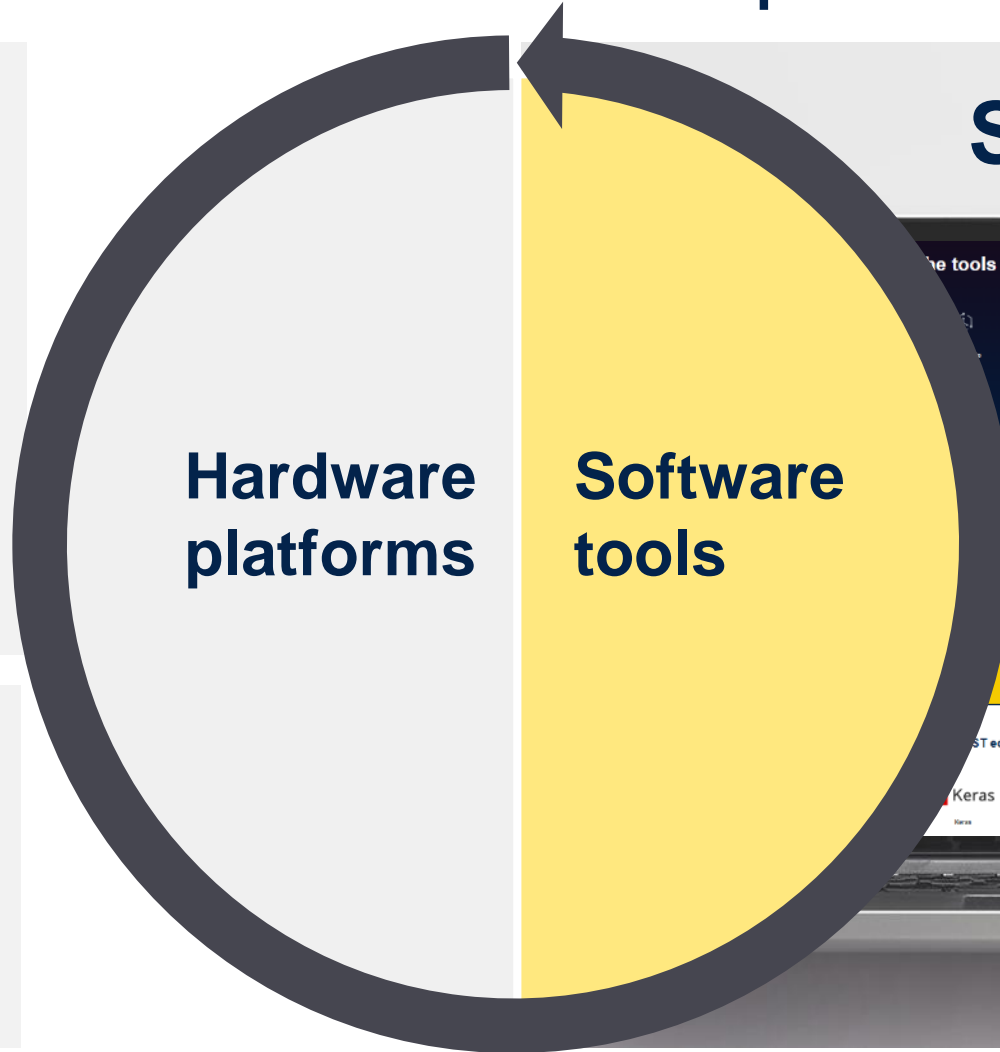
STM32MP2 MPU



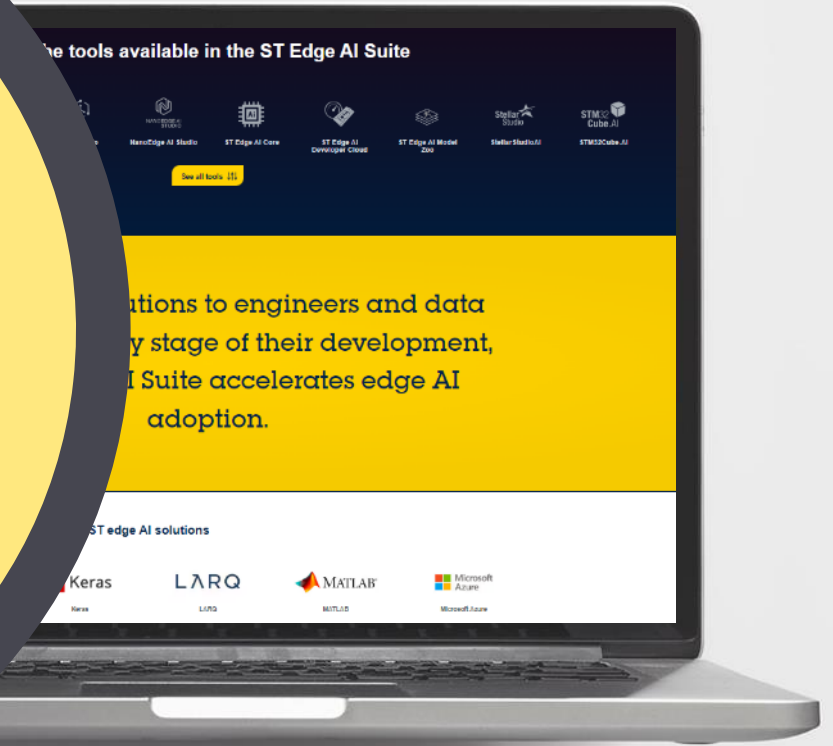
Stellar MCUs



MEMS MLC/ISPU



ST Edge AI Suite

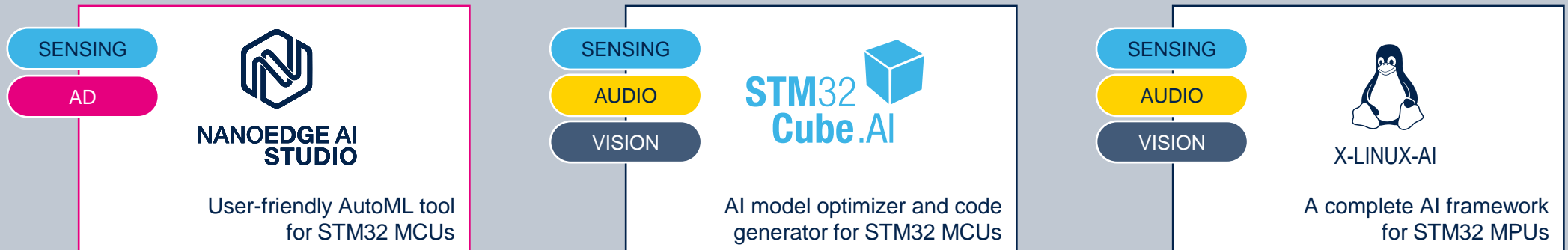


STM32 product offering

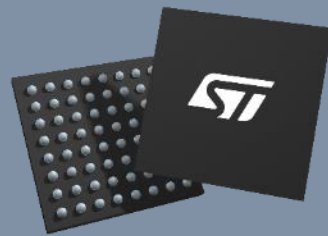
Enabling major edge AI technologies



Software tools for any user profile



Large choice of general purpose & **accelerated** hardware



STM32 MCUs



STM32N6 MCU with AI acceleration



STM32MP1 & STM32MP2 MPUs

ST Neural-ART Accelerator in STM32N6

600x

ML performance uplift*



**ST proprietary
neural processing unit (NPU)**

