

# AI @ The Edge: Vantaggi e Soluzioni

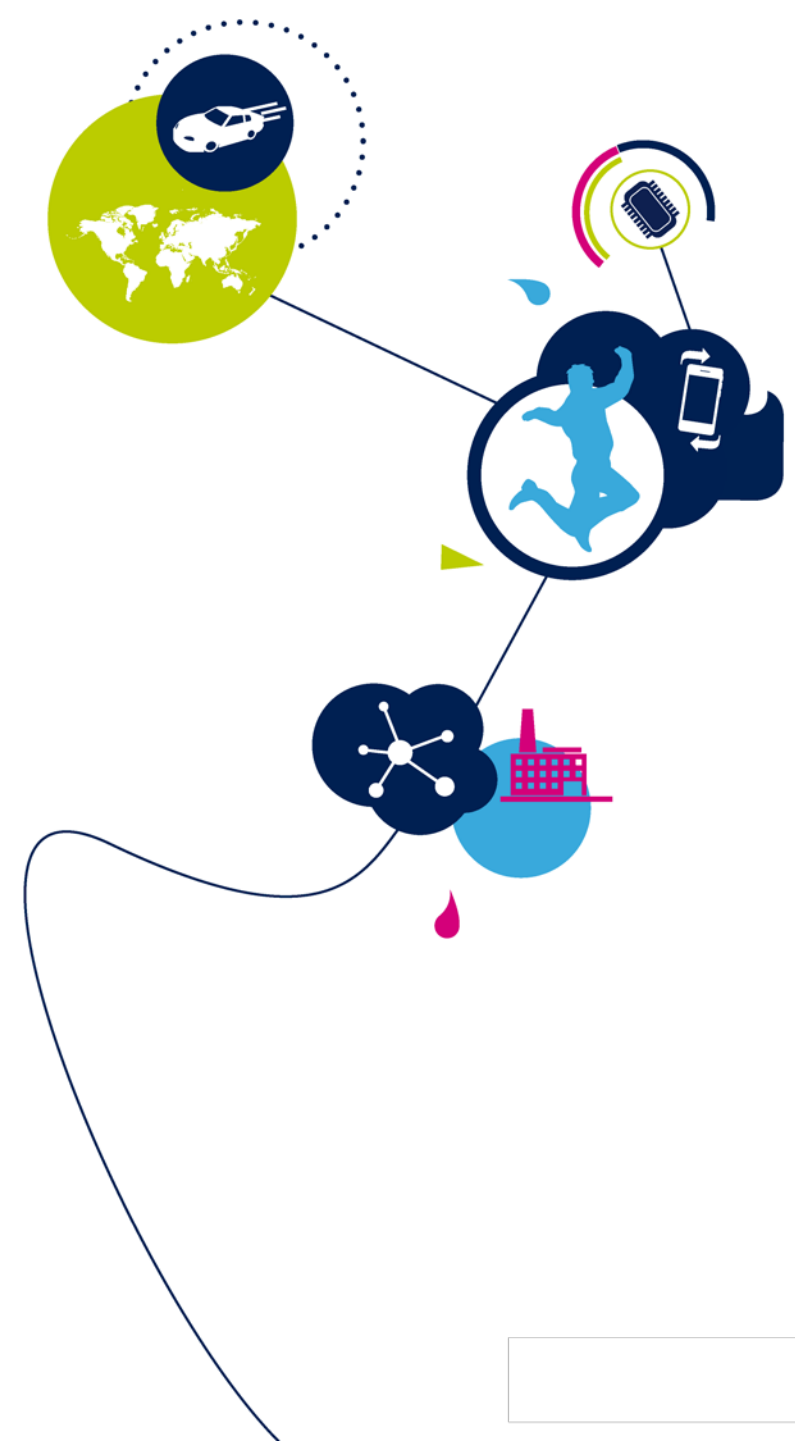
**Viviana D'Alto**

STMicroelectronics

System Research & Applications

Secure Connected Intelligent Systems

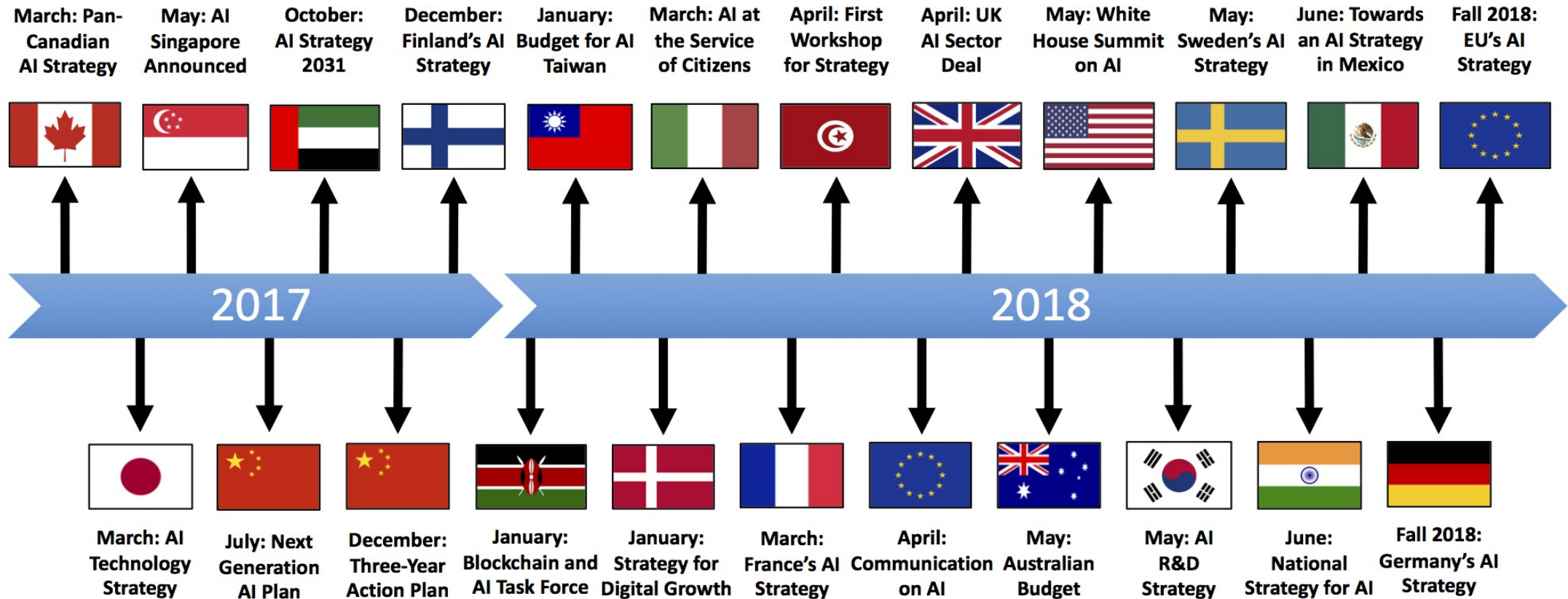
Artificial Intelligence SW & Tools Research Platform Director



# The Race to AI Leadership has begun...

2

## Artificial Intelligence Strategies



# 30 Experts for the Italian AI Strategy

nominated by MISE on Dec.27th, 2018

3



10 experts from the Industry  
(Large companies, start-ups)



10 experts from the R&D  
(University, Think-Thank)



10 experts from the Society  
(Unions, professionals, consumers,  
third sector...)

- Among the world's largest semiconductor companies
- Serving over **100,000** customers across the globe
- 2018 revenues of **\$9.66B**, with year-on-year growth of **15.8%**
- Listed: NYSE, Euronext Paris and Borsa Italiana, Milan
- Signatory of the United Nations Global Compact (UNGC), Member of the Responsible Business Alliance (RBA)



- **~46,000** employees worldwide
- **~ 7,400** people working in R&D
  - **~ 18,000** patents; **~9,600** patent families; **~ 550** new filings (in 2018)
- **11** manufacturing sites
- Over **80** sales & marketing offices





# ST Focus Applications Approach

5



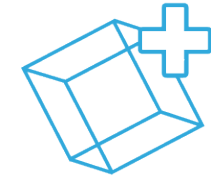
Smart Driving



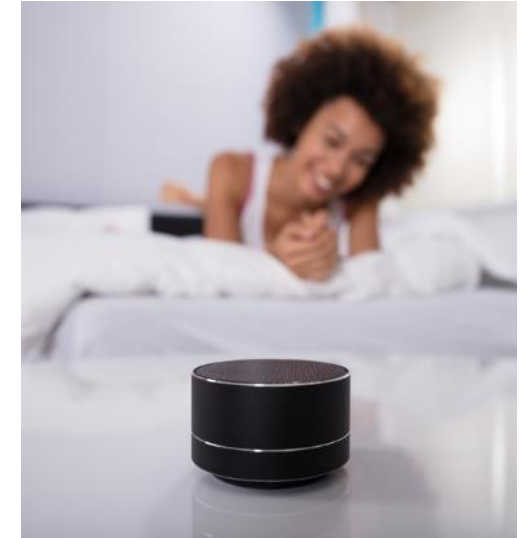
Smart Industry

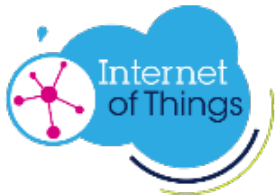


Smart Home & City

















Smart Things





# All IoT Devices have the same needs

6

	Processing & Security	Sensing & Actuating	Connectivity	Signal Conditioning & Protection	Power & Energy Management
 Smart Things	 Ultra-Low Power to High Performance	 Full range of sensors and actuators	 10 cm to 10 km	 Nano Amps to Kilo Amps	 Nano Watt to Mega Watt
 Smart Home	Scalable Security solutions 				
 Smart City					
 Smart Industry					

