STM32 for AI applications

Redefining microcontroller performance to drive innovation in industrial and consumer applications

Daniel Wang

Technical Marketing Manager

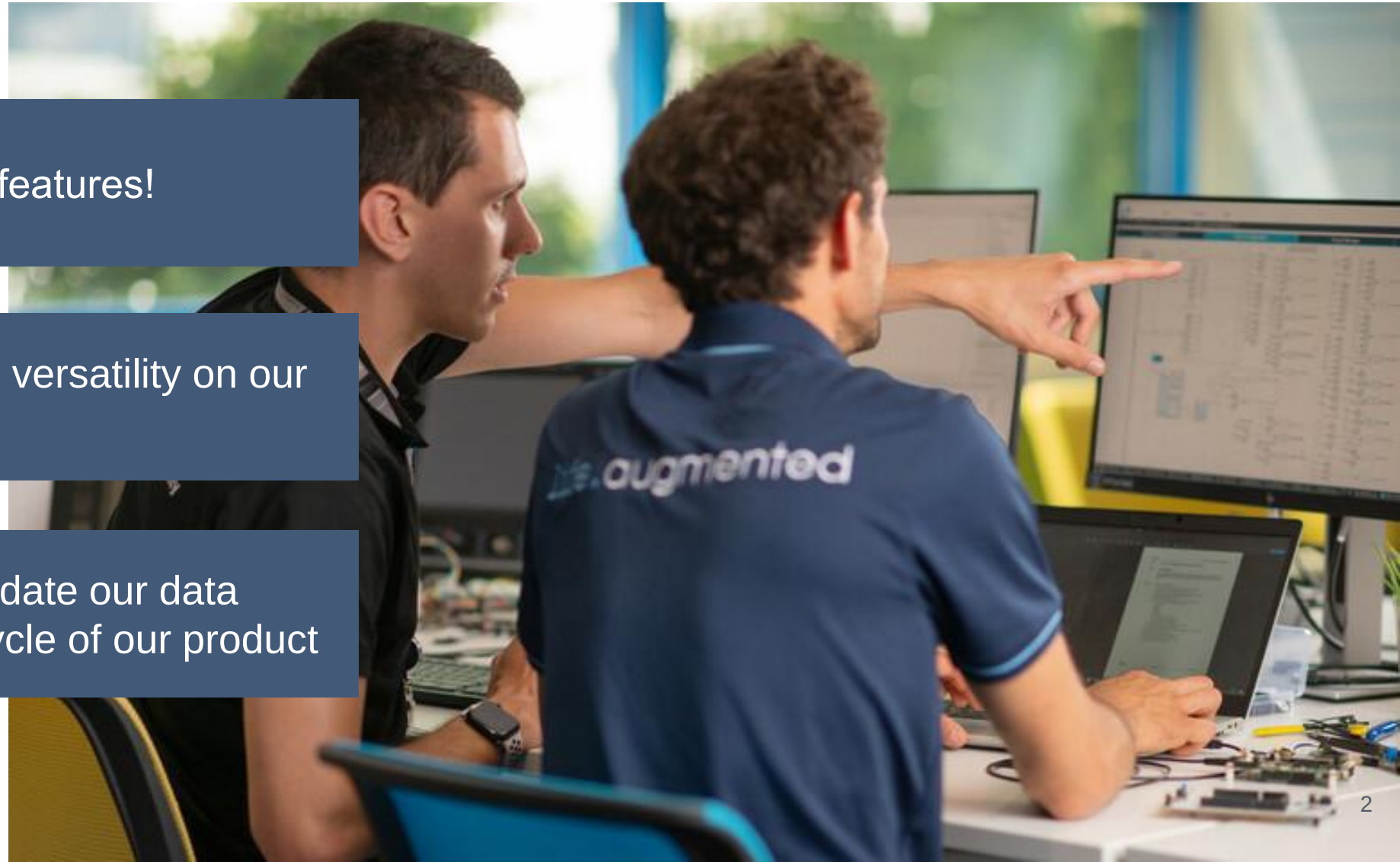
STMicroelectronics

What is the market telling us?

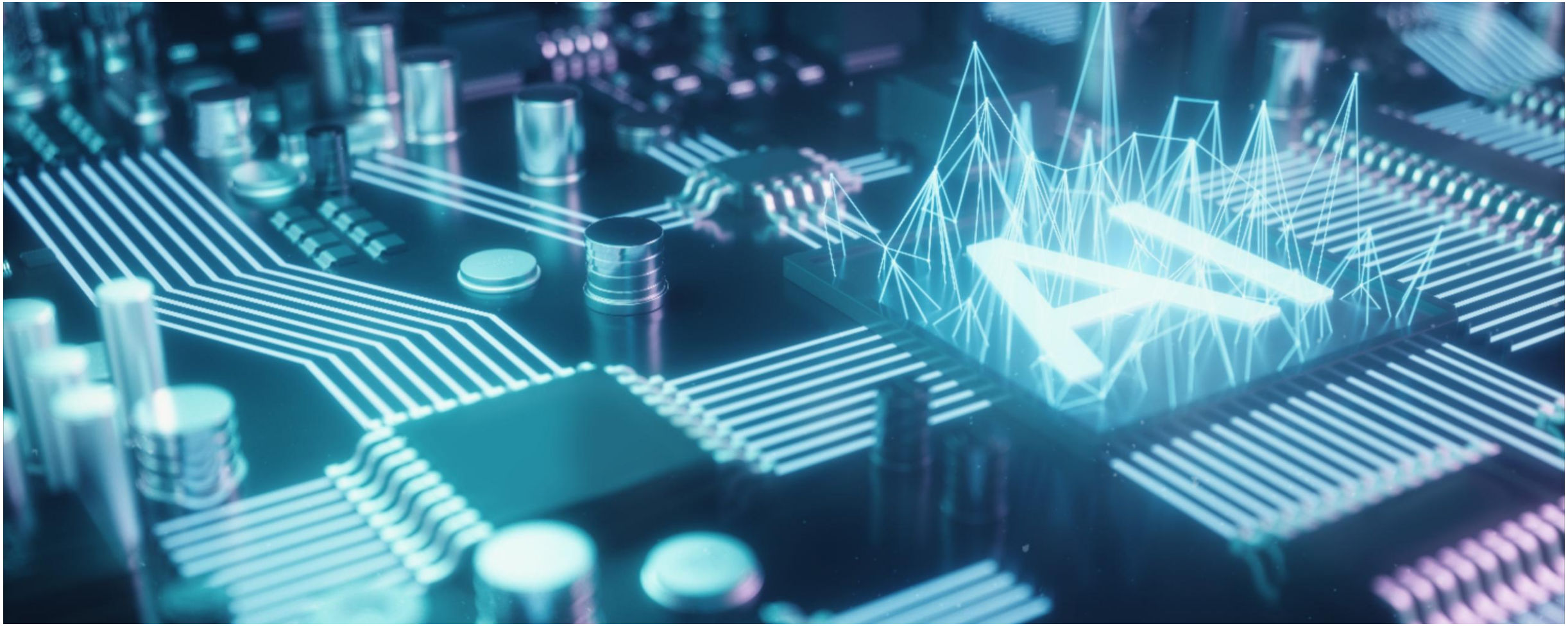
> We’re looking for new features!

> We want to have more versatility on our data analysis

> We want to be able update our data analysis over the lifecycle of our product

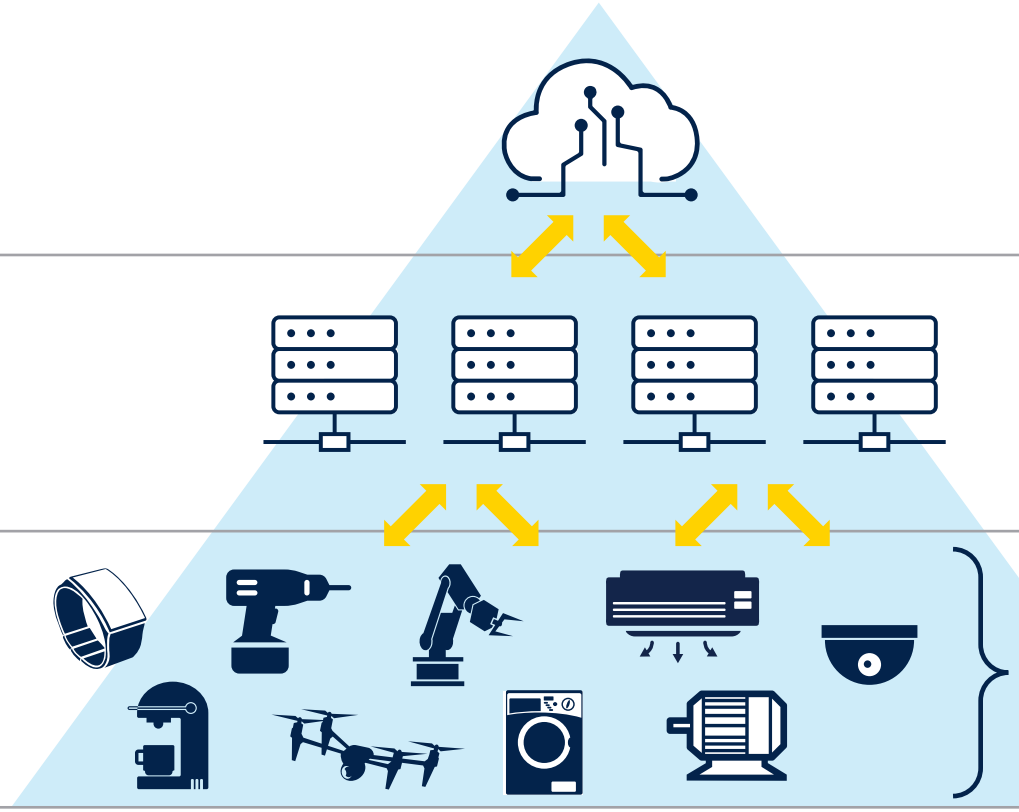


This is where AI opens new horizons for embedded design!