**Specialized hardware accelerator
performs advanced AI inference**

**Achieving edge AI performance
within MCU power budget**

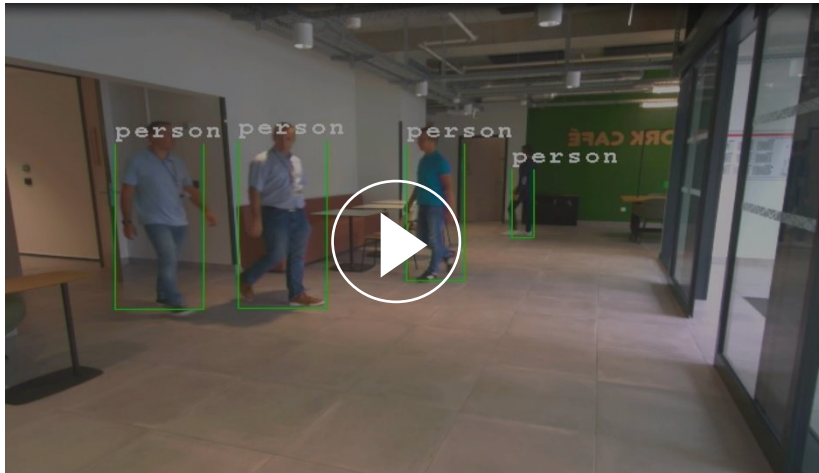


How the STM32N6 changes the game



An MPU-like end-user experience. Available on an MCU.

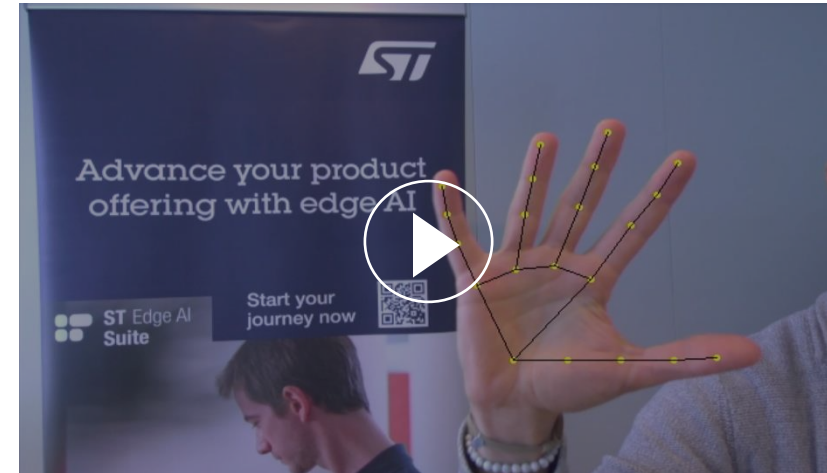
People detection



Pose estimation



Hand landmark



Software ecosystem

Jump-start your project

with an STM32 MCU

Follow these few simple steps to get started.

Start a project

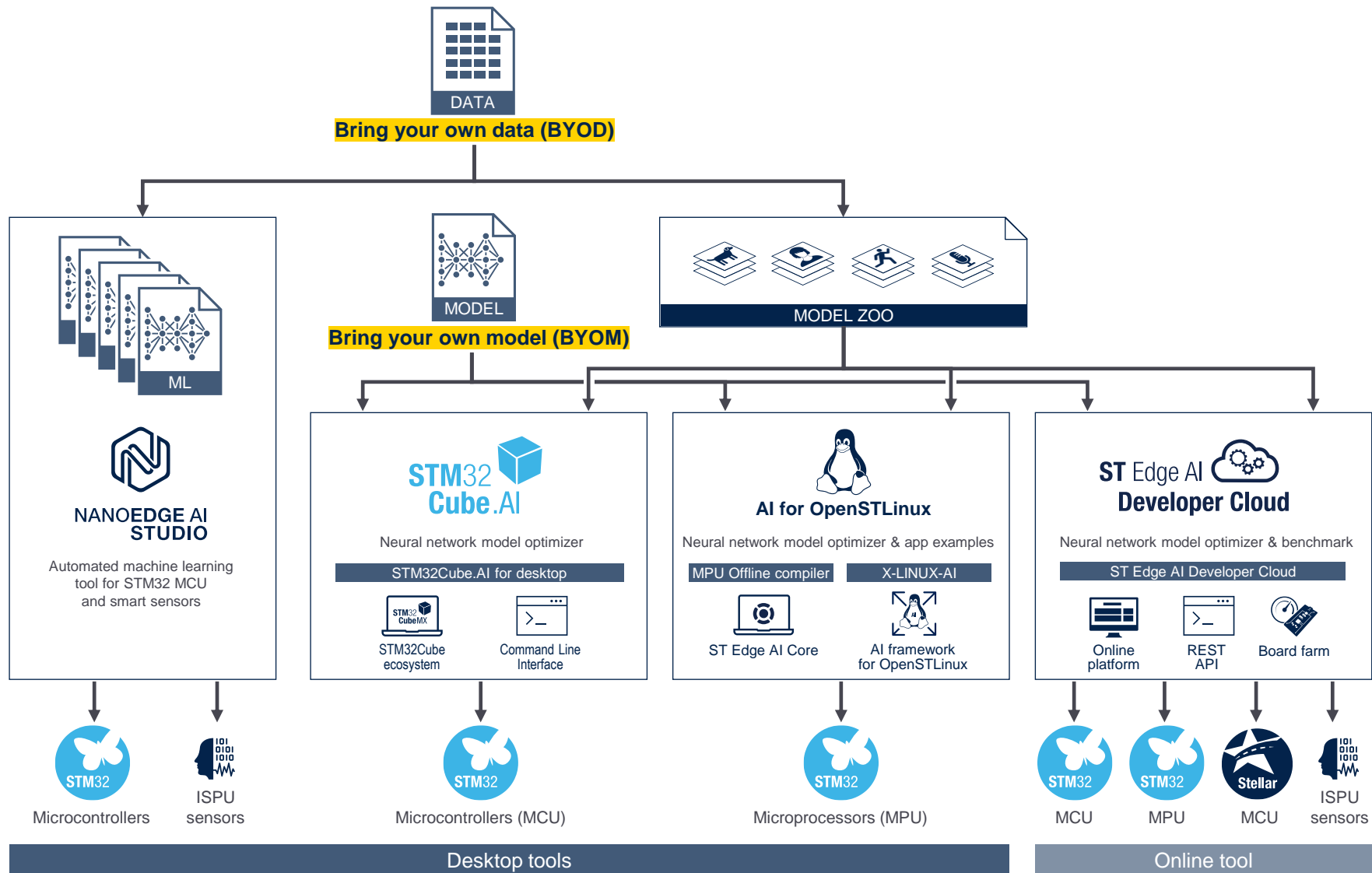
Microcontrollers, boards
and hardware tools,
software tools and
embedded software