# Product Portfolio 7

## Processing

New STM32L4 ultra-low-power MCUs  
advanced audio and energy efficiency



STM32F413 MCU  
The new King of STM32F4 Access lines



On-chip USB High-Speed PHY  
expands STM32F7 MCU ecosystem



STM8S general-purpose MCU  
Now available in 8-pin package



Best-in-class infotainment processor  
for all classes of vehicles



## Security

Boost your contactless-payment designs  
with all-in-one ST53G module



STM32™ IoT secure solution  
with STSAFE™ and ProvenCore-M™



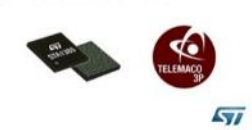
STM32H7 MCUs boost protection  
for smart connected devices



HW and SW for complete drop-in  
solution for IoT security



Advanced secure microprocessors  
protect connected cars



## Sensing & Actuating

Versatile 3-axis accelerometer  
for IoT and wearable devices



New FlightSense™ ranging sensor  
multi-zone, multi-target detection



Smart motion sensors  
for always-on activity tracking



Tiny low-dropout regulator  
ultra-low noise and high-precision sensing



10-bar water-resistant  
MEMS pressure sensor



## Connectivity

Bluetooth® low energy System-on-Chip  
for smartphone-controlled applications



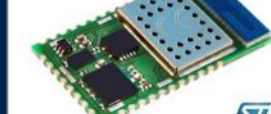
STM32 hardware tools  
boost LoRa® technology



Turnkey PLC chipset  
for smart-energy infrastructure



SPWF04 Wi-Fi module



Sub-1GHz transceiver connects  
Smart Things to the Cloud

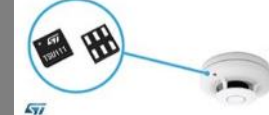


## Conditioning & Protection

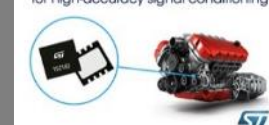
TSX low-power comparators  
better performance and robustness



Tiny nano-power op amp  
Enables longer battery life



3 MHz chopper op amp  
for high-accuracy signal conditioning



USB Type-C™ and PD controllers  
save space and enhance operation safety



Industry-best clamping voltage  
ESD protection in Q201



## Motor Control

Super-Junction MOSFET-based  
600V intelligent power modules



Tiny 2.6A brushed DC motor driver  
for portable, battery-powered devices



Intelligent motion control  
for Smart Industry



MOSFET-based  
SLIMM™-nano modules



Next-gen automotive door-zone controllers  
boost reliability and power efficiency

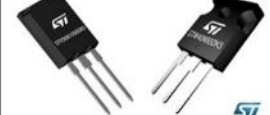


## Power & Energy Management

1200V SiC diodes  
Industrial and automotive-grade



MDmesh™ DK5  
VHV MOSFETs with fast recovery diode



Configurable interleaved PFC controller  
unlocks digital-power advantages



Digital DC-DC multiphase controllers  
for energy-efficient power delivery



Automotive Power MOSFETs in  
PowerFLAT™ 5x6 dual-side cooling





# Artificial Intelligence: The Big Wave

8



## Analysis

### Where am I?

Scene classification (audio, video, environmental sensors)

### Which objects are in the scene, where are they?

Video object detection/classification

### What am I doing?

Activity recognition (audio, video, inertial sensors)

### What's happening?

Event recognition (audio, video, inertial sensors, environmental sensors)



## User Interaction

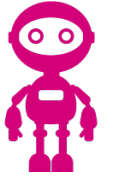
- **Command detection** (audio)
- **Speech Recognition** (audio)
- **Gesture Recognition** (inertial sensors, video)
- **User identification and mood detection** (audio, video)



## Continuous Learning

**How can I detect unpredictable, unclassified events in dynamic environments?**

Recurrent networks (audio, video, inertial sensors, environmental sensors)

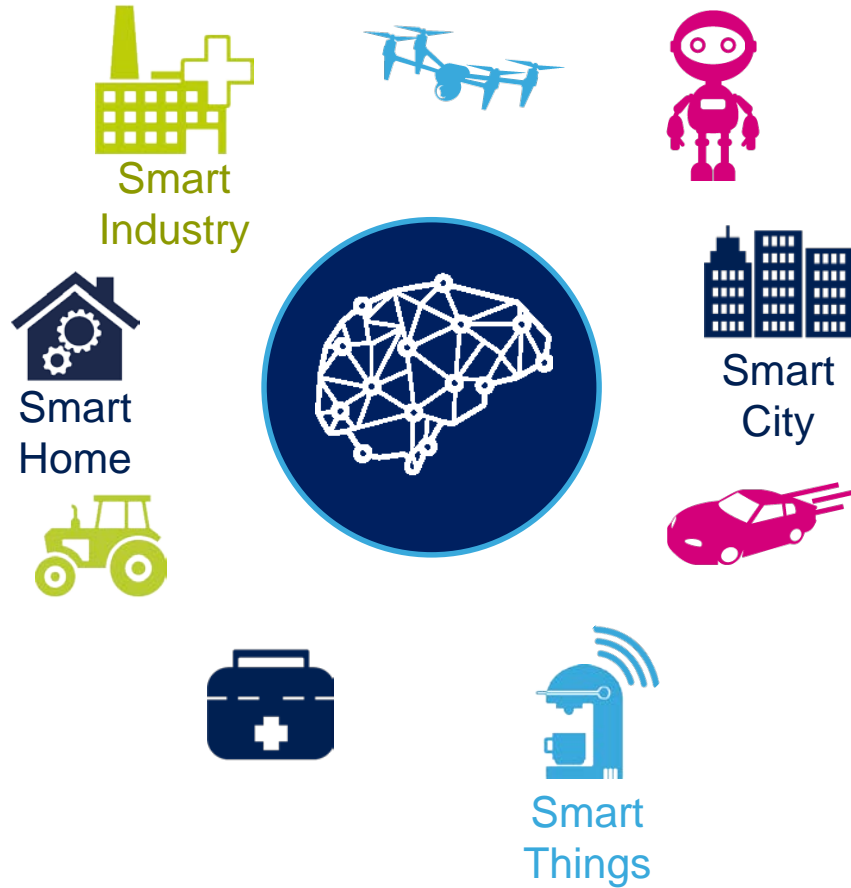




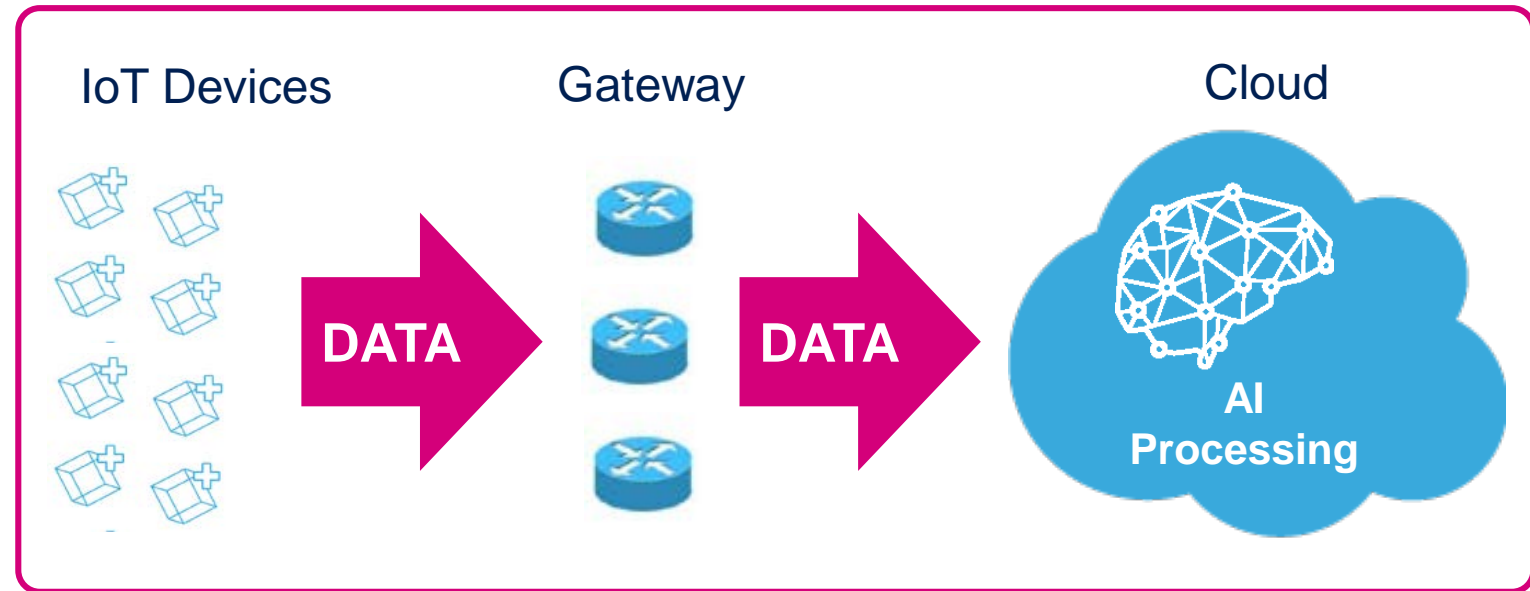


# Artificial Intelligence Enables Huge Opportunities... and Challenges

9



First came the Centralized Artificial Intelligence...