Distributed Artificial Intelligence approach

Leverage billions of devices at the Edge!



Data Center Cloud

Analytics, storage, compute

Edge Nodes

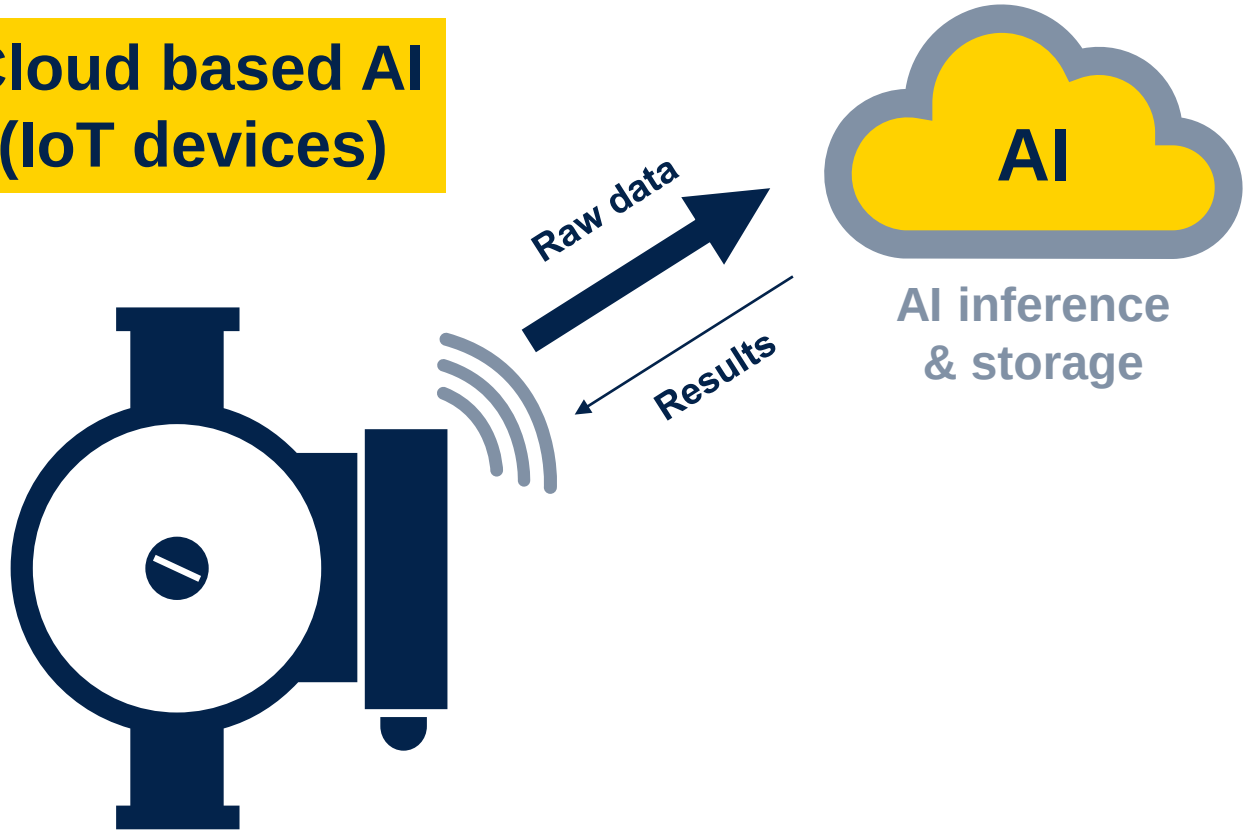
IoT gateways, micro datacenters

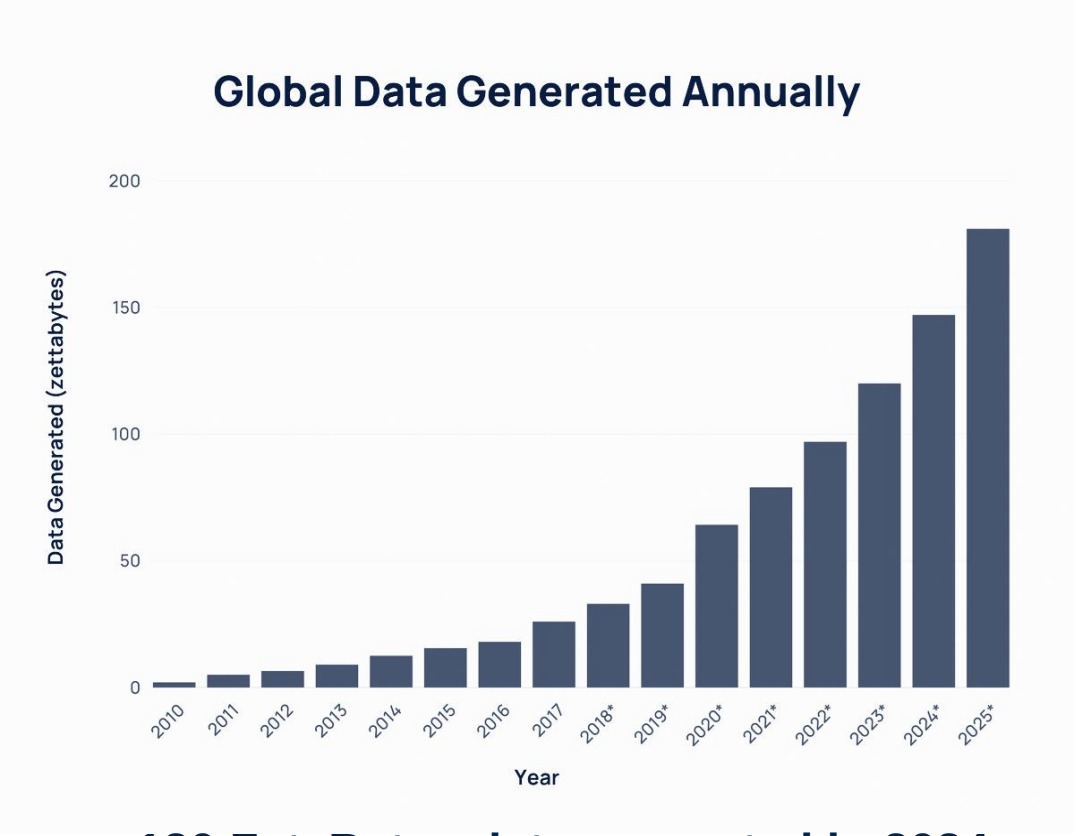
Edge Things

Real time, local processing

Cloud processing for AI & IoT: Generating a tsunami of data

Cloud based AI (IoT devices)



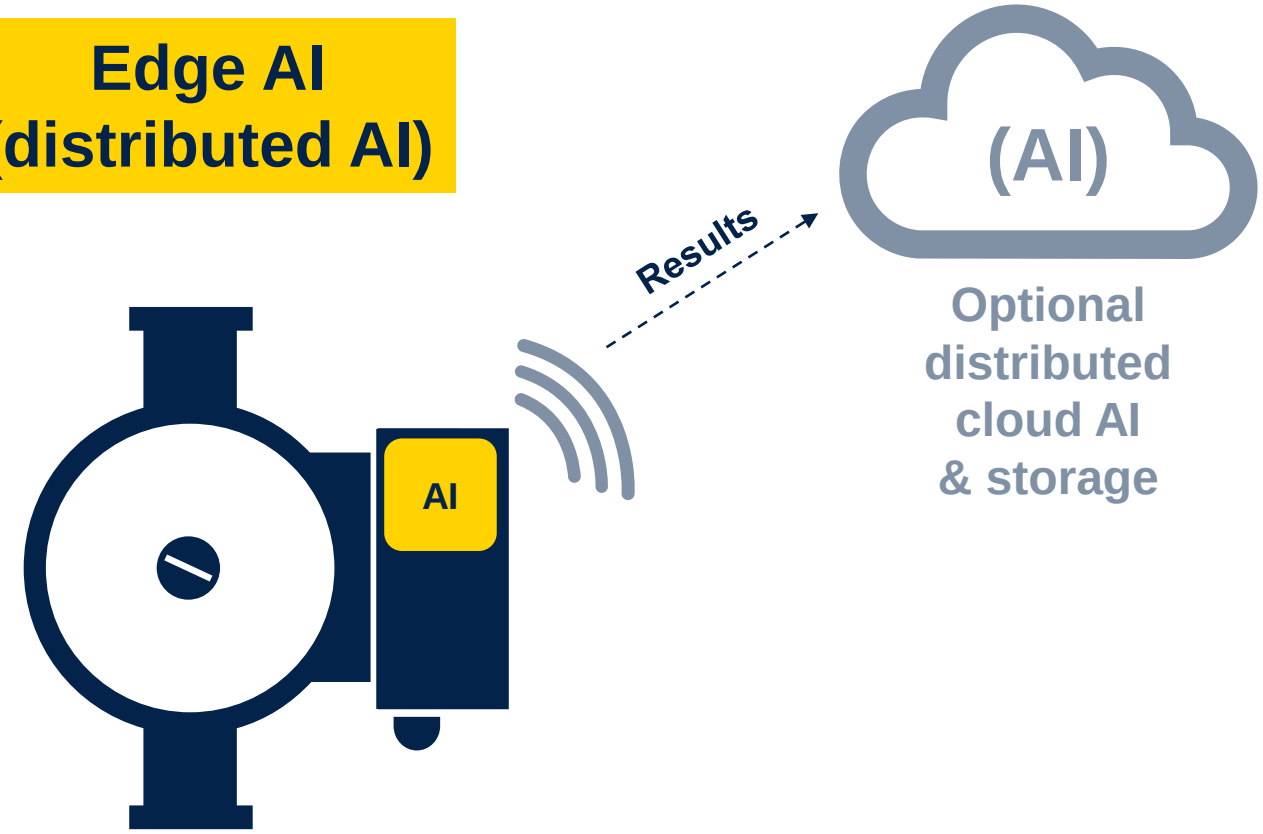


120 ZetaBytes data generated in 2024
> 180 ZetaBytes in 2025

Source: explodingtopics.com

The rise of edge AI: AI at device level

Edge AI (distributed AI)