MCU.	Tools.	Community.
3,300+	500+	135,000+
part numbers.	software & boards.	members.



life.augmented

A broad collection of free tools

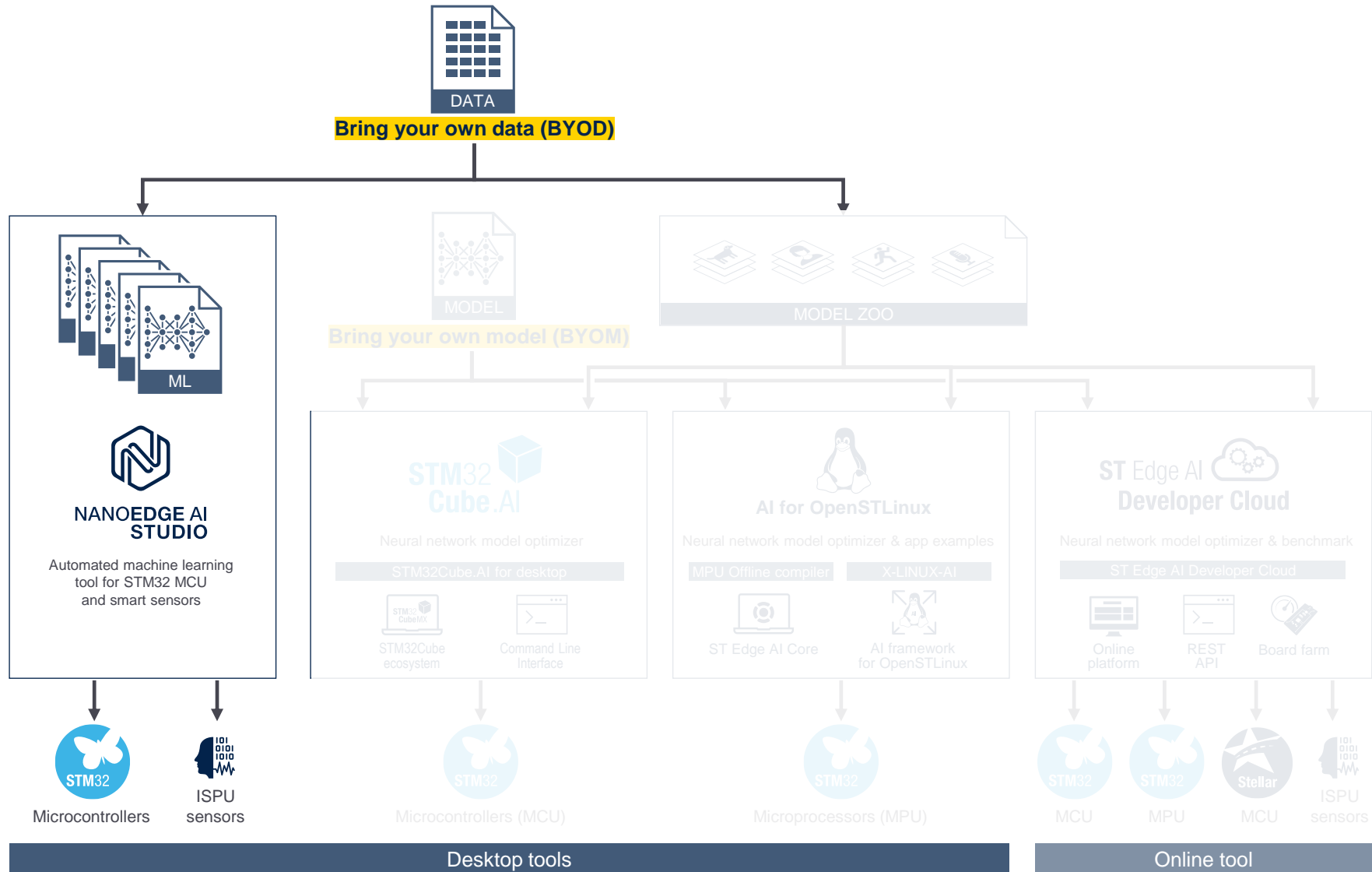


NanoEdge AI Studio, the AutoML tool



life.augmented

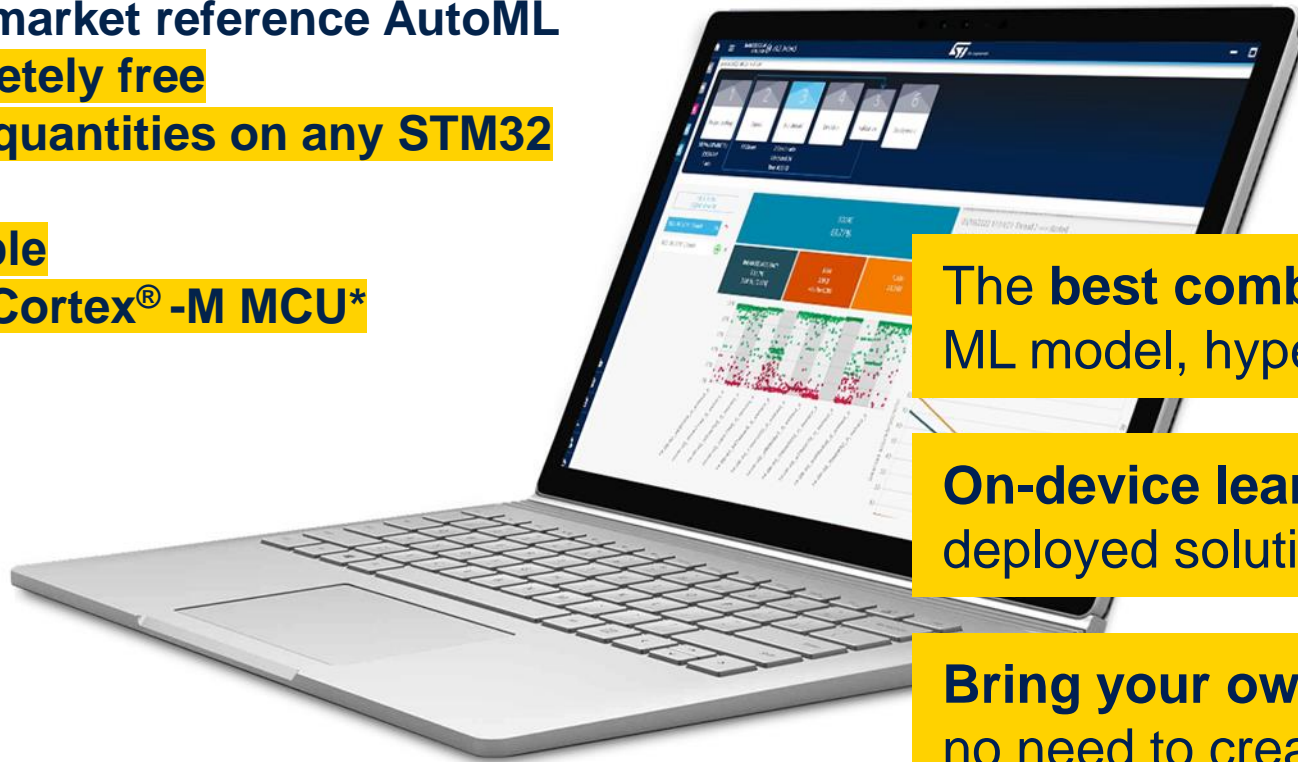
Software tool: NanoEdge AI Studio



Simplified edge 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 a
deployed solution without retraining