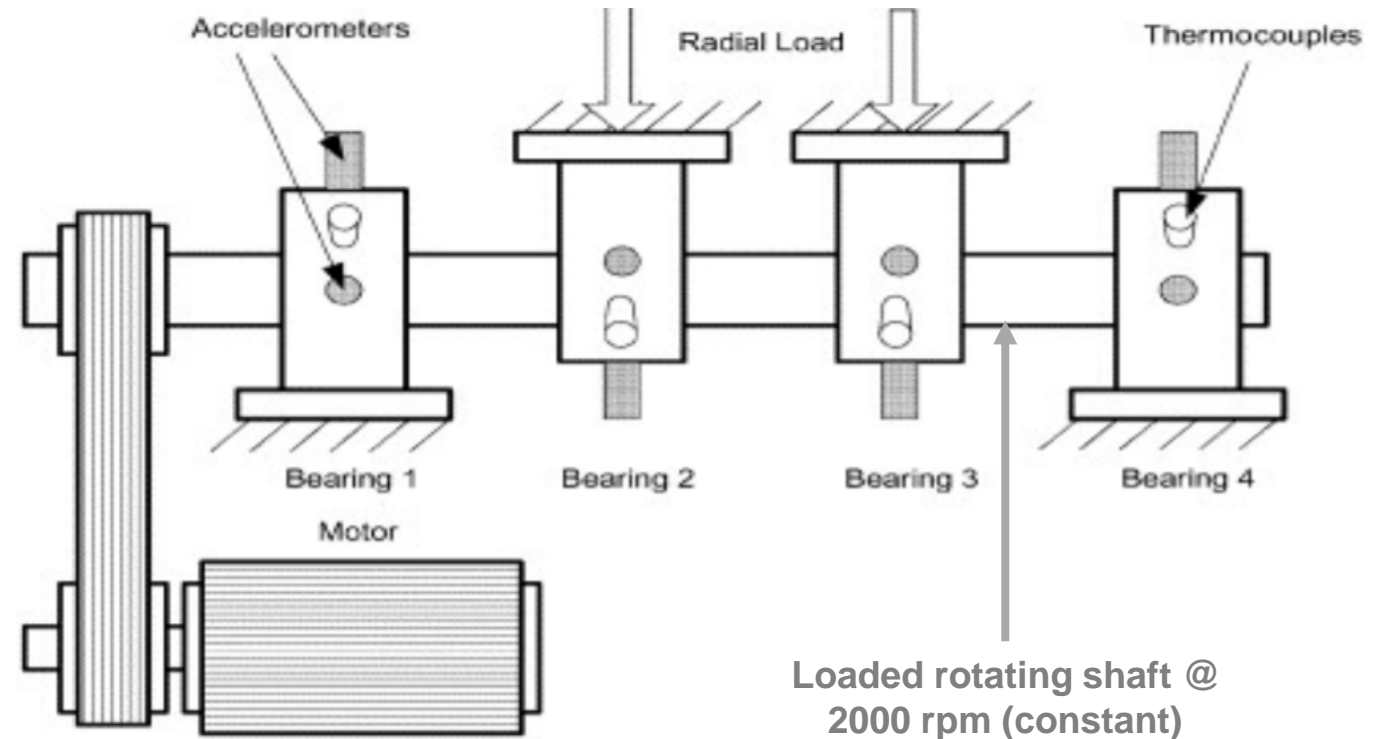


# Data Challenge (1)

## Cloud Predictive Maintenance

10

- 4 Bearings
- 2 Accelerometers (X,Y)/bearing
- 2 Bytes/Axis @ 20KHz
- Total Data Rate: 320 KB/sec, leading to **27.648 GB/day for this single Set-Up in a single equipment of a single Industry,** if in continuous mode



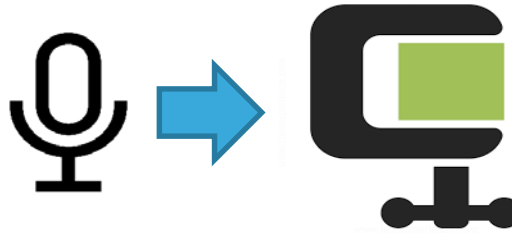
# Data Challenge (2)

## The Cloud Voice Recognition Example

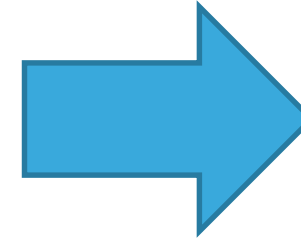
11

### Average Person

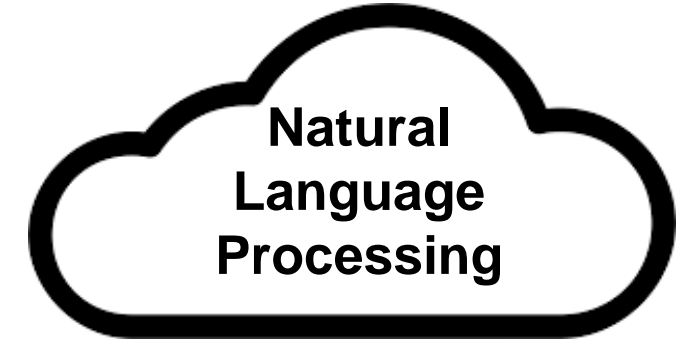
- 16000 utterances<sup>1</sup>/day @ 163 words/minute<sup>2</sup>;
- ≈ 98 minutes speech/day.



Audio coding @  
128Kbps



≈94 MB/day



1 Million People

≈94 TB/day



10 Million  
People

≈1 PB/day



100 Million  
People

≈9 PB/day

**Cloud based Voice  
Recognition means  
huge data bandwidth  
and computation  
capabilities**

# Data Challenge (3)

12

## The Cloud Voice Recognition Example

KLINT FINLEY BUSINESS 06.08.17 12:09 PM

### HOW GOOGLE COPEs WHEN EVEN IT CAN'T AFFORD ENOUGH GEAR

....

But when Dean and Hölzle ran the numbers, they realized that if every Android user in the world used about three minutes of voice recognition time per day, Google would need *twice* as much computing power to handle it all. The world's largest computing infrastructure, in other words, would have to double in size.





# The Sustainability Challenge

13

## 'Tsunami of data' could consume one fifth of global electricity by 2025

Billions of internet-connected devices could produce 3.5% of global emissions within 10 years and 14% by 2040, according to new research, reports **Climate Home News**

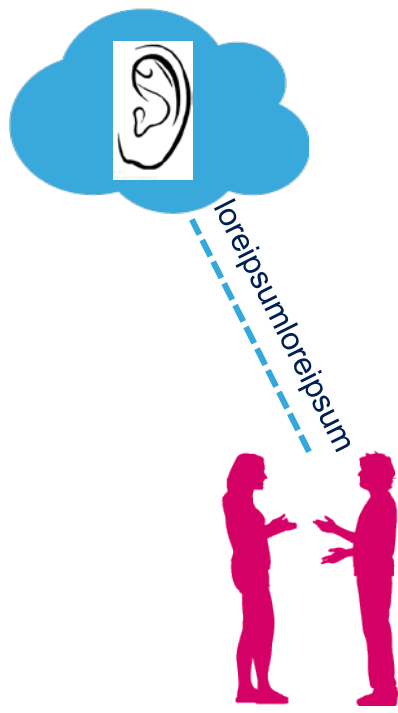


The communications industry could use 20% of all the world's electricity by 2025, hampering attempts to meet climate change targets and straining grids as demand by power-hungry server farms storing digital data from billions of smartphones, tablets and internet-connected devices grows exponentially.

He expects industry power demand to increase from 200-300 terawatt hours (TWh) of electricity a year now, to 1,200 or even 3,000TWh by 2025. Data centres on their own could produce 1.9 gigatonnes (Gt) (or 3.2% of the global total) of carbon emissions, he says.