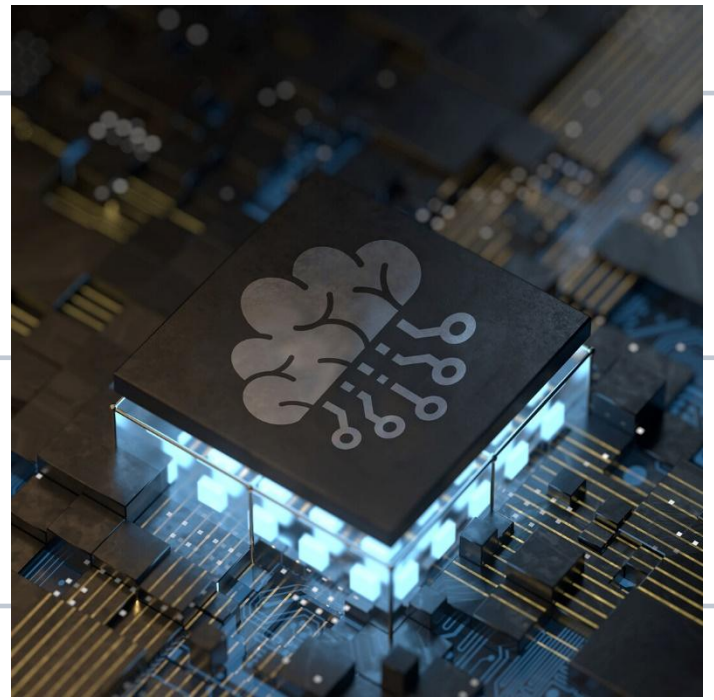
Artificial intelligence at the Edge

Moving part of Artificial Intelligence closer to the data acquisition brings several benefits

Ultra-low latency
Real-time applications

More reliability

Security of data
No sharing in the cloud



Privacy by design
GDPR compliant

Sustainable on energy
Low-power consumption

Better user experience

From DMIPS to TOPS, the paradigm shift
Opening a new range of embedded AI applications

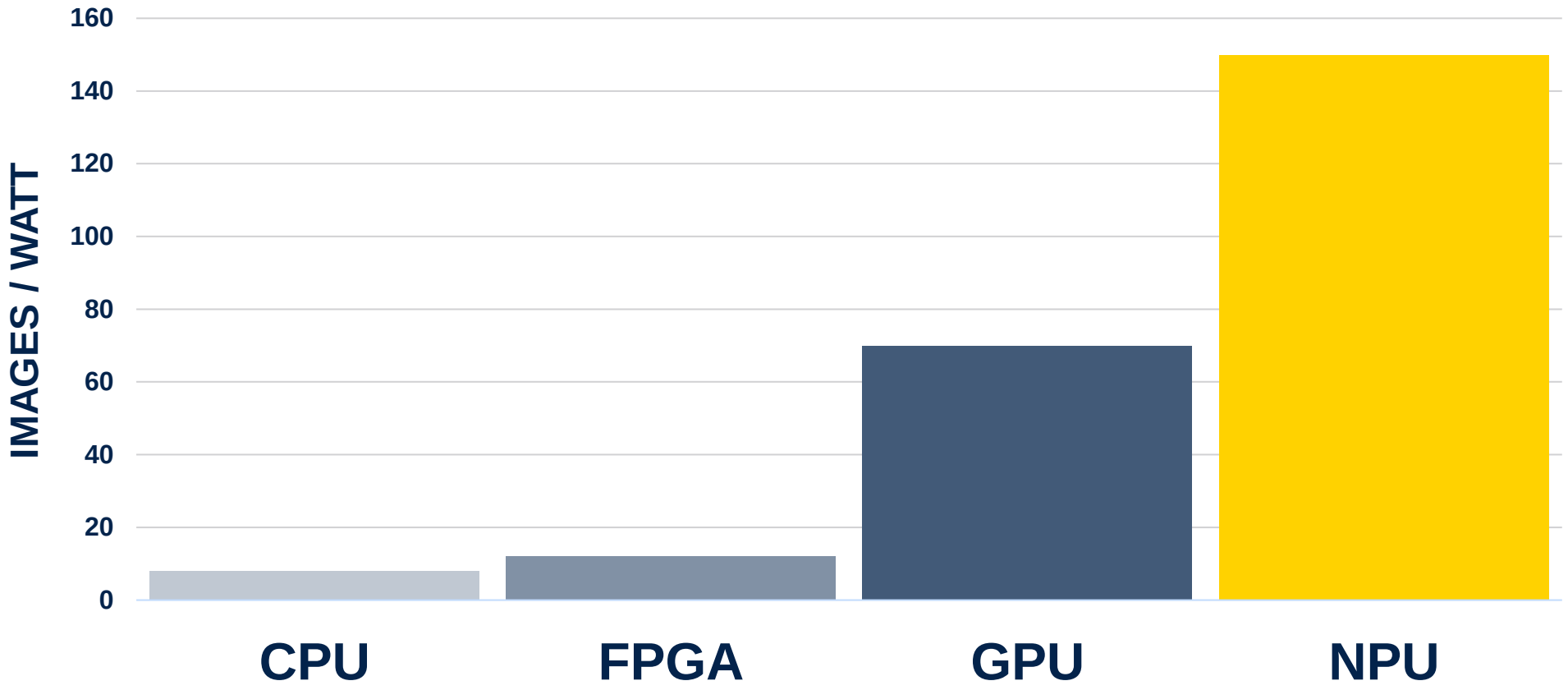
- Object segmentation localization
- Pose estimation
- Object classification
- Speech recognition
- Sound analysis
- Face/people detection
- Wake word
- Time series classification
- Anomaly detection

Microcontrollers (Arm® Cortex® -M)

Microcontrollers with NPU accelerator

Mono-modality workloads	**Multi-modality workloads**
Static single subjects	**Faster moving multiple subject**
Low power	**High efficiency**
Optimal light conditions	**Open light conditions**
Acceptable precision	**High precision**
Low resolution and framerate	**Higher resolution and framerate**

Edge AI acceleration requires new architectural solution: the NPU



GPU: graphic accelerator

NPU: neural processing unit (AI accelerator)

Source W. Dally

The STM32 portfolio

Five product categories

- **Wireless MCU** — Short- and long-range connectivity
- **Ultra-low-power MCU**
- **Mainstream MCU**
- **High-performance MCU**

  32-bit general-purpose microcontrollers: from 75 to 3,360 CoreMark score
- **Embedded MPU** — 32- and 64-bit microprocessors