Bring your own data approach:
no need to create edge AI models

**under a special license agreement*

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



AD
Anomaly
Detection

1C
1-Class
Classification

nC
n-Class
Classification

E
Extrapolation

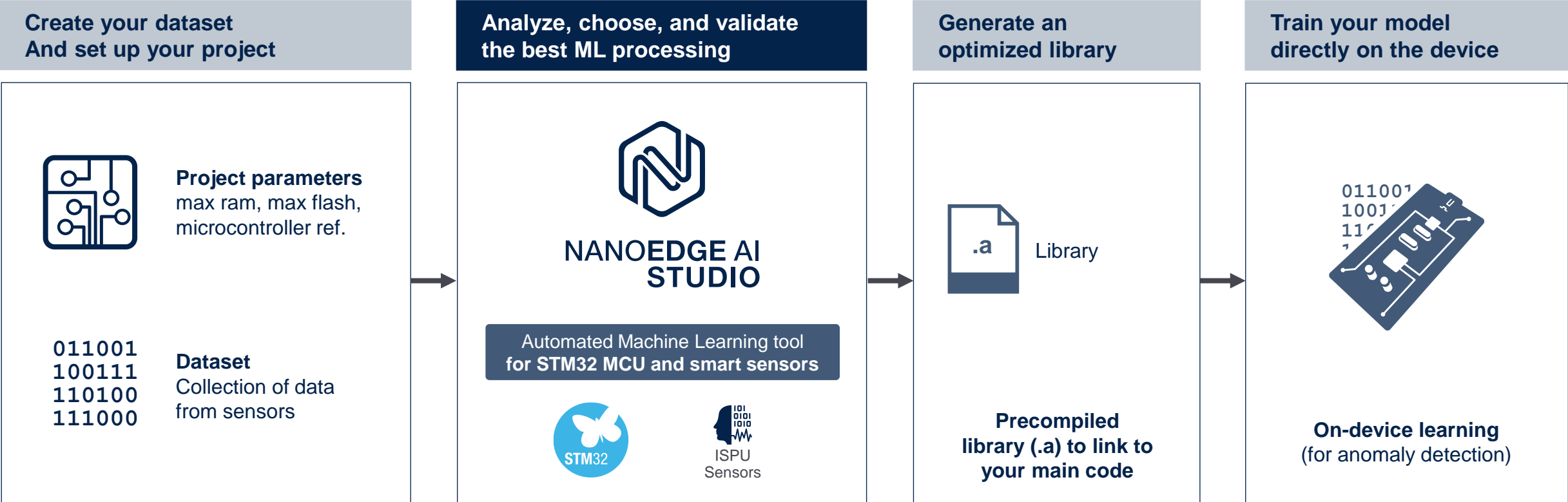
“
I want to anticipate product failures

“
I need to detect any outliers

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

“
I need to predict future states

NanoEdge AI Studio workflow



All the tools needed for data preparation



DM
Data
Manipulation

“

I want to clean my data quickly and efficiently”

The screenshot shows the NANOEDGE AI STUDIO v3.2.0 interface. At the top, there are three tabs: File, Action, and Result, all with checkmarks. The main area is divided into three panels:

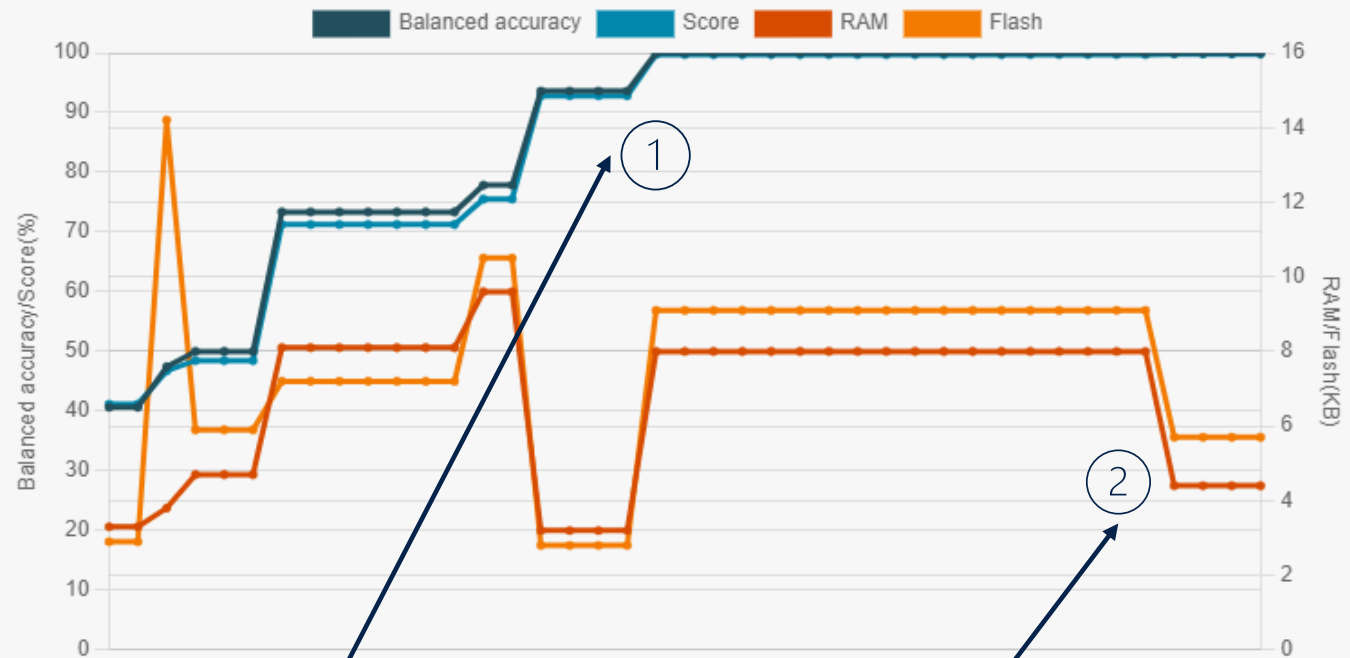
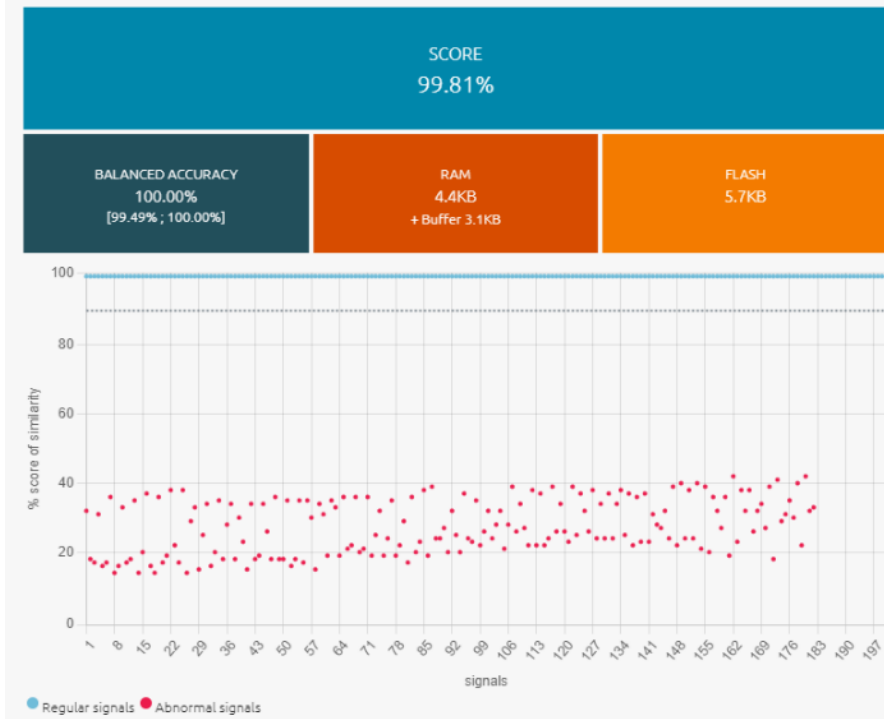
- File Panel:** Shows a file named "Log_ISPU_Spacer_2022-06-13-163838.log". It has a toggle for "Ignore first header line" and displays "182 Lines 768 Columns". The "Delimiter" is set to "Space". A "File preview" table is shown below.
- Action Panel:** Shows an "Extract lines" action. A slider is set to 182. Below the slider, a "RUN" button is visible. A preview of the extracted data shows a grid of values, with the top-right corner highlighted in blue.
- Result Panel:** Shows the result of the action: "182 Lines 768 Columns". It includes "RUN NEW ACTION" and "SAVE AS" buttons.