A 2016 **Berkeley laboratory report** for the US government estimated the country's data centres, which held about 350m terabytes of data in 2015, could together need over 100TWh of electricity a year by 2020. This is the equivalent of about 10 large nuclear power stations.



# The Privacy Challenge

14

PRIVACY ON THE CLOUD

## An Oregon family's encounter with Amazon Alexa exposes the privacy problem of smart home devices

By Youyou Zhou • May 25, 2018

Here's the latest nightmare scenario for the tech-phobic: A woman in Portland, Oregon [found out](#) that her family's home digital assistant, Amazon's Alexa, had recorded a conversation between her and her husband without their permission or awareness, and sent the audio recording to a random person on their contacts list.

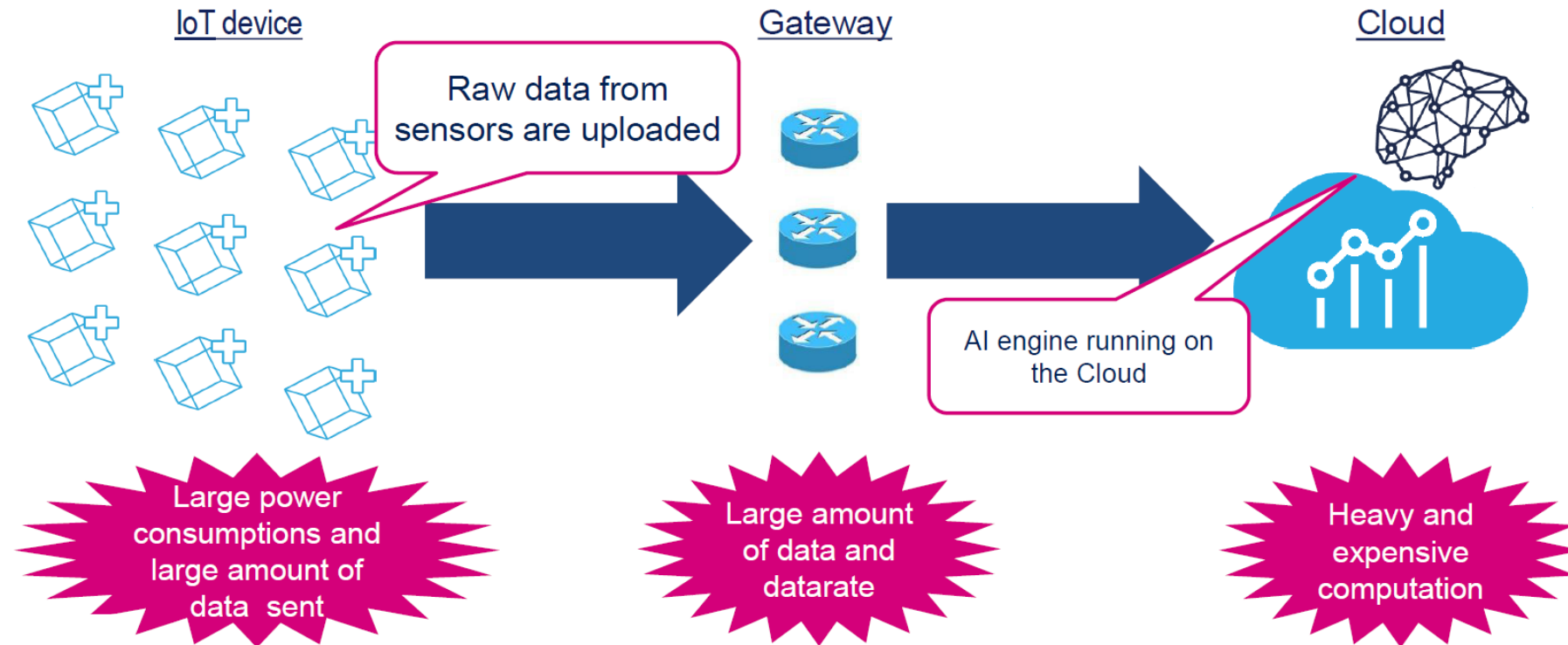
Echo woke up due to a word in [background conversation](#) sounding like "Alexa." Then, the subsequent conversation was heard as a "[send message](#)" request. At which point, Alexa said out loud "To whom?" At which point, the background conversation [was interpreted as a name in the customers contact list](#). Alexa then asked out loud, "[contact name], right?" Alexa then interpreted background conversation as "right". As unlikely as this string of events is, we are evaluating options to make this case even less likely."



# From Centralized AI Approach....

15

- ↓ Scalability
- ↑ Latency
- ↑ Bandwidth
- ↓ Node intelligence
- ↓ Privacy
- ↓ Security
- ↓ Energy Saving

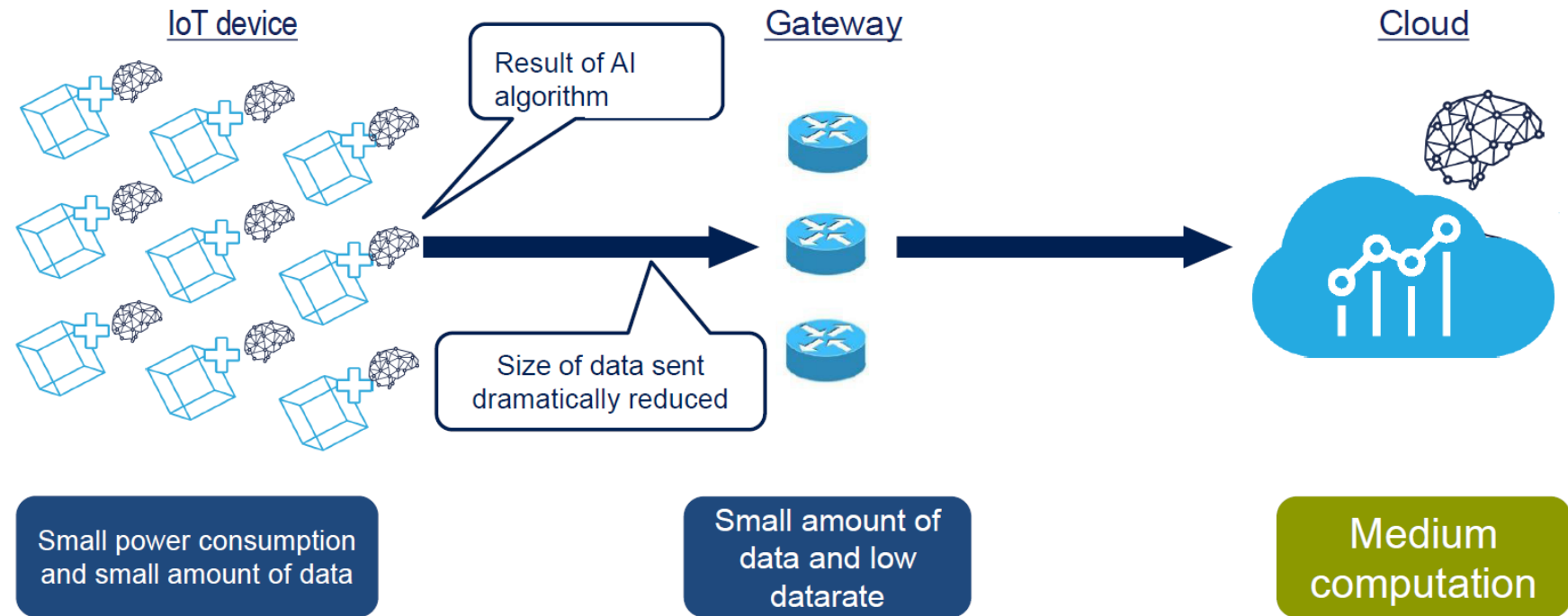




# ...to Distributed AI and AI@ the Edge/Node

16

- ↑ Scalability
- ↓ Latency
- ↓ Bandwidth
- ↑ Node Intelligence
- ↑ Privacy
- ↑ Security
- ↑ Energy Saving
- ↑ Data Value
















# AI @ the Edge Flavours

17

Low	Medium	High
   <ul style="list-style-type: none"><li>• Sensor analysis</li><li>• Activity recognition (motion sensors)</li><li>• Stress analysis or attention analysis</li></ul>	   <ul style="list-style-type: none"><li>• Audio &amp; sound</li><li>• Speech Recognition</li><li>• Object detection</li></ul>	   <ul style="list-style-type: none"><li>• Objects detection / classification / tracking</li><li>• Natural Language Understanding / Speech Synthesis</li></ul>