Enabling edge AI solutions

Scalable security

MPU portfolio

MCU portfolio

STM32 portfolio

LONGEVITY 10 YEARS COMMITMENT

MPU

STM32MP1
1 GHz Cortex-A7
209 MHz Cortex-M4

STM32MP2
Dual 1.5 GHz Cortex-A35
400 MHz Cortex-M33

High-performance MCUs

STM32F7
1,082 CoreMark
216 MHz Cortex-M7

STM32H7
3347 CoreMark
Up to 600 MHz Cortex -M7
240 MHz Cortex -M4

STM32N6
3,360 CoreMark
800 MHz Cortex -M55
Neural processing unit

STM32F2
398 CoreMark
120 MHz Cortex-M3

STM32F4
608 CoreMark
180 MHz Cortex-M4

STM32H5
1,023 CoreMark
250 MHz Cortex-M33

Mainstream MCUs

STM32F3
245 CoreMark
72 MHz Cortex-M4

STM32G4
569 CoreMark
170 MHz Cortex-M4

Mixed-signal MCUs

STM32C0
114 CoreMark
48 MHz Cortex M0+

STM32F0
106 CoreMark
48 MHz Cortex-M0

STM32G0
142 CoreMark
64 MHz Cortex-M0+

STM32F1
177 CoreMark
72 MHz Cortex-M3

Ultra-low-power MCUs

STM32L0
75 CoreMark
32 MHz Cortex-M0+

STM32U0
140 CoreMark
56 MHz Cortex-M0+

STM32L4
273 CoreMark
80 MHz Cortex-M4

STM32U3
393 CoreMark
96 MHz Cortex-M33

STM32L4+
409 CoreMark
120 MHz Cortex-M4

STM32L5
443 CoreMark
110 MHz Cortex-M33

STM32U5
651 CoreMark
160 MHz Cortex-M33

Wireless MCUs

STM32WL
162 CoreMark
48 MHz Cortex-M4
48 MHz Cortex-M0+

STM32WB0
156 CoreMark
64 MHz Cortex-M0+

STM32WB
216 CoreMark
64 MHz Cortex-M4
32 MHz Cortex-M0+

STM32WBA
407 CoreMark
100 MHz Cortex-M33

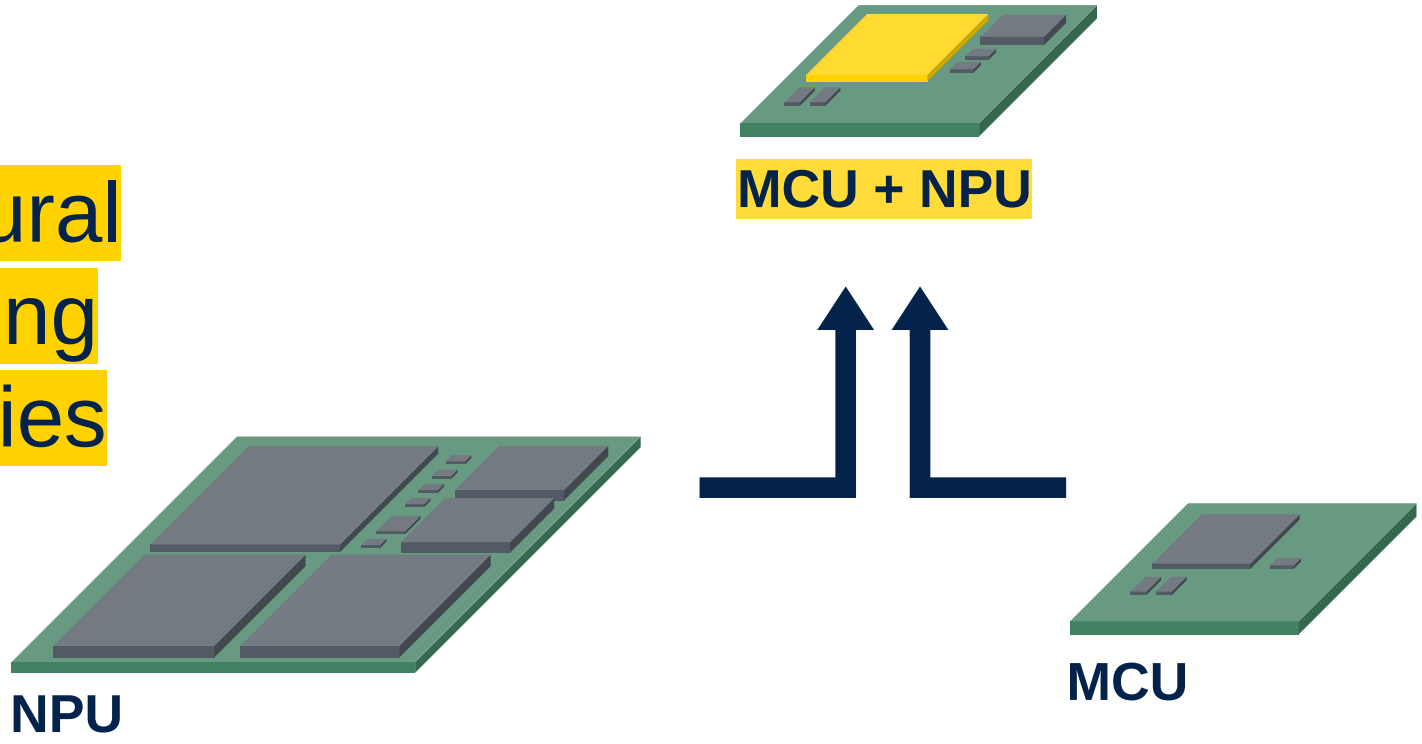
Latest product generation
Radio coprocessor only
New series or lines introduced in 2025

The first high-performance STM32 MCU with AI acceleration

Enabling unmatched edge AI performance on an MCU

Benefit from extended neural network computing capabilities while leveraging the advantages of an MCU.

High neural processing capabilities



Small footprint

Lower power

Lower cost

Lower BOM

Faster boot/wkup

STM32N6 feature overview

Dedicated embedded neural processing unit (NPU)

- 600 GOPS NPU
- 3 TOPS/W power consumption

Arm® Cortex®- M55 core

- 1280 DMIPS / 3360 CoreMark
- New DSP extensions (MVE)

Embedded RAM

- 4.2 Mbytes of embedded RAM for real-time data processing and multitasking

Computer vision pipeline

- Parallel and MIPI CSI-2 camera module I/F
- Dedicated image processor (ISP)

Extended multimedia capabilities

- 2.5D graphics accelerator
- H.264 encoder, JPEG encoder/decoder

Extended security features

- Arm® TrustZone® for the Cortex®-M55 core and the NPU
- Target certifications SESIP3, PSA L3

* 600 GOPS NPU vs 1 GOPS NN peak processing capabilities on STM32H7

Optimize your application with the large embedded memory

Large embedded RAM

4.2 Mbytes

Fast external memory I/F

Hexa-SPI	Octo-SPI	FMC
Up to 800 Mbytes/s	Up to 400 Mbytes/s	Up to 664 Mbytes/s

Large contiguous embedded memory

- Ideal for running neural networks or graphic applications
- External RAM becomes optional

Fast serial I/F for external memories

- Allows the use of fast and cost-effective memory
- Hexa-SPI for fast access to RAM
- Octo-SPI for secured flash memory

Flashless configuration

- Adaptability to application requirements
- Enabling cost flexibility

Flexible memory controller

PSRAM, SDRAM, NOR, NAND

Improved security with on-the-fly encryption

Hardware-accelerated crypto engine on all interfaces

Elevating graphics performance

Graphic accelerators

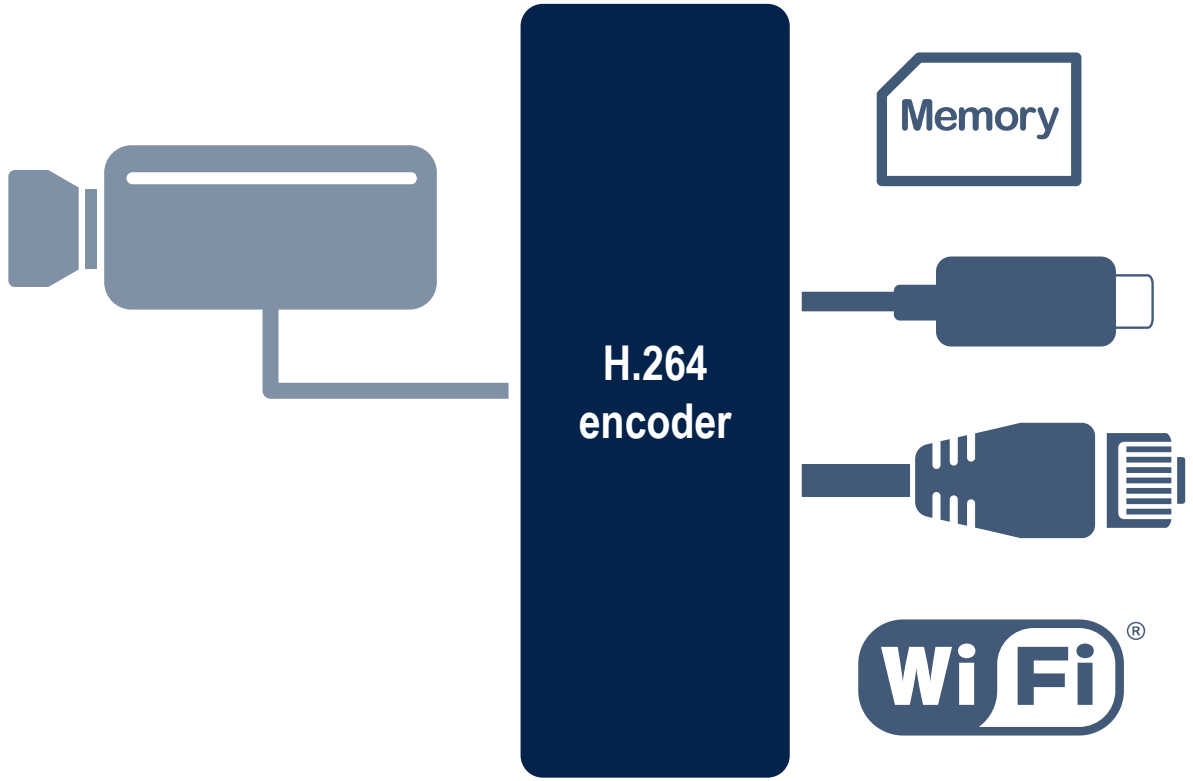
NeoChrom GPU	• 2.5D GUI acceleration • Perspective correct texture mapping (scale, rotate, flip)
Chrom-ART Accelerator	Efficient 2D graphics sub-system
JPEG codec	MJPEG video coding & decoding
Chrom-GRC	Framebuffer optimization

Elevating multimedia experiences

Multimedia unit

H.264 encoder

- 1080p15 and 720p30
- Real-time streaming over USB, over Ethernet, Wi-Fi



Geared for computer vision applications

Enabling fast & efficient image acquisition and processing thanks to a widely adopted camera interface and embedded ISP.

CAMERA PIPELINE

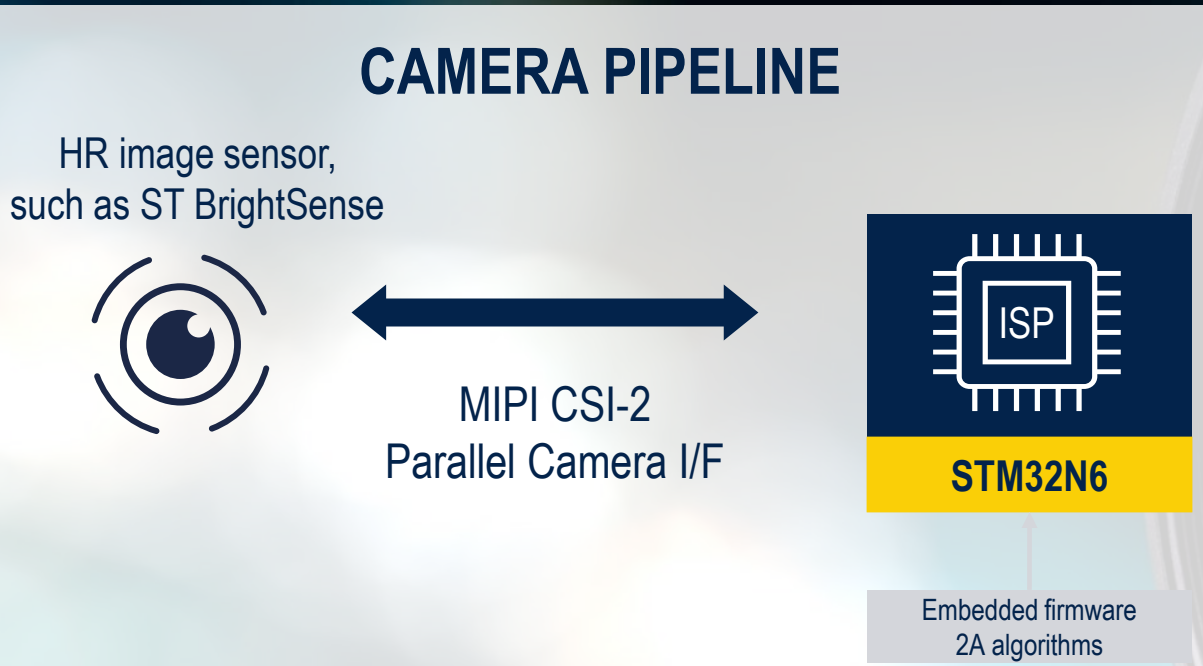


Image signal processor (ISP)