Below the File panel, there is an "Add file(s)" section with a dashed box and the text "Drop files or click to import".

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

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

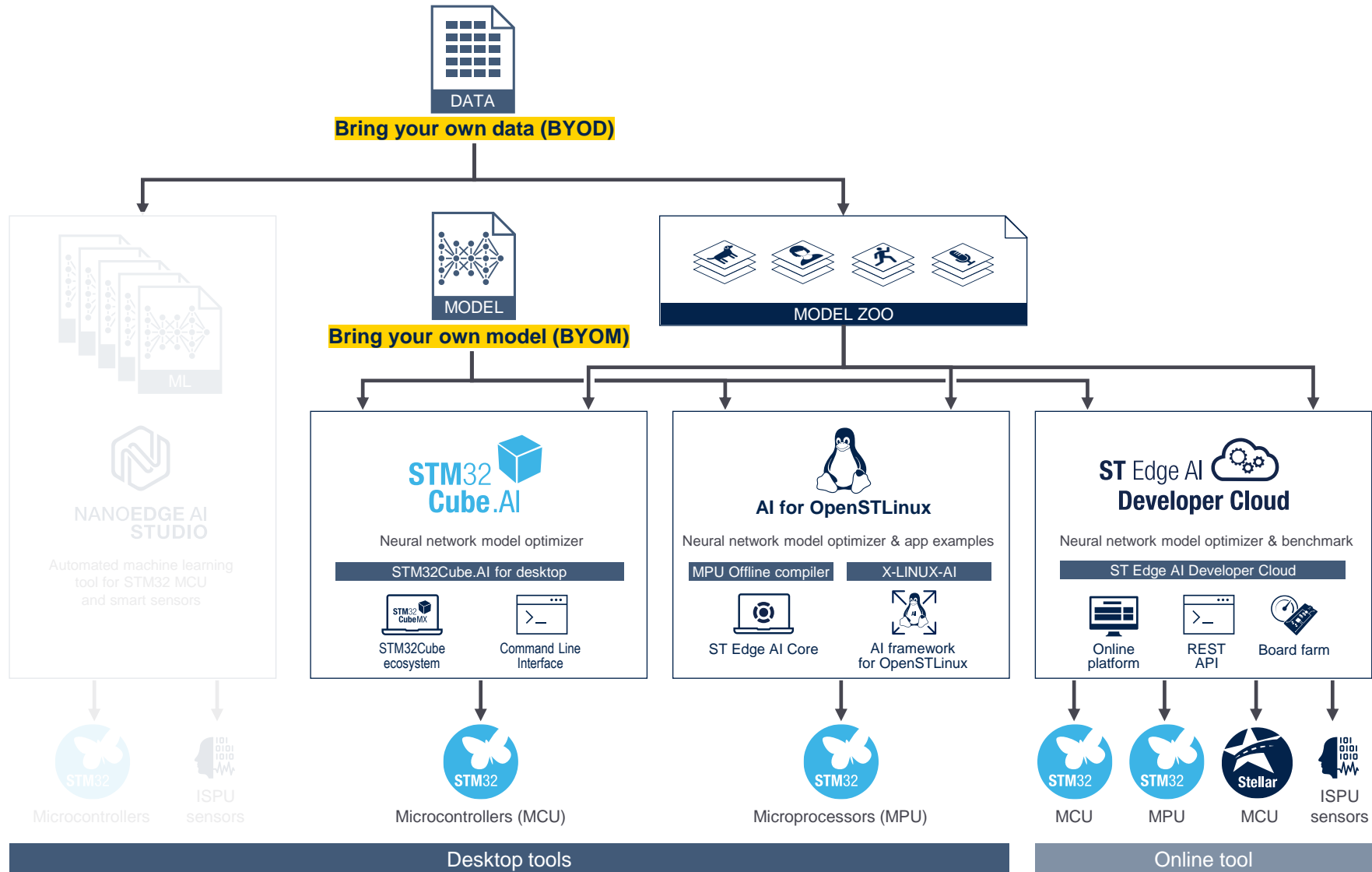
STM32Cube.AI & ST Edge AI Developer Cloud