10s MOPs

GOPs

0.5-1 TOPs

1-2 TOPs

MCU

From IP embedded in MCU/MPU to dedicated SOC

Sagemcom

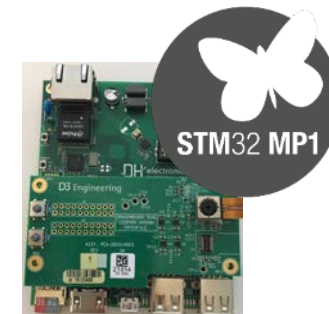
Industrial Maintenance Module



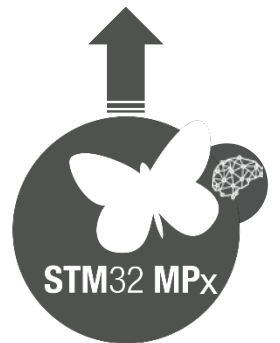
nnf  
no new folk studio



OpenMV Machine Vision Camera



Arrow Avenger96 Board + Camera Module



ST  
life.augmented

# The Key Steps Behind Neural Networks

18



Neural Network (NN) Model Creation



Operating Mode

Capture data



1

2



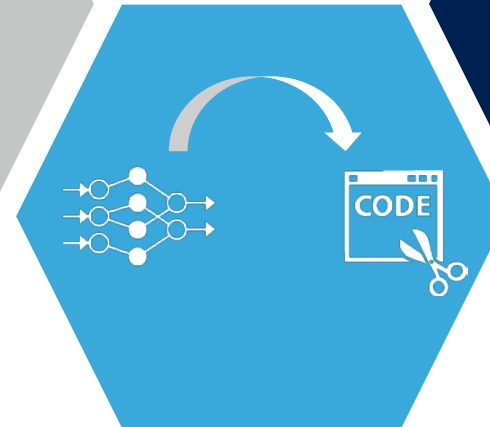
Clean, label Data  
Build NN topology

Train NN Model



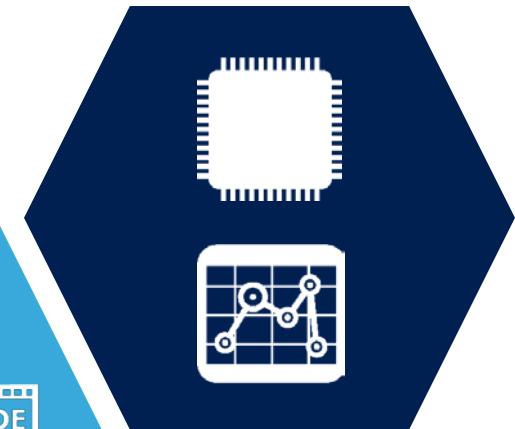
3

4



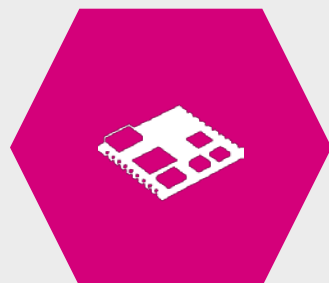
Convert NN into  
optimized code for MCU

Process & analyze new  
data using trained NN



5

# Capture and Label Data to Train NN



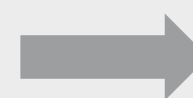
Embedded motion



Labelling controlled  
by smartphone application

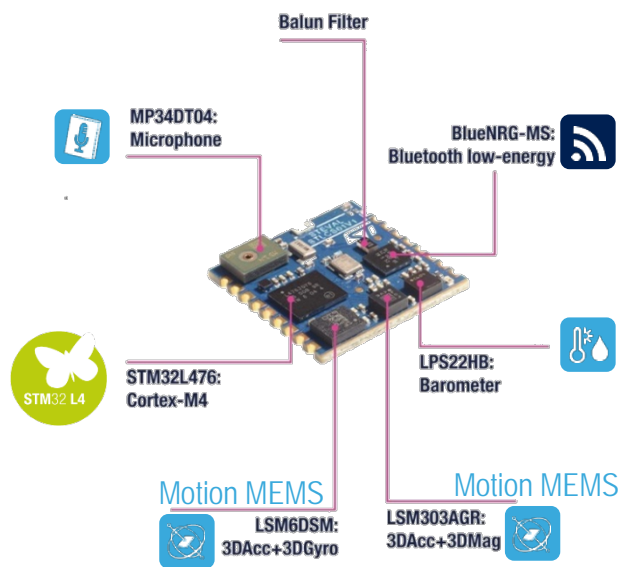


Data stored on the device  
SD card for future learning



5 classes

Stationary, walking, running,  
biking, driving



# The Key Steps Behind Neural Networks

20



Neural Network (NN) Model Creation



Operating Mode

Capture data



1

2



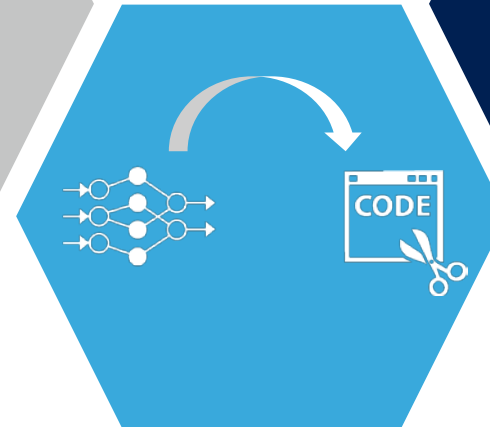
Clean, label Data  
Build NN topology

Train NN Model



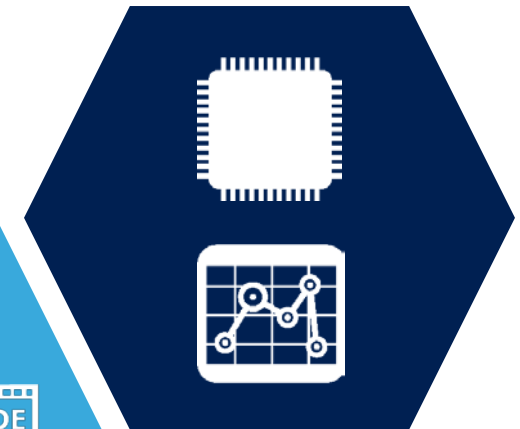
3

4



Convert NN into  
optimized code for MCU

Process & analyze new  
data using trained NN



5



# Artificial Intelligence with STM32

## STM32Cube.AI

21

STM32Cube.AI SW tool allows our customers to innovate...

Define & Train  
NN Model



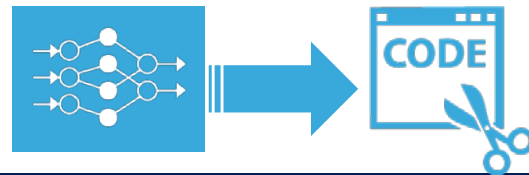
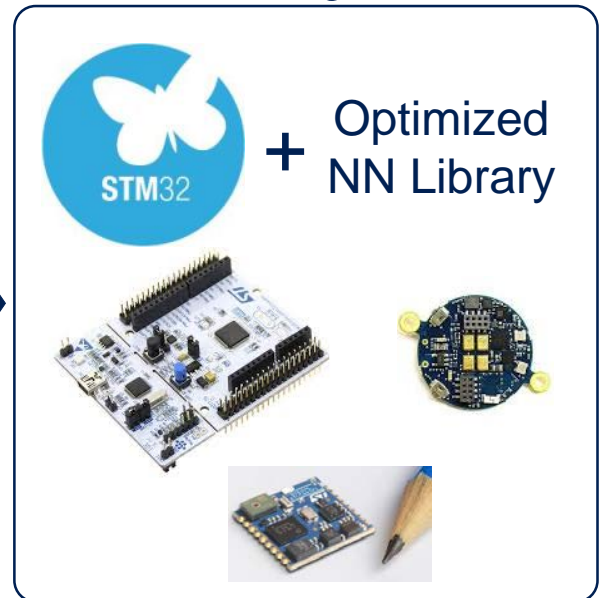
Pre-trained  
Neural Network  
Model

Convert NN into  
optimized code for MCU

STM32  
Cube.AI

Automatically  
generated,  
optimized  
Neural Network  
Library for  
STM32

Process & analyze  
new data using trained NN



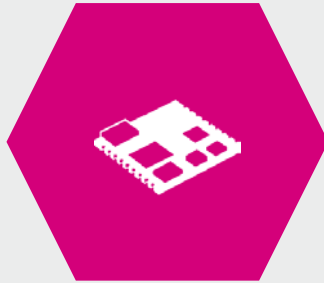
... bringing AI into the STM32 Portfolio