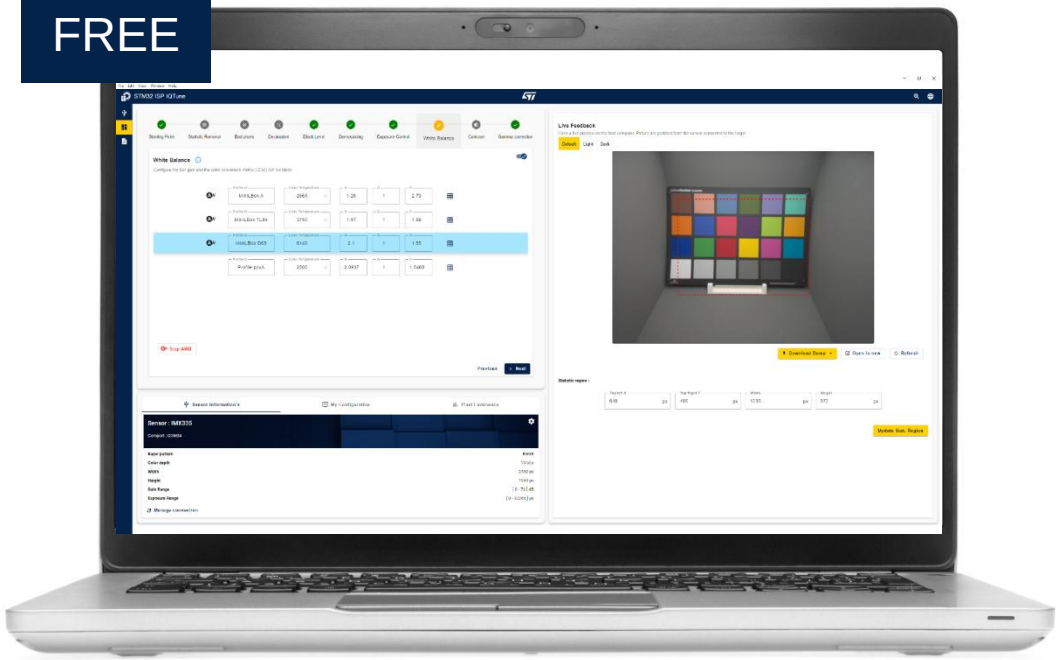
- Dimensioned for 5 Mpixel camera at 30 FPS
- Generates 3 different outputs from the same input for sending to the multimedia encoder or to the NPU
- ISP IQTune Software tool to tune ISP for cost savings and design flexibility

Embedded firmware on Arm® Cortex® core

- 2A for auto white balance and auto exposure
- Image processing library

Configure the image signal processor for free

FREE



Access now

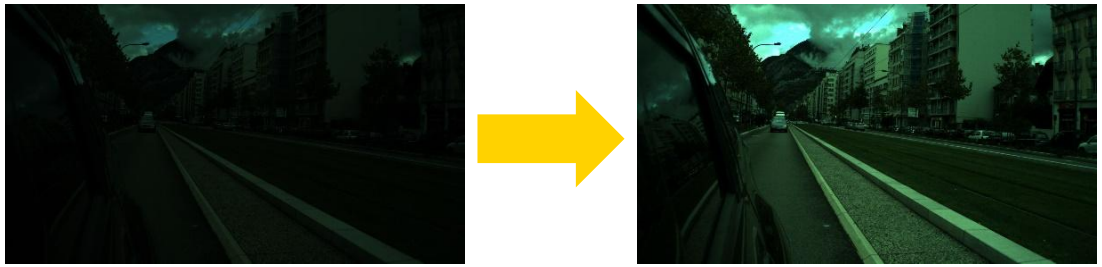
Industry-first software tool for ISP tuning on MCUs and MPUs.

Save ISP tuning cost and gain efficiency.

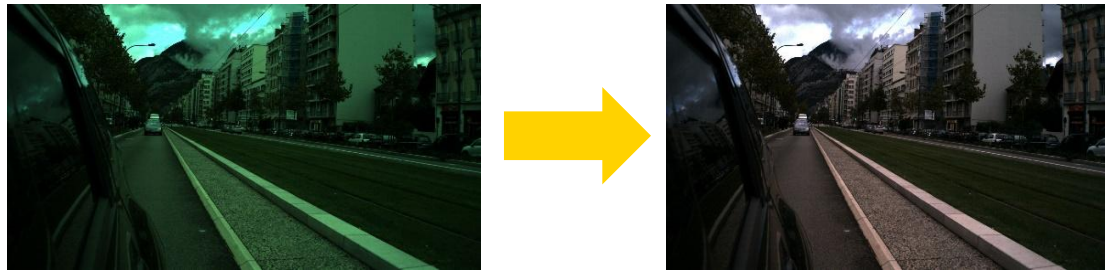
Flexibility to configure the ISP to your application requirements.

Main corrections

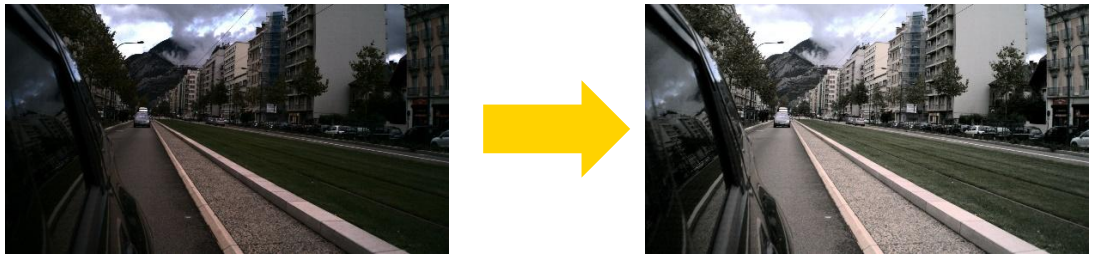
AEC (exposure)



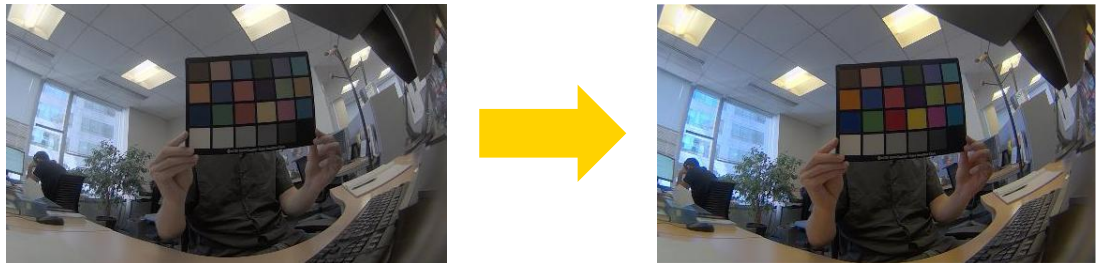
AWB (white balance)



ACE (contrast)



CCM (color)



Smoother and richer graphics with NeoChrom GPU

The NeoChrom GPU offloads the CPU from the graphic computations, freeing up the memory and boosting performance.

Fully supported by **TouchGFX** and partner GUI software.

NeoChrom GPU

Scale/animate bitmaps

Full screen transitions

360° Bitmap rotations

Text animation

Vector graphics (software)

Perspective correct texture mapping

Fast 2D bitmap copy color format conversion

MJPEG videos

STM32 MCUs for building UIs

State-of-the-art mechanisms for high security levels

Extensive security mechanisms to protect AI algorithms, ensure hardware robustness to attacks, and enable a multitenant approach.

psacertified™ level three

Target certification

SESIP™3

Target certification

Cryptography for hardware robustness,
including MCU and NPU memory on-the-fly encryption & decryption (*)