The screenshot displays the STM32Cube.AI Developer Cloud interface. At the top, a progress bar shows five steps: Optimize (active), Quantize, Benchmark, Results, and Generate. Below this, the 'Model currently selected' is 'MOBilenet_18_025.H5'. The interface is divided into four columns: INPUT (float, 1x224x224x3), OUTPUT (float, 1x1x1x1), MODEL TYPE (float), and MACC (1207196). A 'Select another model' button is present. The 'Select your model optimization options' section includes: 'Balance between RAM size and inference time (-optimization balanced)' (selected), 'Optimize for RAM size (-optimization ram)', 'Optimize for inference time (-optimization time)', 'Use activation buffer for input buffer (-allocate-inputs)' (checked), and 'Use activation buffer for output buffer (-allocate-outputs)' (checked). An 'Optimize' button is at the bottom right of this section. The 'History of optimization results' table shows one entry for '1/19/23, 6:22 PM' with 'Level: Default' and 'Optimization: balanced'. The table columns are: Date & Time, Optimization, Allocate Inputs, Allocate Outputs, MACC, Float size, and RAM size. Summary statistics at the bottom of the table are: Total: 829 KiB (+0.09%) and Total: 650 KiB (+0.80%).

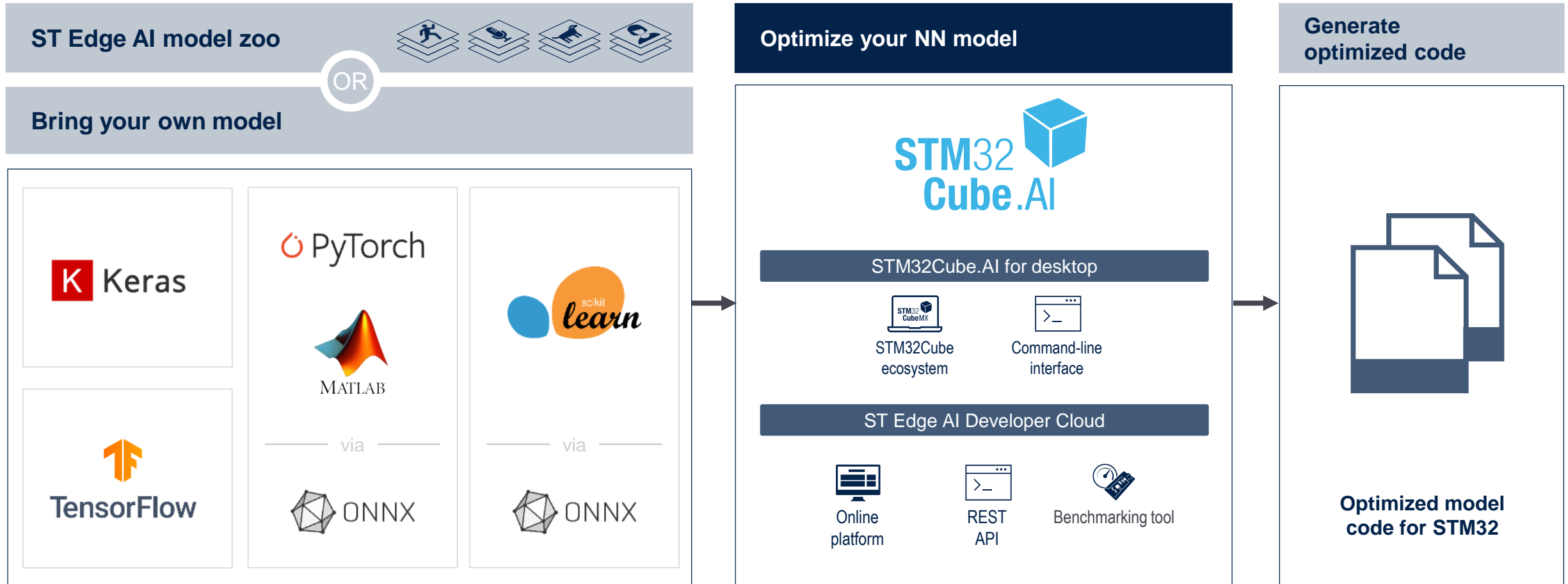
Date & Time	Optimization	Allocate Inputs	Allocate Outputs	MACC	Float size	RAM size
1/19/23, 6:22 PM	Level: Default	balanced	true	true	1207196	54543

Total: 829 KiB (+0.09%)
Total: 650 KiB (+0.80%)

Software tool: STM32Cube.AI



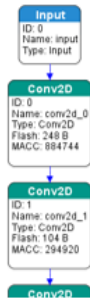
One tool – two versions to deploy AI on STM32



The 3 pillars of STM32Cube.AI

Graph optimizer

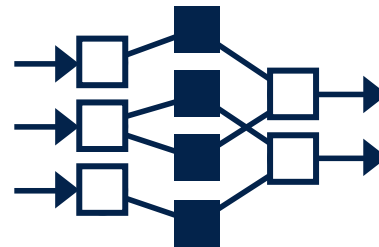
Automatically improve performance through graph simplifications & optimizations that benefit STM32 target hardware architectures



- Auto graphs rewrite
- Node/operator fusion
- Layout optimization
- Constant-folding...
- Operator-level info to fine-tune memory footprint and computation

Quantized model support

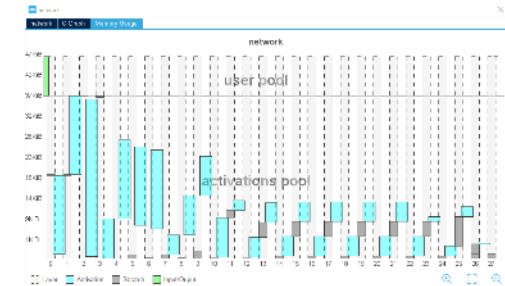
Import your quantized ANN to be compatible with STM32 embedded architectures while keeping their performance



- From FP32 to Int8 or mixed-precision
- Minimum loss of accuracy
- Code validation on target
 - Latency
 - Accuracy
 - Memory footprint

Memory optimizer

Optimize memory allocation to get the best performance while respecting the constraints of your embedded design



- Memory allocation
- Internal/external memory repartition
- Model-only update option

STM32Cube.AI is **free of charge**, available both in graphical interface and in command line.