# Embedded Neural Networks

## The “AI-sensing-1 Function Pack” Example

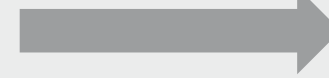
22



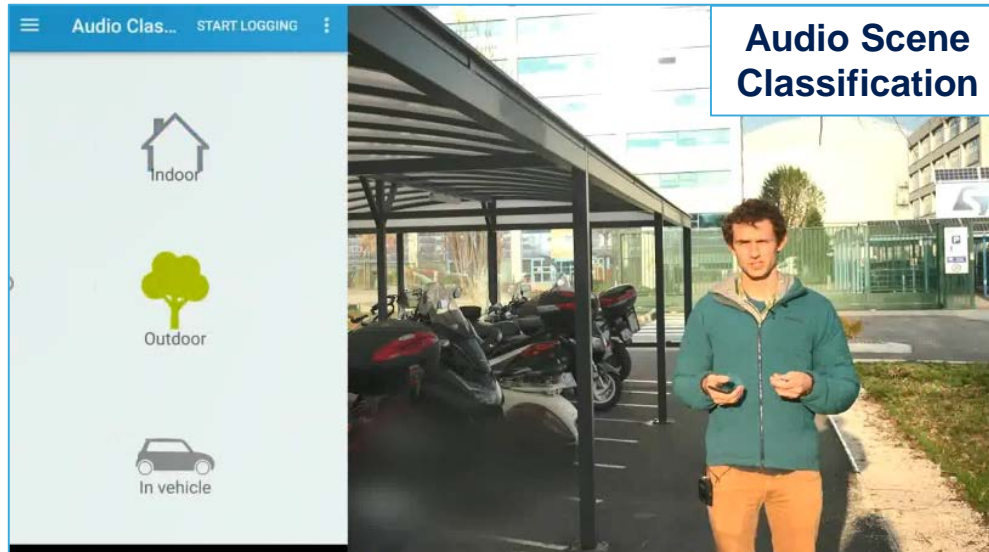
Embedded motion  
pre-processing



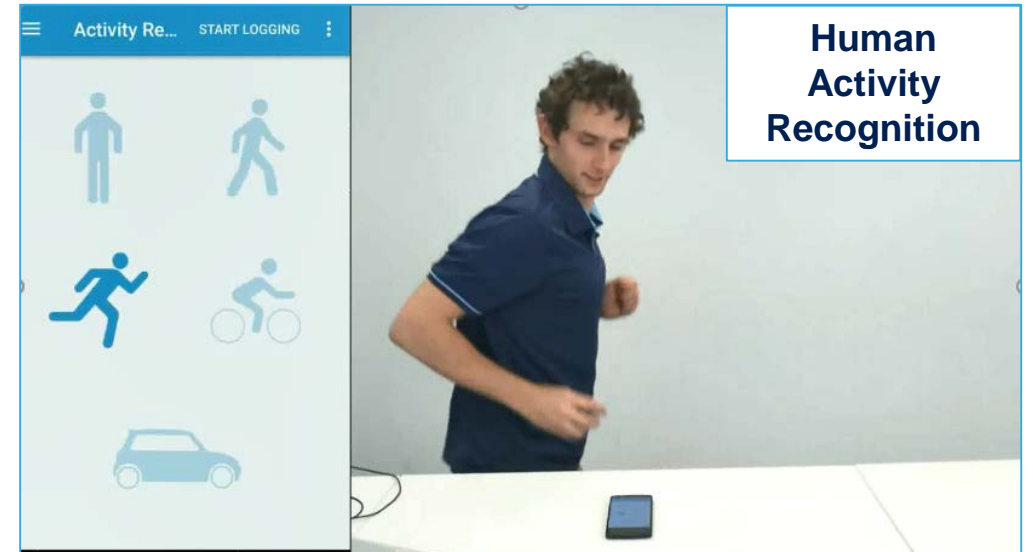
NN & example  
dataset provided



Inference result  
displayed on mobile app



Audio Scene  
Classification



Human  
Activity  
Recognition

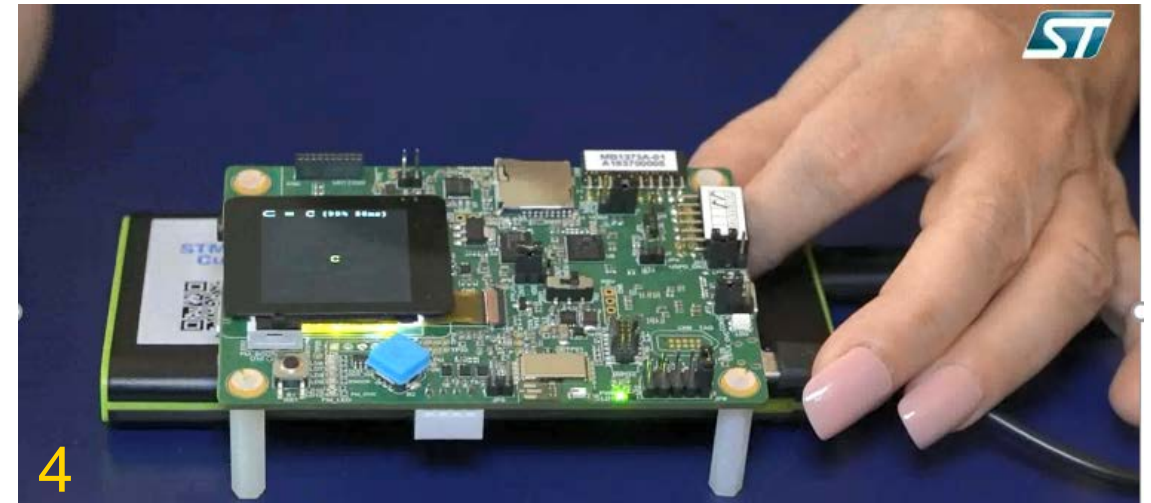
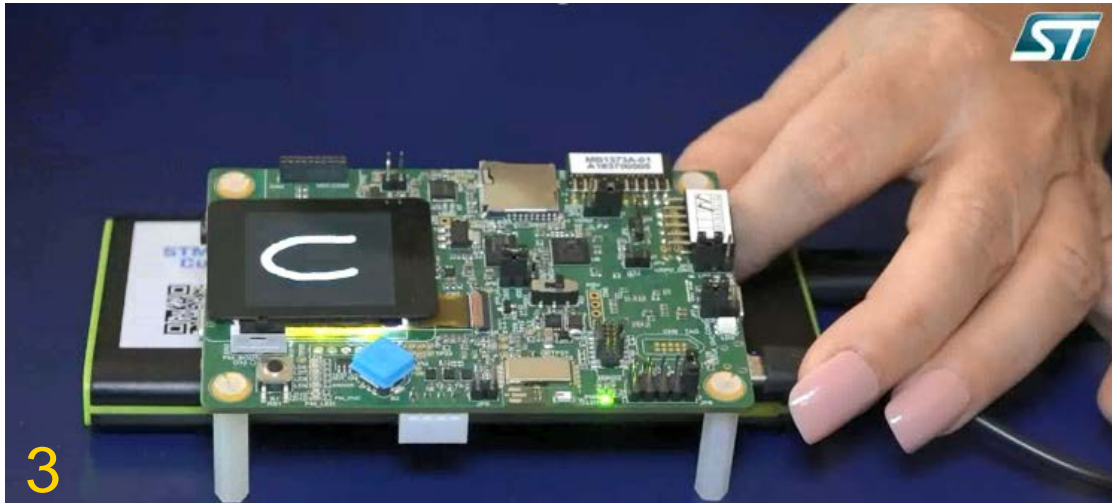
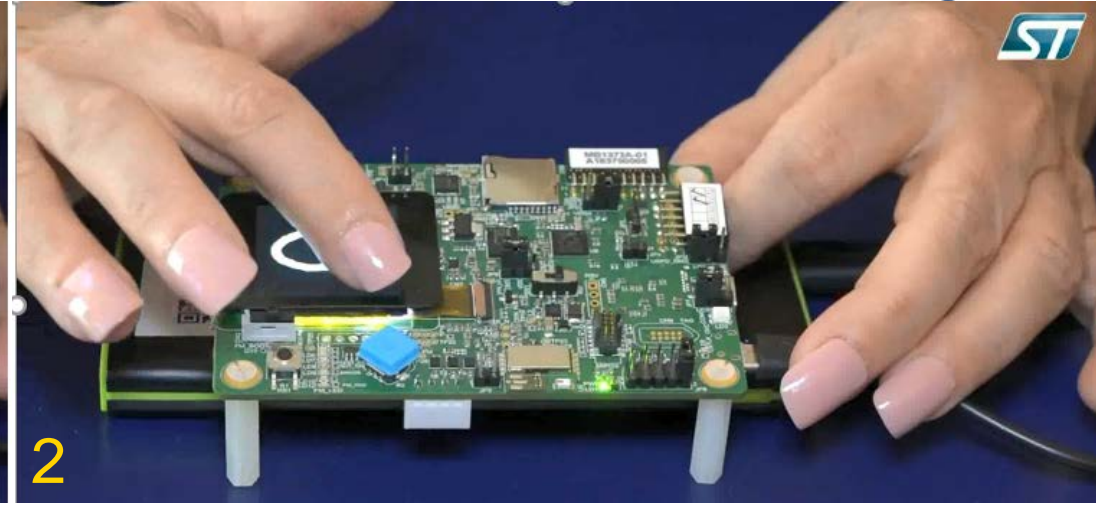
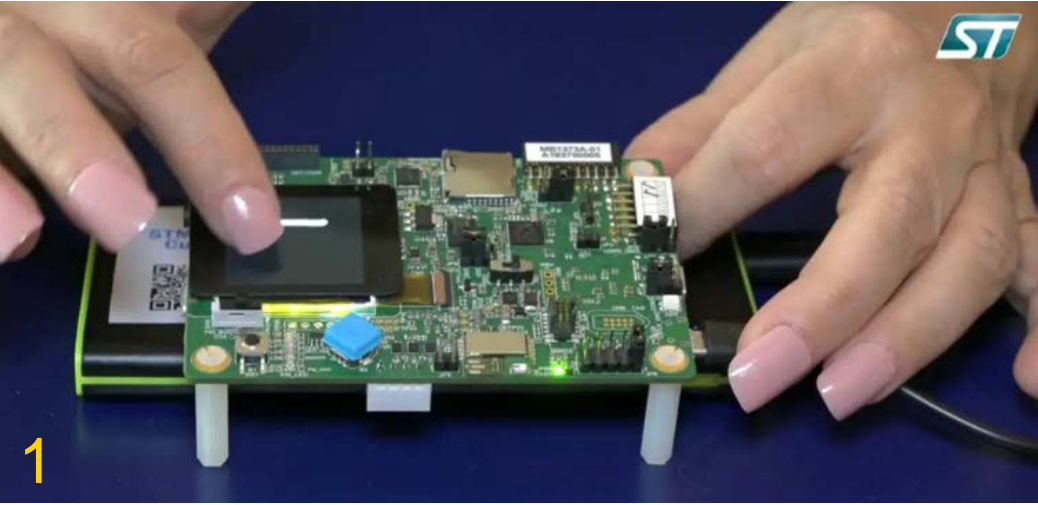