Hardware and temporal code isolation
for runtime protection

Device authentication
during product life cycle

(*) Available soon

Memory protections
against illegal access control

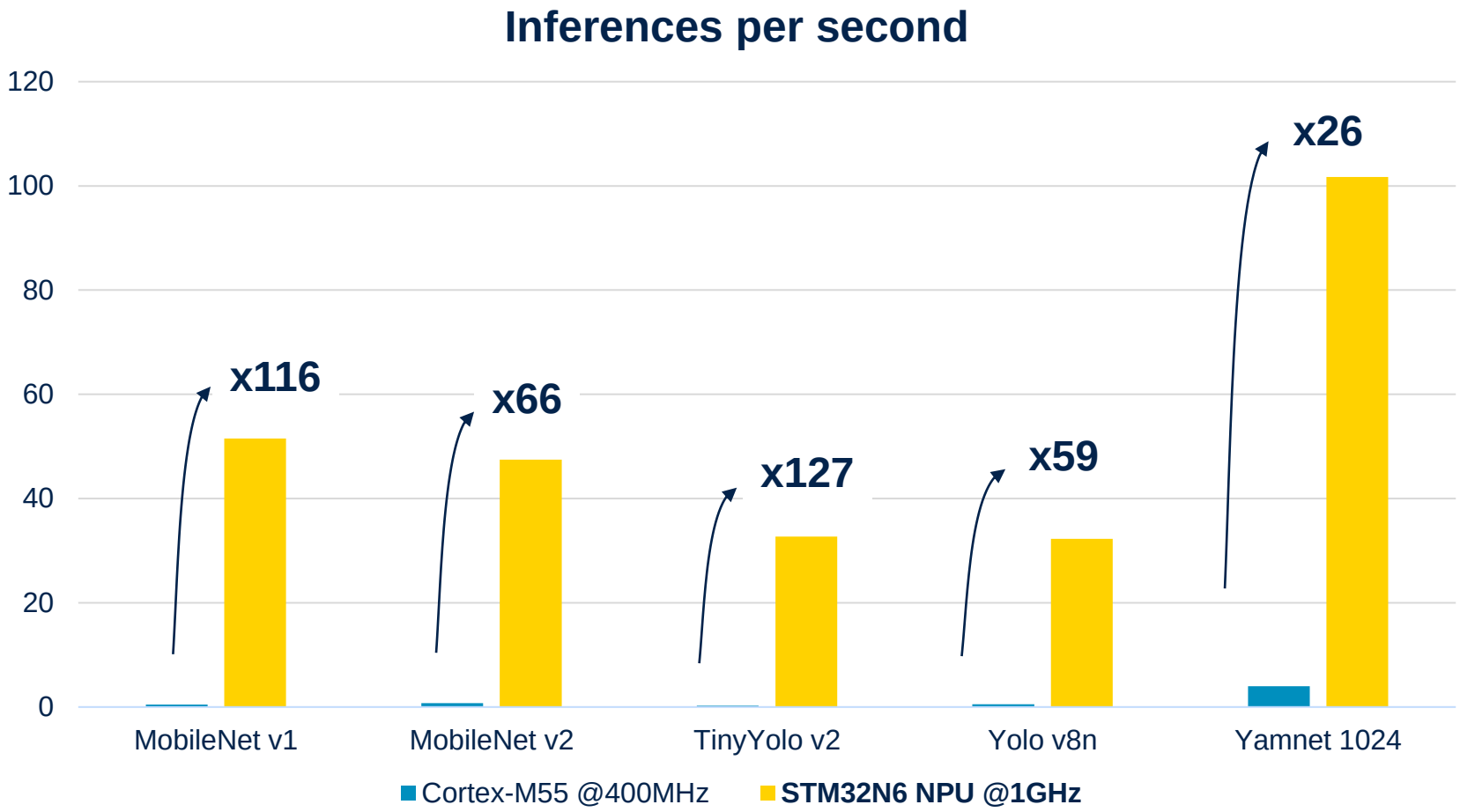
Active tamper protection

Trust anchor
ensuring authenticity & integrity

More insights on the STM32N6 wiki page

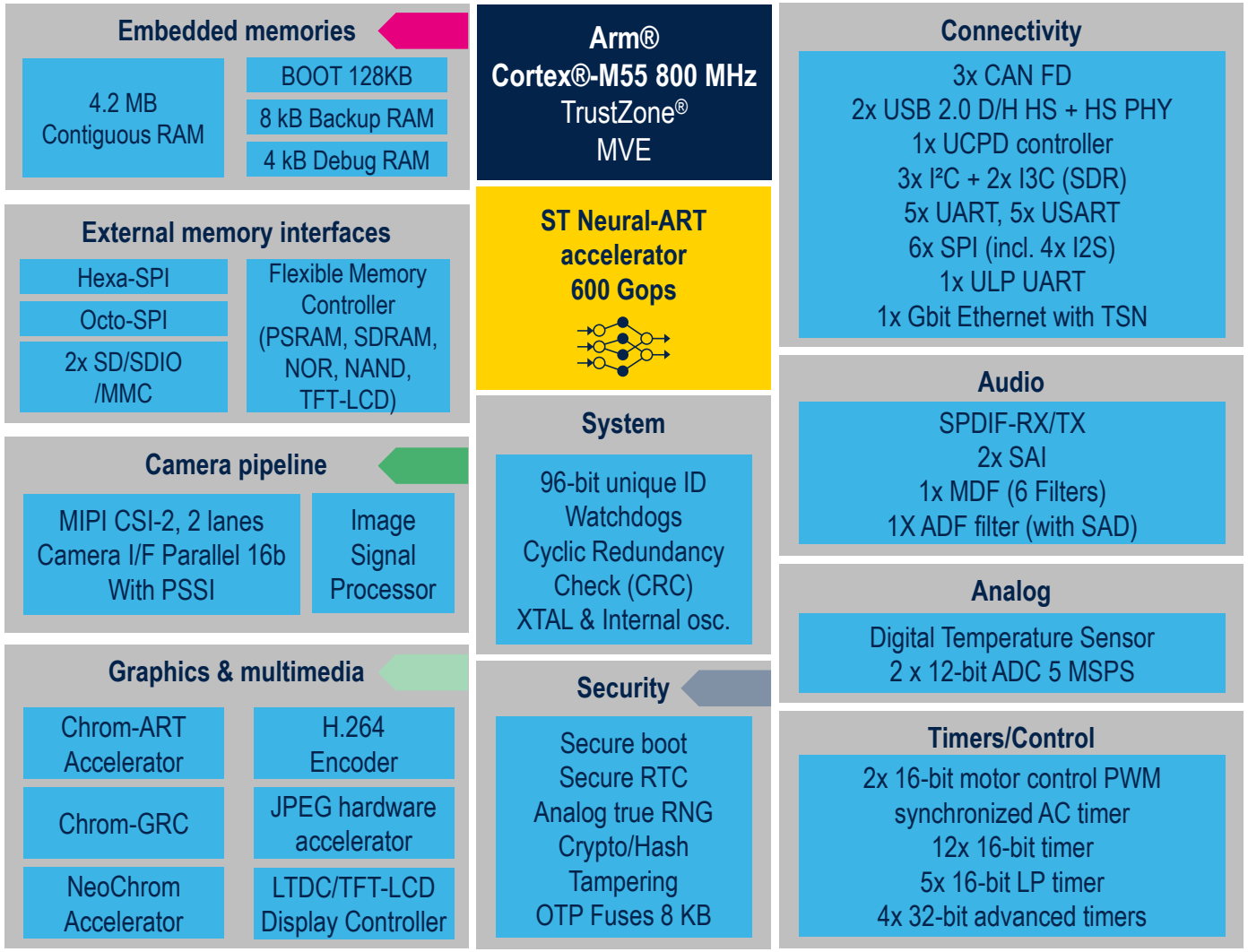
More on STM32Trust security framework.

Neural-ART Accelerator provides a huge performance leap for AI inference



- **MobileNet v1:** image classification
- **MobileNet v2:** image classification
- **TinyYolo v2:** object detection
- **Yolov 8n :** object detection
- **Yamnet 1024:** audio recognition

STM32N6x7 and STM32N6x5 MCUs



Leading edge MCU core

Neural processing unit (STM32N6x7 MCUs only)