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

Discover the ST Edge AI Developer Cloud

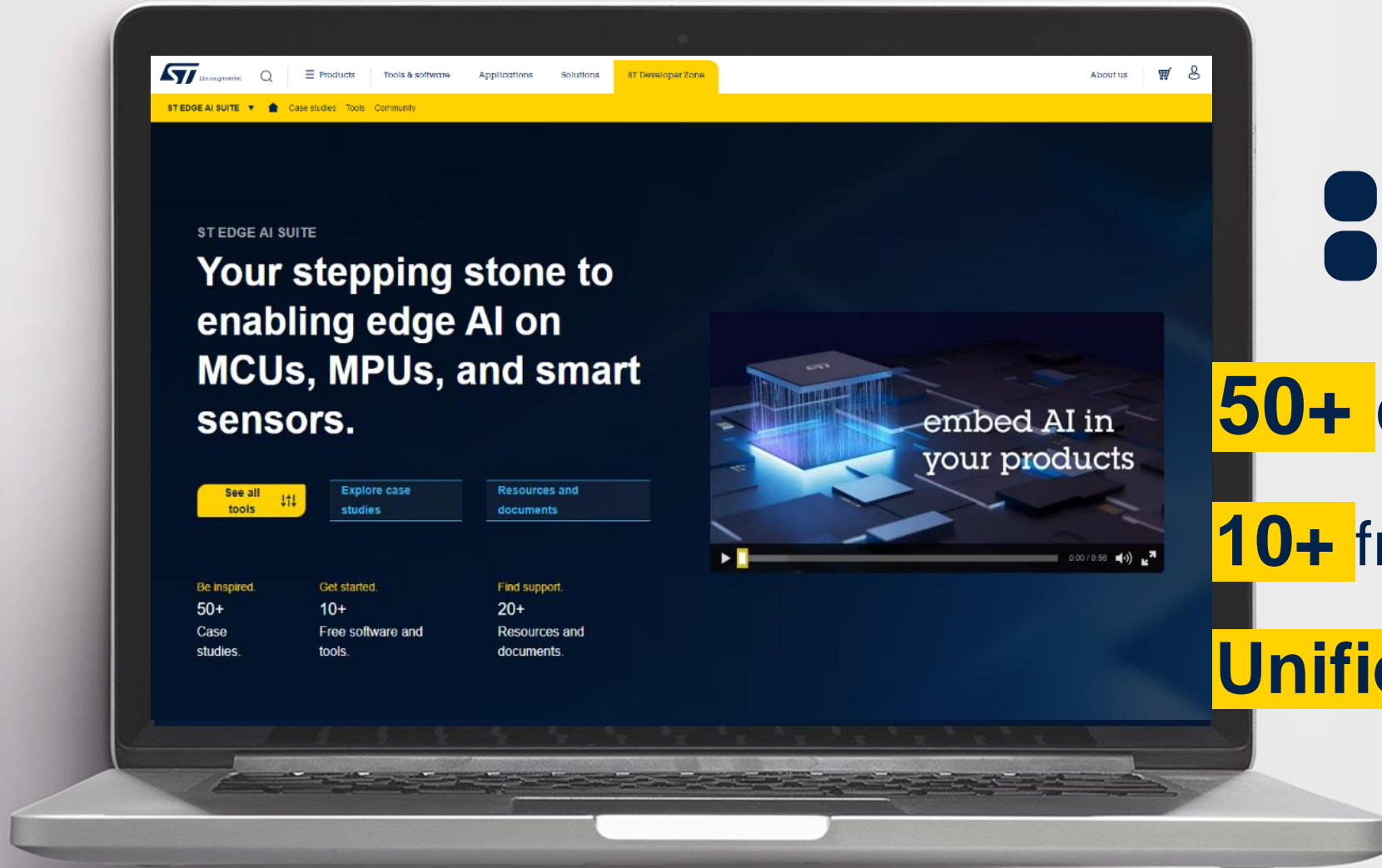
4 steps
benchmark
edge AI models
remotely