# Embedded Neural Networks (2)

## The “Handwritten Character Recognition”

23



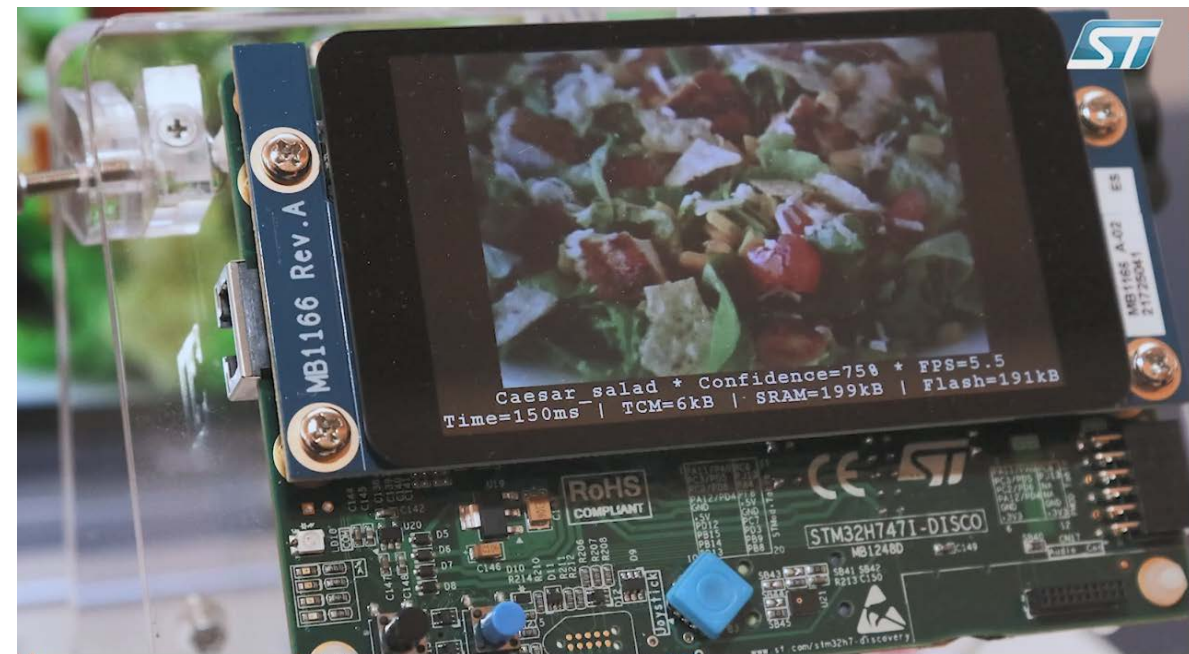
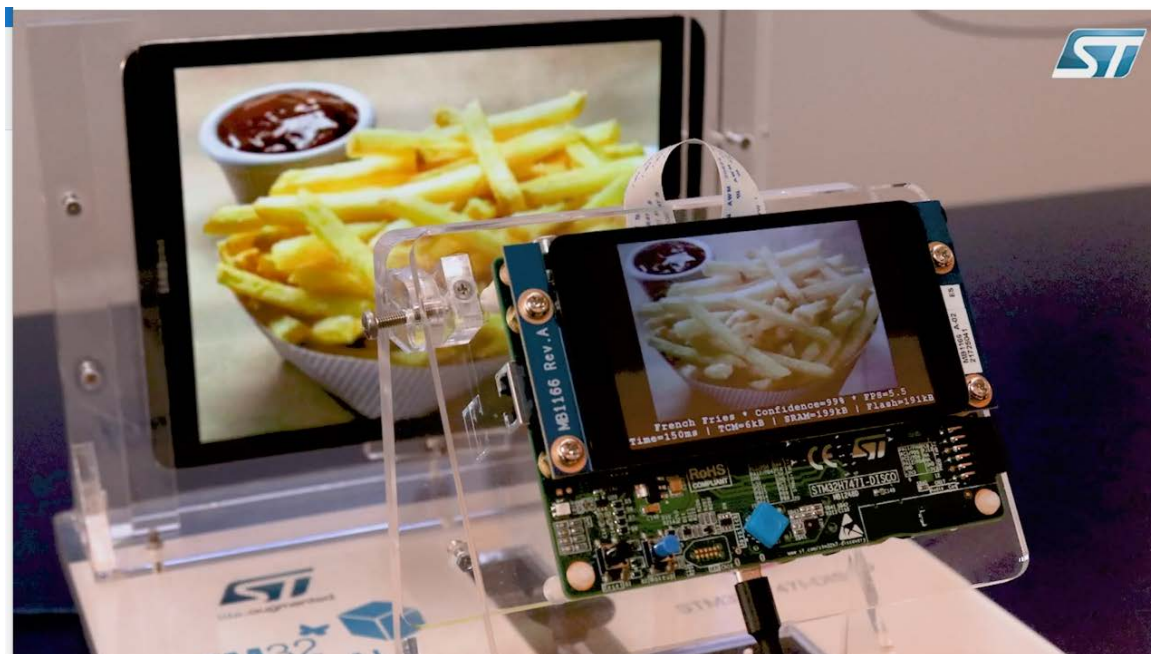




# Embedded Neural Networks (3)

## The “Food Classification” (18 Classes)

24



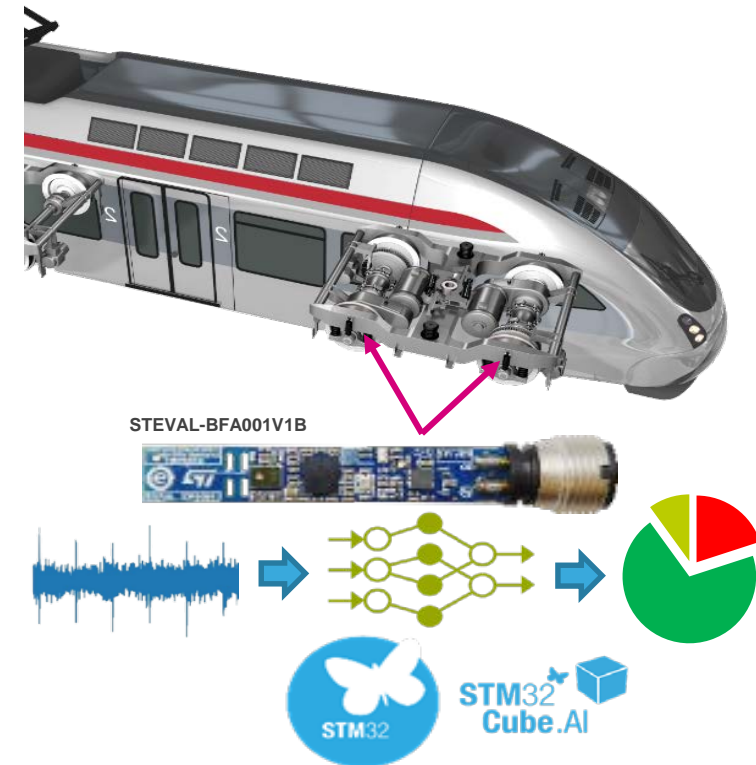




# AI@The Node for Immediate Alerts

25

- AI @the Node allows data analysis close to the sensors sending alerts for critical situations more promptly than centralized/cloud-based solutions
- ST integrated multi-sensors unit STEVAL-BFA001V1B placed close to the ball bearing of a train wheel can detect anomalous vibrations and promptly signal alerts thanks to Artificial Neural Networks implemented on the STM32 MCU present on the board itself
- An automatic tool to optimize and map Neural Networks on STM32 Microcontrollers as STM32Cube.AI simplifies and speeds-up the implementation of AI on Node of sensors
- The Project is a cooperation between STMicroelectronics and Lenord+Bauer