Large embedded memory + flexible I/F

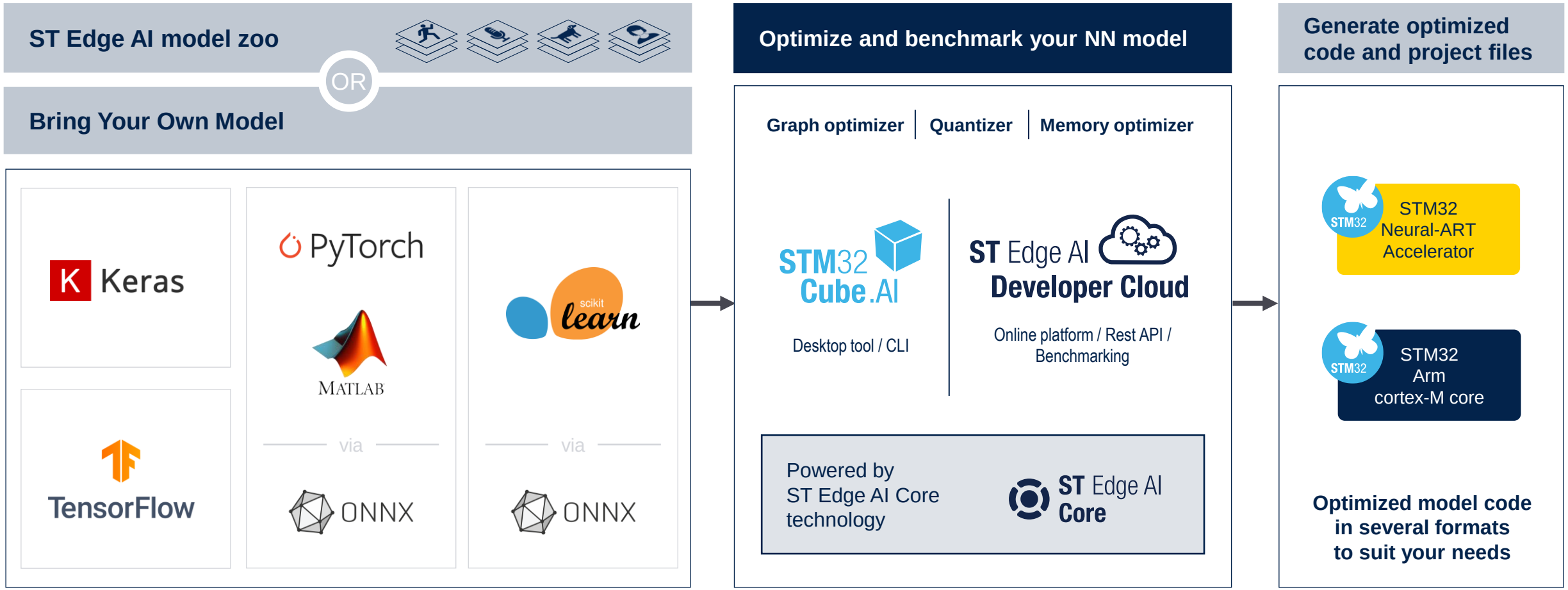
Dedicated camera pipeline

Extended multimedia capabilities

Advanced & certified security

STM32N6 portfolio: one series, two lines

Seamless integration with existing software ecosystem



Embedding innovation across product segments

Drones
Flying & landing

Smart industry
Anomaly detection

Smart homes
Event detection

Smart farming
Animal well-being

Personal healthcare
Body measurements

White goods
Smart control

Automotive
Environment sensing

Smart buildings
Building automation

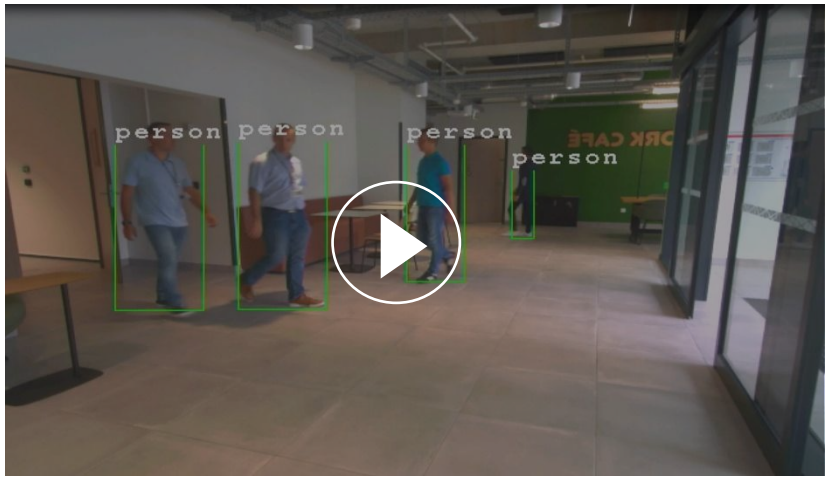
Robots
Collision detection

Personal electronics
Wearables

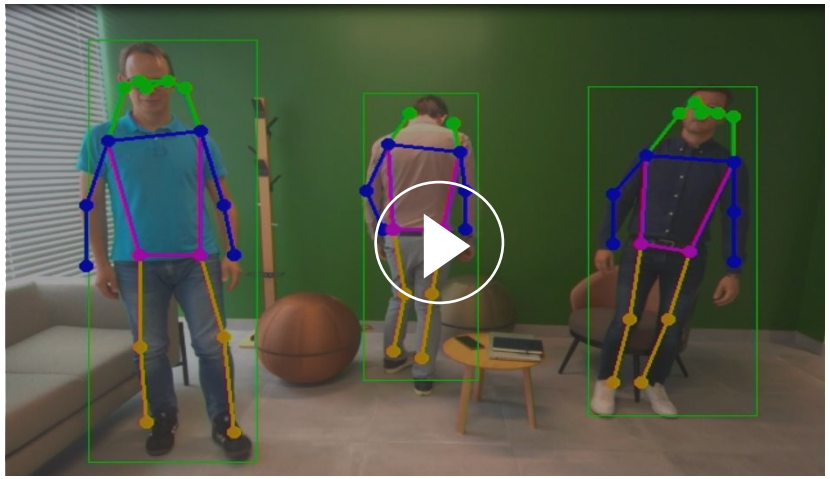
How the STM32N6 changes the game

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

People detection



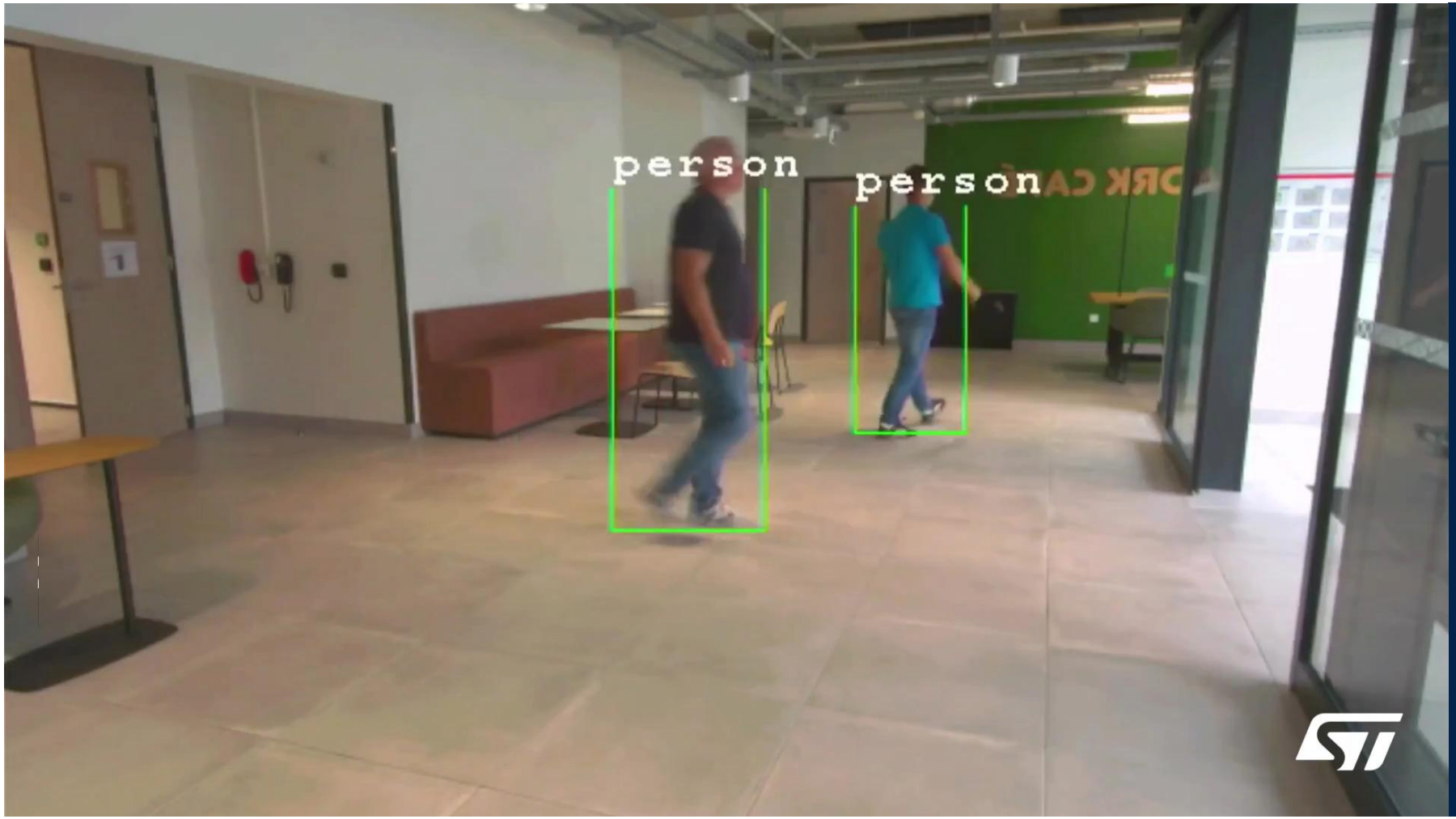
Pose estimation



Hand landmark

Watch all videos on YouTube

High-accuracy people detection at a distance in varied ambient conditions



KEY METRICS

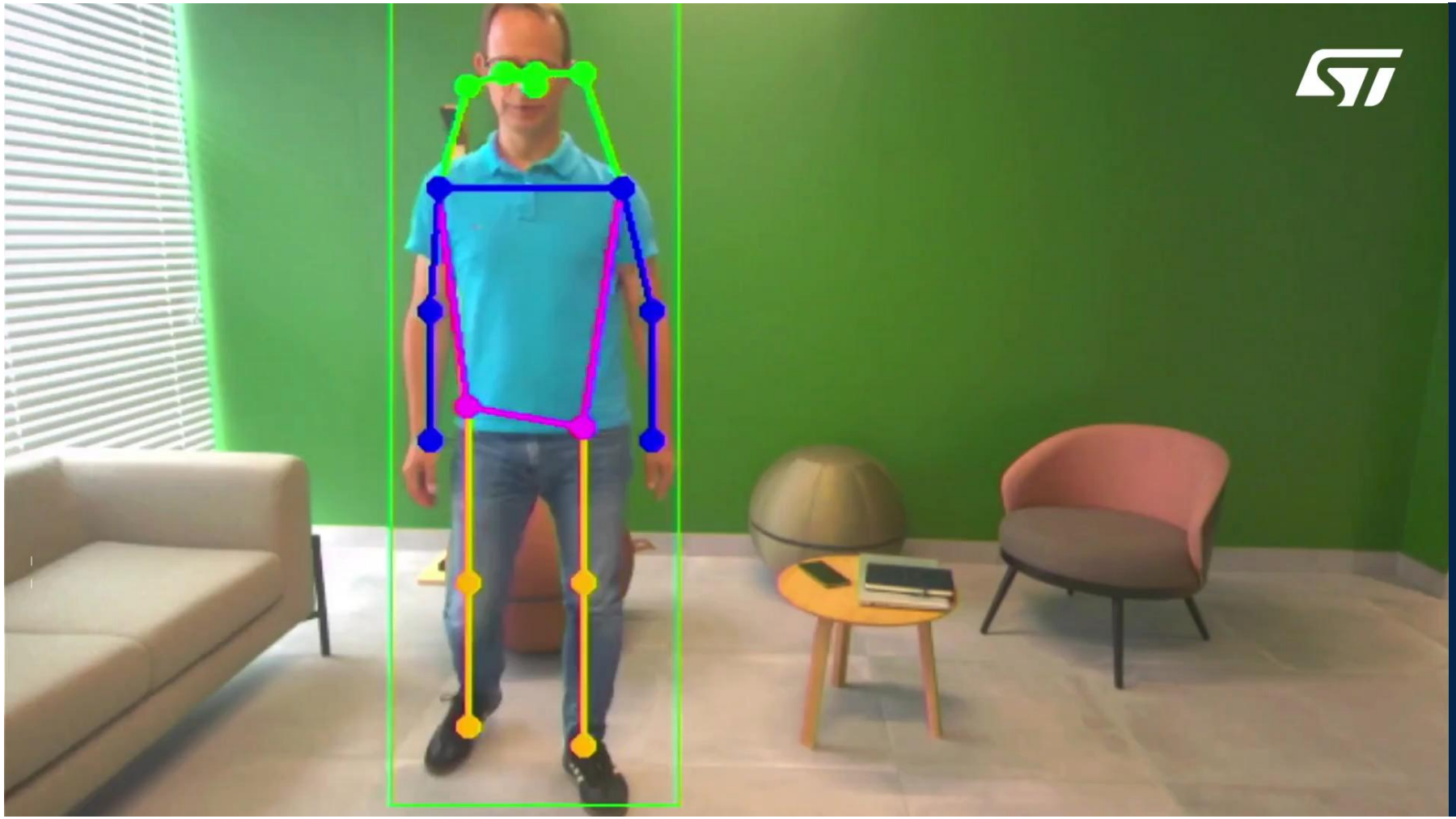
Yolo v8

- 320 x 320
- 26 FPS
- 2.9 MB weights
- 1.6 MB activations

KEY APPLICATIONS

- Smart doorbells
- Room occupancy
- Alarm systems

High-accuracy multipose estimation



KEY METRICS

Yolo v8n

- 256 x 256
- 26 FPS
- 3.35 MB weights
- 2.59 MB activations

KEY APPLICATIONS

- Behavior analysis
- Activity monitoring
- Fall detection

Precise system control with hand landmark

KEY METRICS
(Two models in parallel)