ST Edge AI Suite



life.augmented

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**

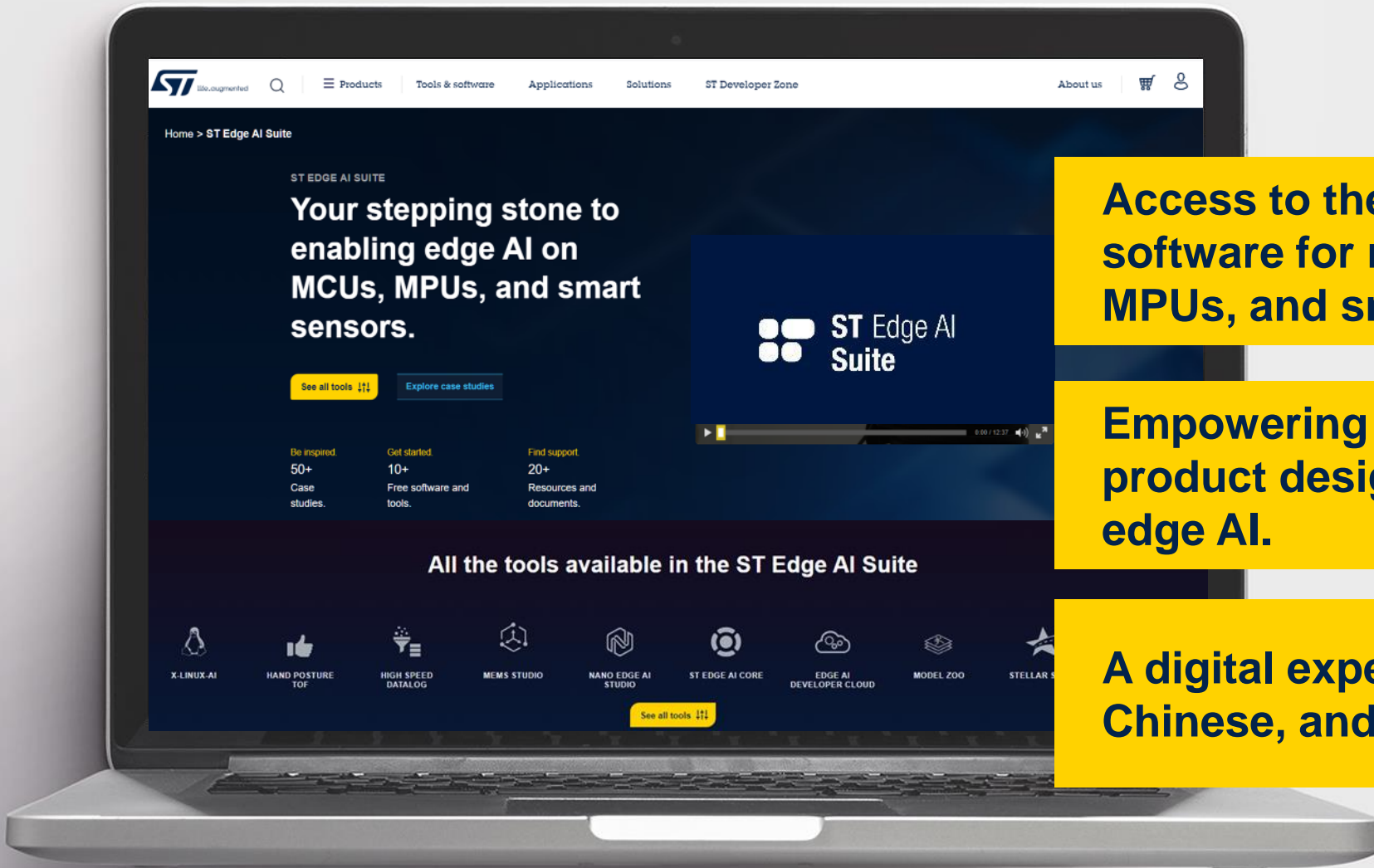
50+ case studies

10+ free software tools

Unified AI core technology



ST Edge AI Suite

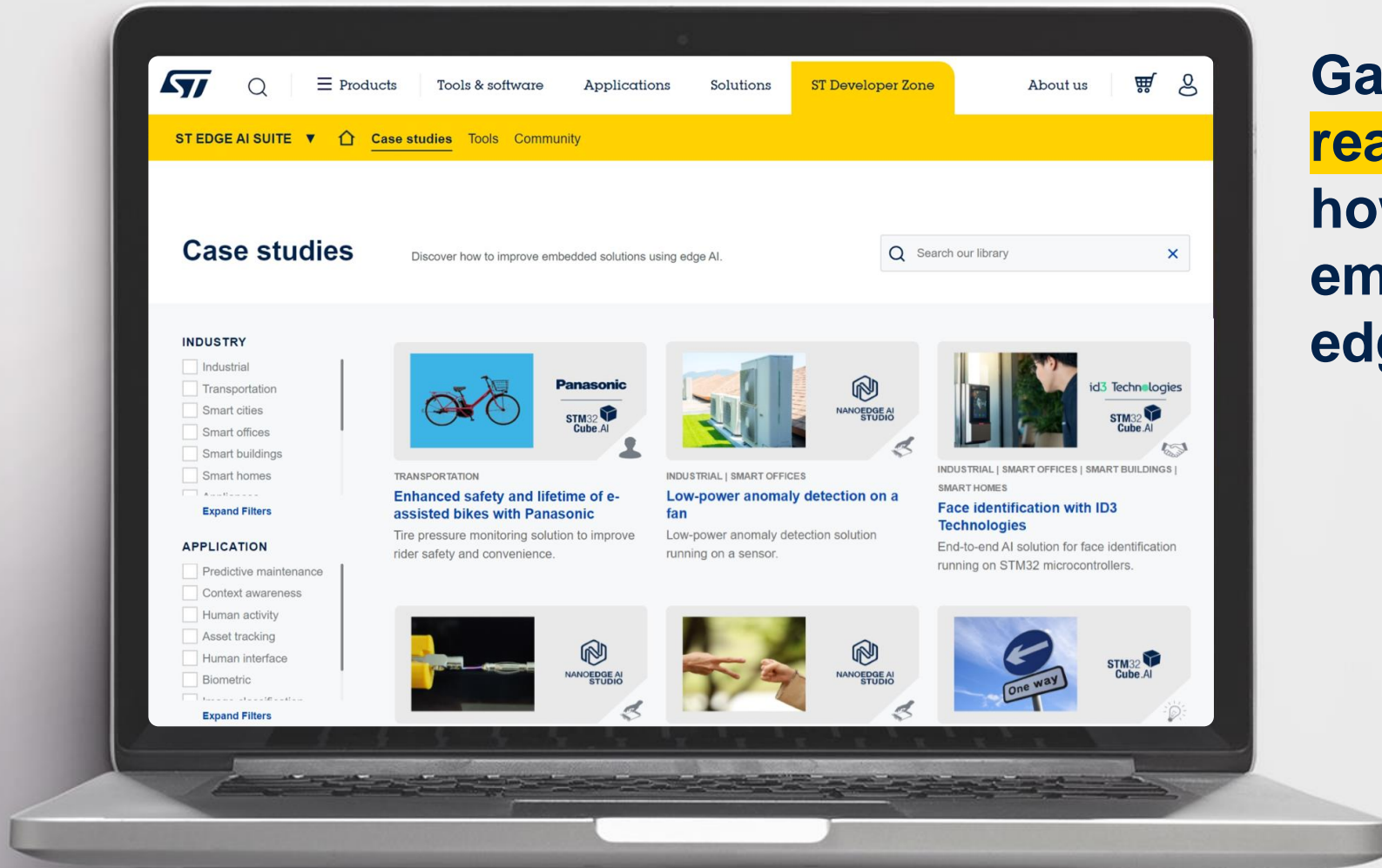


Access to the enabling hardware and software for running edge AI on MCUs, MPUs, and smart sensors.

Empowering developers, data scientists, and product designers to harness the power of edge AI.

A digital experience available in English, Chinese, and Japanese.

Explore inspiring case studies



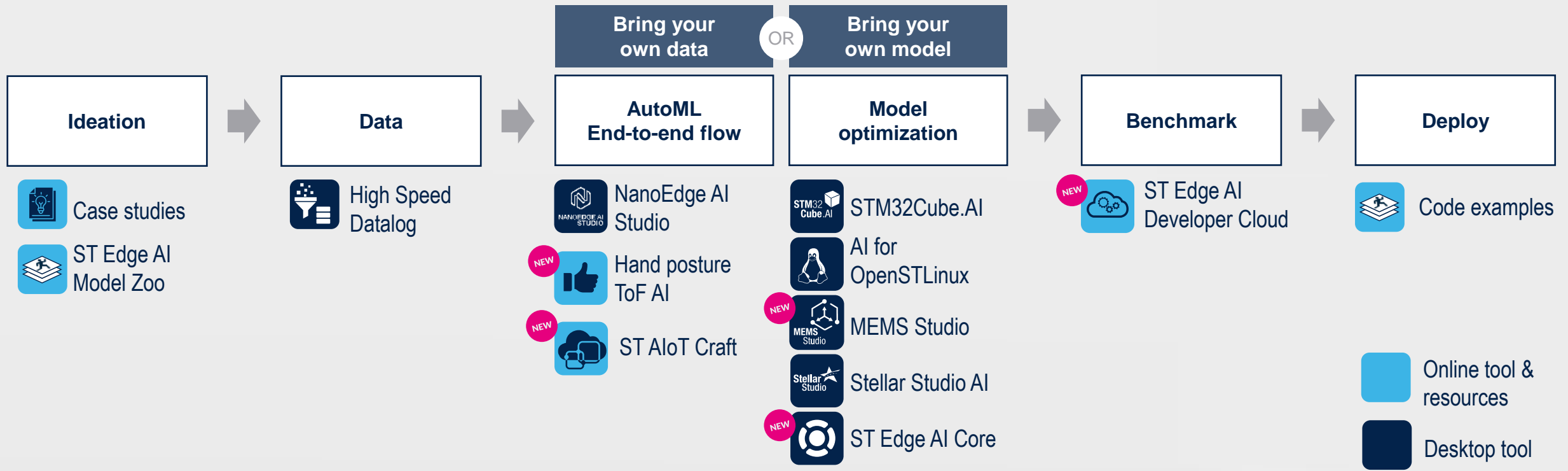
Gain **valuable insights from real-world case studies** on how to improve your embedded solutions using edge AI.



[Edge AI case studies](#)

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/stm32ai

© 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.

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



life.augmented