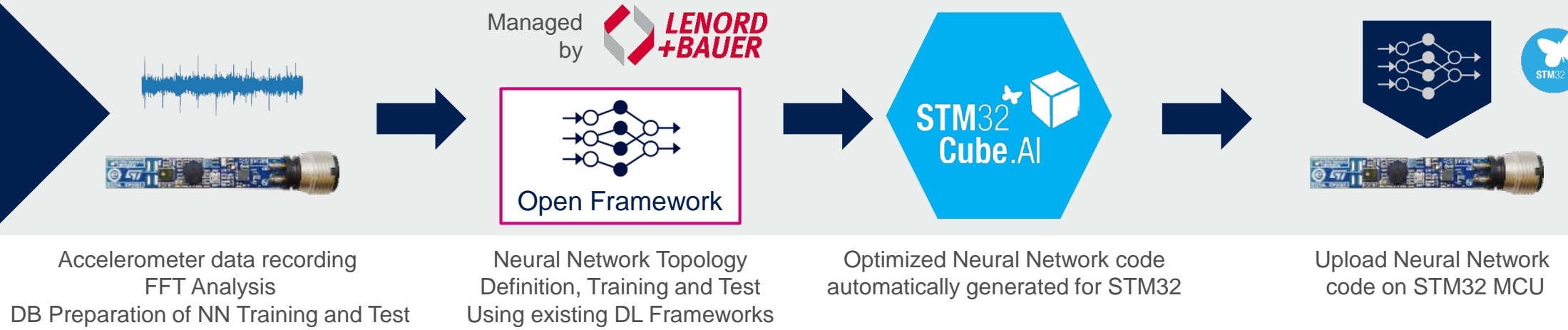


# Data Analysis with AI

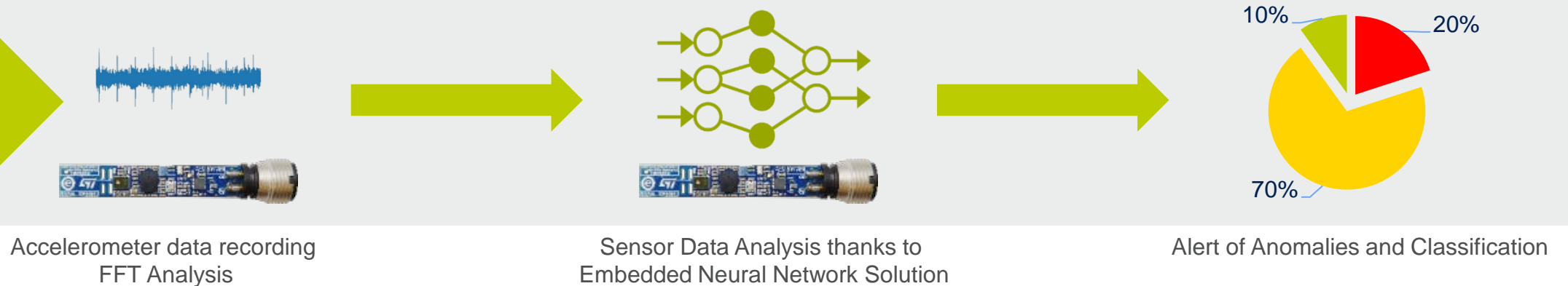
## Training and Embedding a Neural Network

26

### Training



### Operating



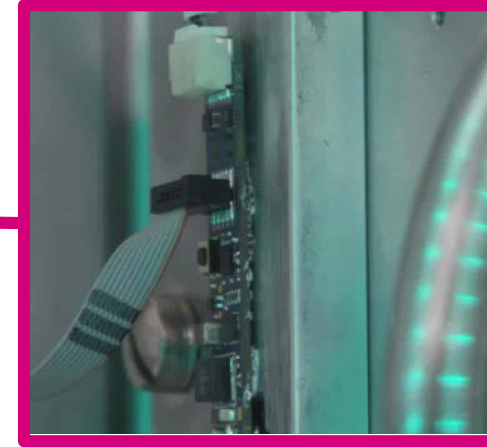
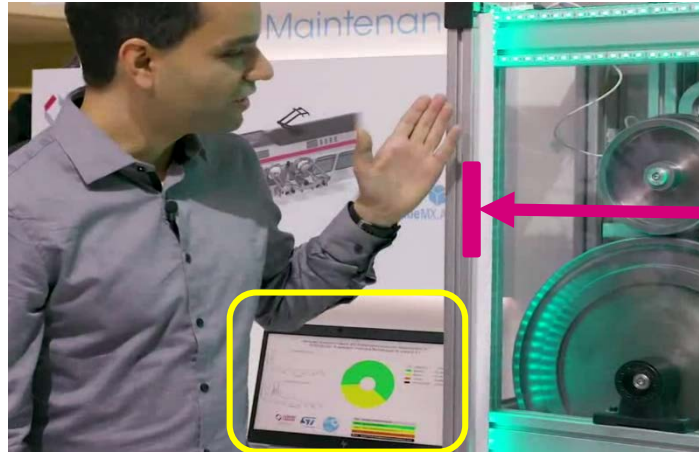


# AI@The Node for Immediate Alerts

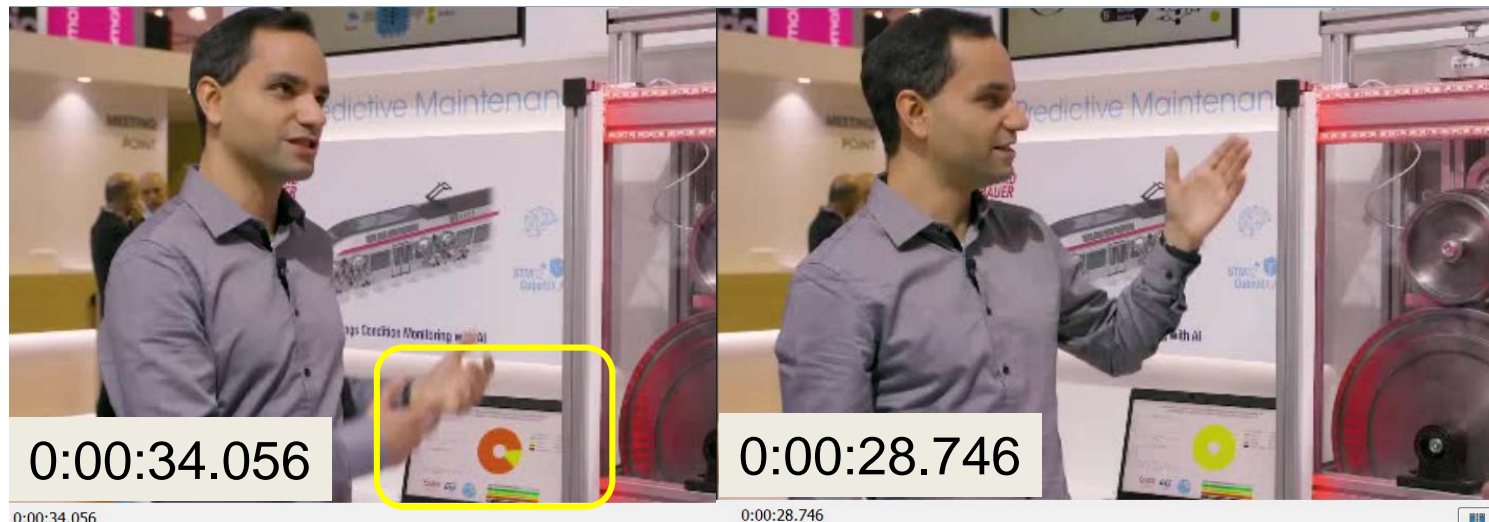
27

2. This system reproduces a train wheel

2. The Laptop gets data from STEVAL-BFA001V1B sensors and analyses them with a neural network in remote.



3. The same sensors data are processed on the STM32 MCU on board the STEVAL-BFA001V1B by the same ANN used in the laptop (i.e. AI @ the Node).



4. An anomalous vibration is detected and classified as critical by the Neural Network on STM32 (red LED lights in the “train wheel” structure) several seconds in advance than the detection by the NN running on PC (red graph on laptop display)

- **Artificial Intelligence is a breakthrough innovation**, but we need to preserve our privacy, grant high security standards, define scalable, sustainable and ethic solutions, while leading to economic benefits
- **A Distributed Intelligence approach increases AI robustness, scalability, responsiveness, security**, by implementing AI solutions as close as possible to Sensors recording data in our environment, home, city, industry, car and personal devices
- **AI @ The Node is a fundamental ingredient** to reach the above targets, and ST is providing solutions to be included in several application scenarios



life.augmented