Palm detector
- 192 x 192
- 1.1 MB weights
- 1.1 MB activations

Hand landmark
- 224 x 224
- 3.2 MB weights
- 1 MB activations

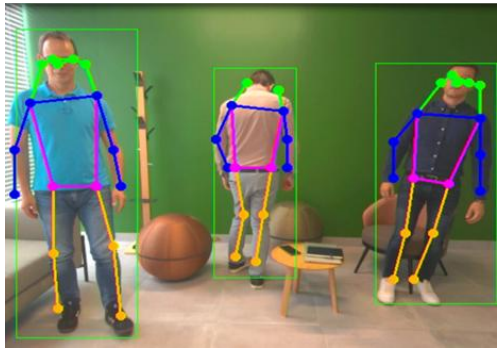
KEY APPLICATIONS

- Touchless HMIs
- Smart appliances
- Smart industry

Get started with edge AI examples

People detection

- Application example showing a people detection use case.
- Demonstrating typical AI computer vision application: camera capture, pre-processing, single model inference and post-processing.
- RTOS-based application example.



Multipose estimation

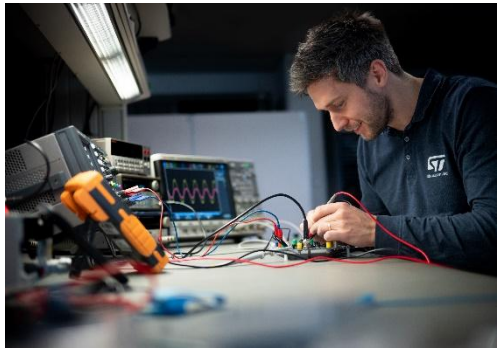
- Application like people detection but built around a multi-pose estimation use case.
- RTOS-based application example.

Hand landmark detection

- Application example showing hand landmark detections.
- Demonstrating the execution of two NN models consecutively.
- RTOS-based application example.

H264 encoding / USB UVC streaming

- Demonstrating a more complete application involving several STM32N6 multimedia features: NPU to perform the inference, H264 encoding and USB video device class stream output data to a PC.
- RTOS-based application example.

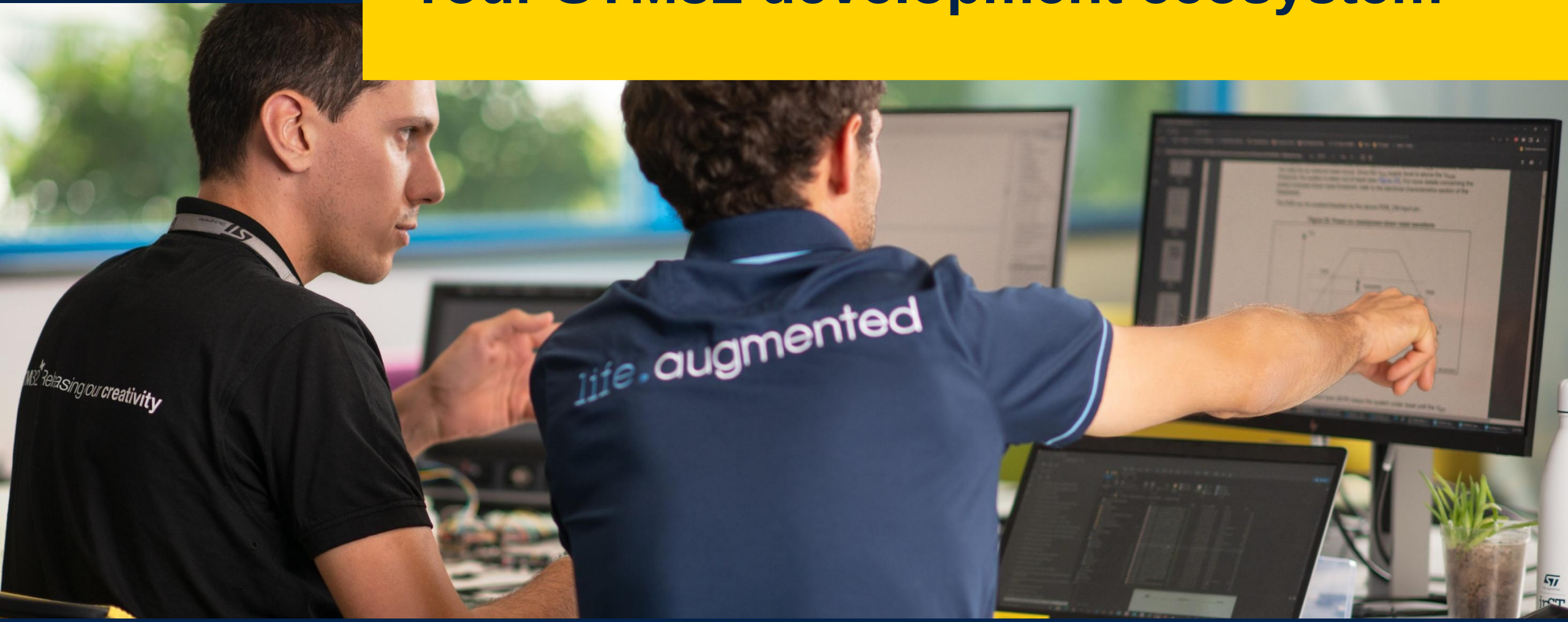


Power measurement

- Demonstrating low power optimizations.
- Enabling easy power measurement on STM32N6 discovery board.
- Bare-metal application example.

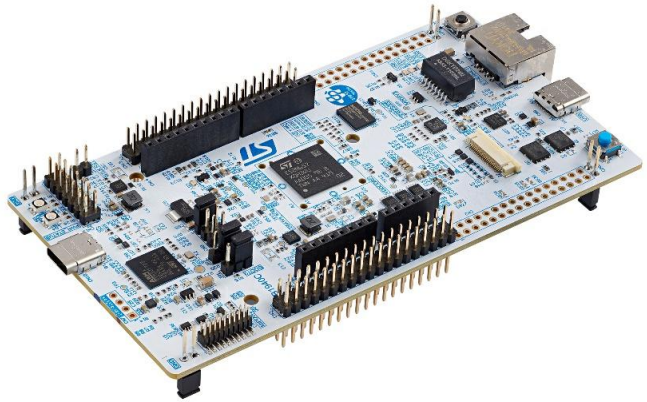
Access the source code here

Your STM32 development ecosystem



Development tools for STM32N6 series

Jump-start your evaluation, prototyping, and design



NUCLEO-N657X0-Q

Affordable prototyping

STLINK v3, ST morpho, ARDUINO®, MIPI CSI-2 connector, USB 2.0, 1GB Ethernet

Camera connector compatible with Raspberry.

STM32N6570-DK

Advanced prototyping including AI

STLINK v3, ST morpho, Arduino®, MIPI connector, USB 2.0, 1 Gbyte Ethernet, 32 Mbytes HexaRAM, Audio Jack, SD card

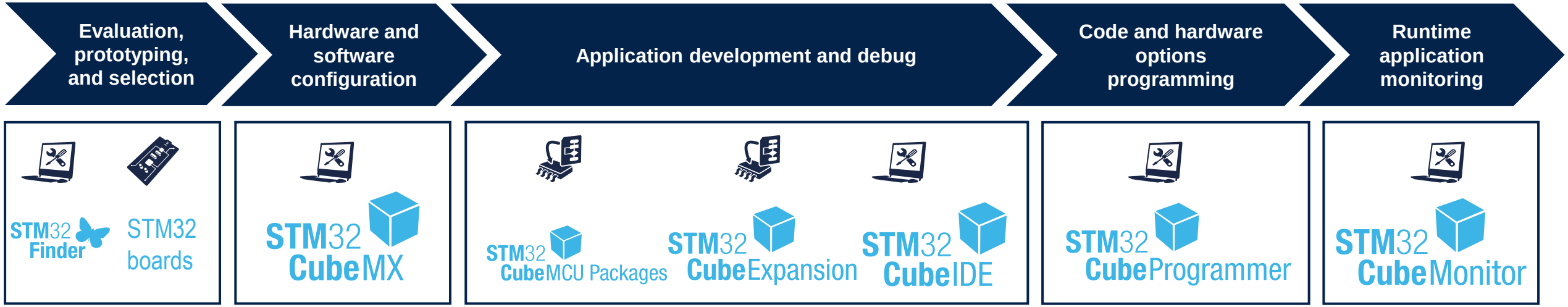
B-CAMS-IMX expansion board

Rolling shutter camera, M12 removable lens, multizone direct Time-of-Flight sensor, inertial motion unit, Raspberry Pi compatible 22-pin connector.

Included in discovery kit.

Leveraging STM32Cube framework

Tools and software supporting you during all your design steps



Complemented with open-source frameworks and partner solutions

FreeRTOS, ThreadX, Zephyr, IAR, Arm Keil, Visual Studio Code

Our technology starts with You

Find out more at www.st.com