tinyML® Talks

Enabling Ultra-low Power Machine Learning at the Edge

“Standardized AI Architectures for Secure TinyML”

Andrea Basso – Research director, Synesthesia

June 20, 2023

www.tinyML.org

TINY ML TALKS webcast

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

TINY ML TALKS webcast

Executive Strategic Partners

TINY ML TALKS webcast

EDGE IMPULSE

The Leading Development Platform for Edge ML

edgeimpulse.com

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

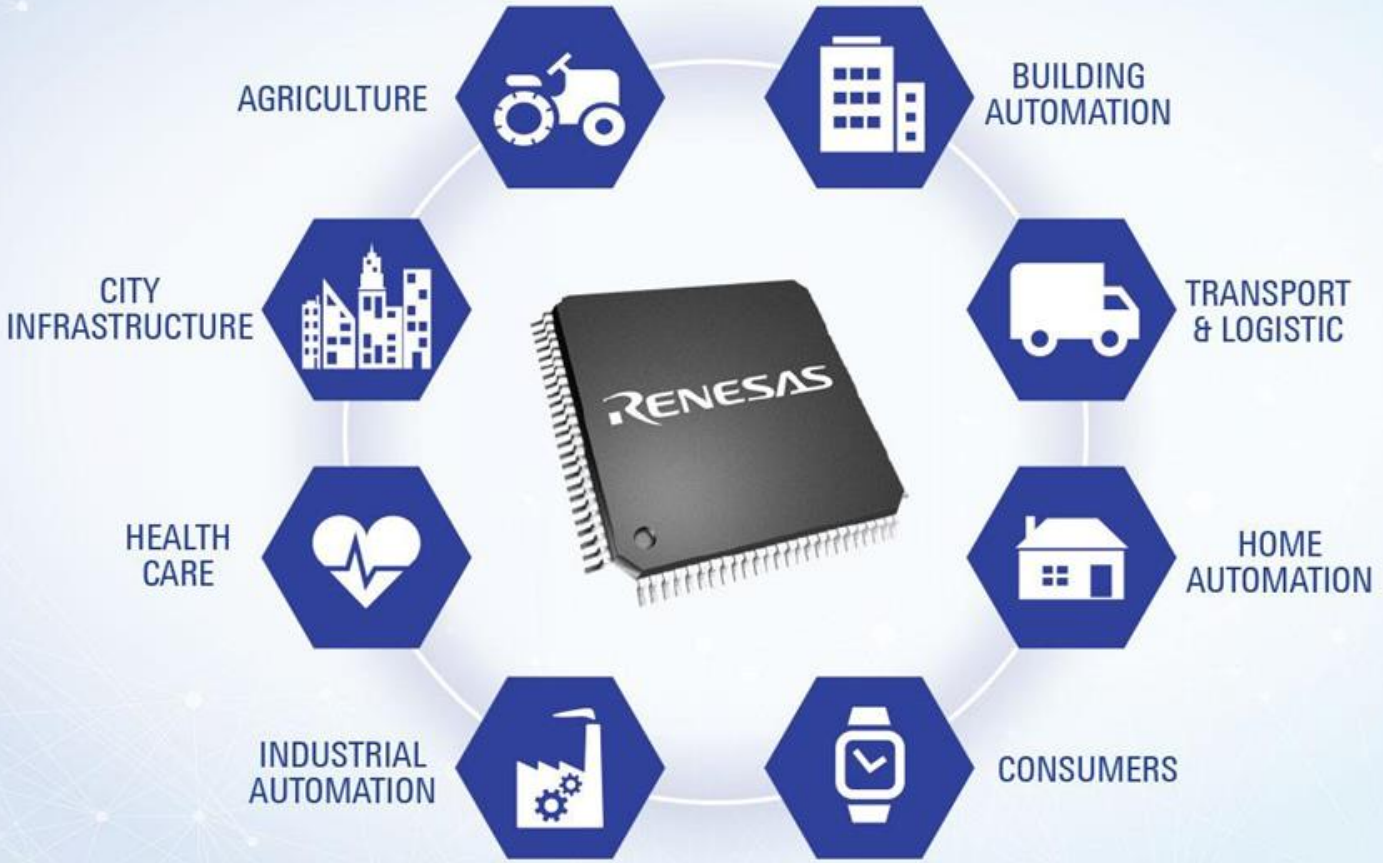
Making Edge AI A Reality

www.syntiant.com

TINY ML TALKS webcast

Platinum Strategic Partners

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



renesas.com

AITRIOS

DEPLOY VISION AI AT THE EDGE AT SCALE

SONY

TINY ML TALKS webcast

Gold Strategic Partners

ANALOG DEVICES

AHEAD OF WHAT'S POSSIBLE™

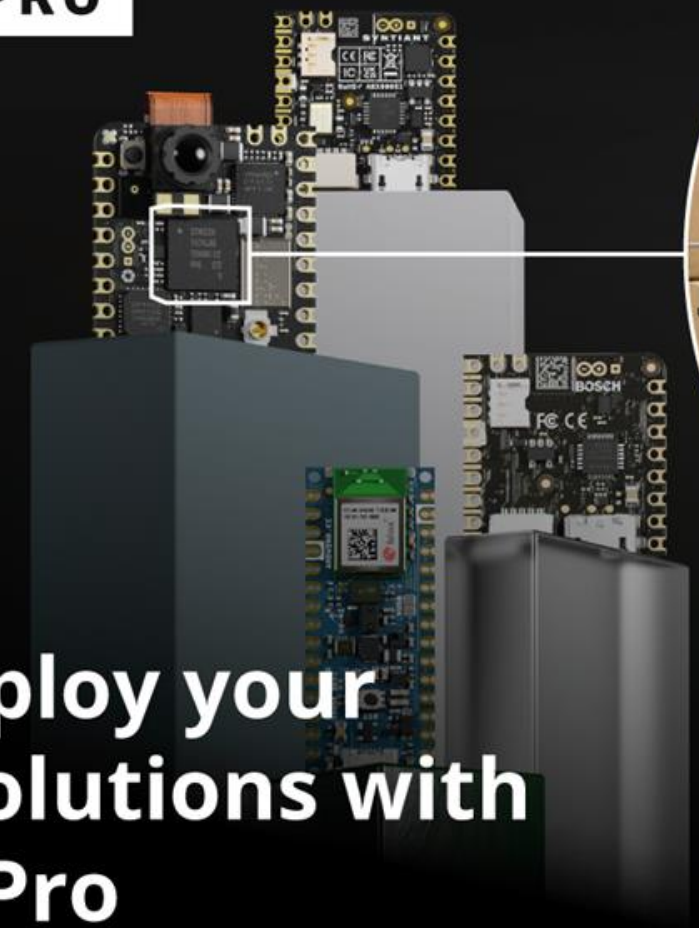
ANALOG DEVICES

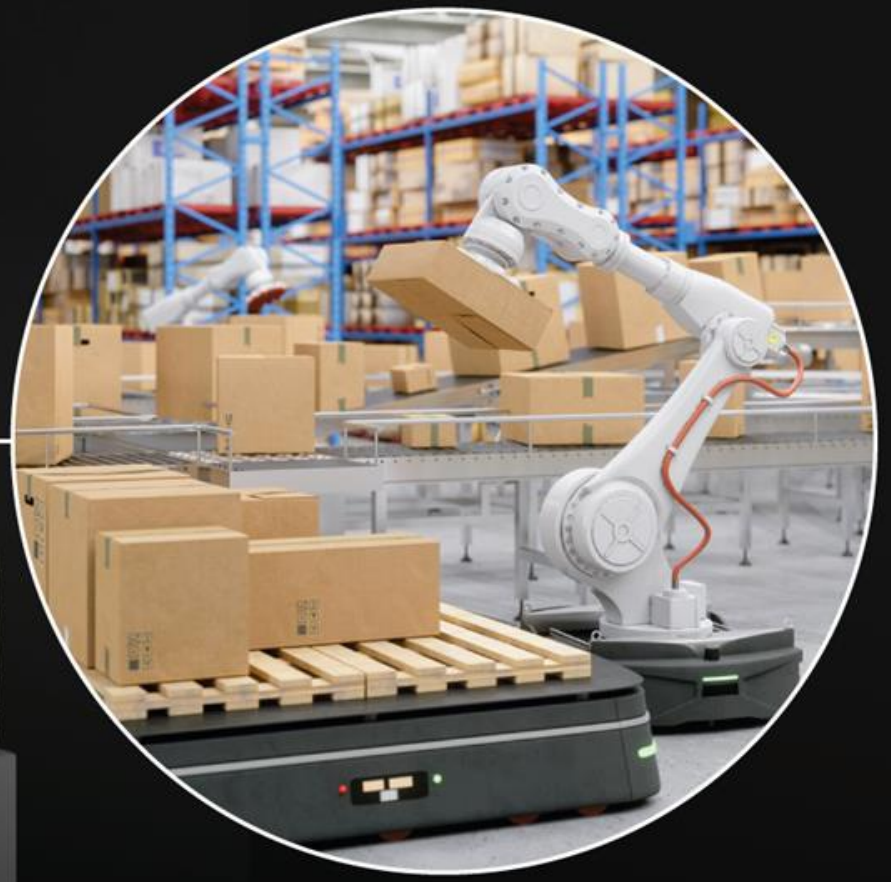
AHEAD OF WHAT'S POSSIBLE™

Where what if becomes what is.

Witness potential made possible at analog.com.

ARDUINO PRO™





Easily deploy your tinyML solutions with Arduino Pro

arduino.cc/pro

Made In Italy

arm AI

Powering tinyML Innovation

Arm AI Virtual Tech Talks

The latest in AI trends, technologies & best practices from Arm and our Ecosystem Partners.

Demos, code examples, workshops, panel sessions and much more!

Fortnightly Tuesday @ 4pm GMT/8am PT

Find out more:
www.arm.com/techtalks

Driving decarbonization and digitalization. Together.

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

www.infineon.com

innatera

NEUROMORPHIC INTELLIGENCE FOR THE SENSOR-EDGE

www.innatera.com


Microsoft

The Right Edge AI Tools Can Make or Break Your Next Smart IoT Product

SensiML

Analytics Toolkit Suite

sensiml.com/tinyML

STMicroelectronics provides extensive solutions to make tiny Machine Learning easy

life.augmented

www.st.com/ai

synaptics

ENGINEERING EXCEPTIONAL EXPERIENCES

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

www.synaptics.com

TINY ML TALKS webcast

Silver Strategic Partners

brainchip

GREENWAVES TECHNOLOGIES

Groovety Inc.

IBM

imagimob

Nota AI

NXP

POLYN TECHNOLOGY

Qeexo

Schneider Electric

SILICON LABS

Join Growing tinyML Communities:

15.2k 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/

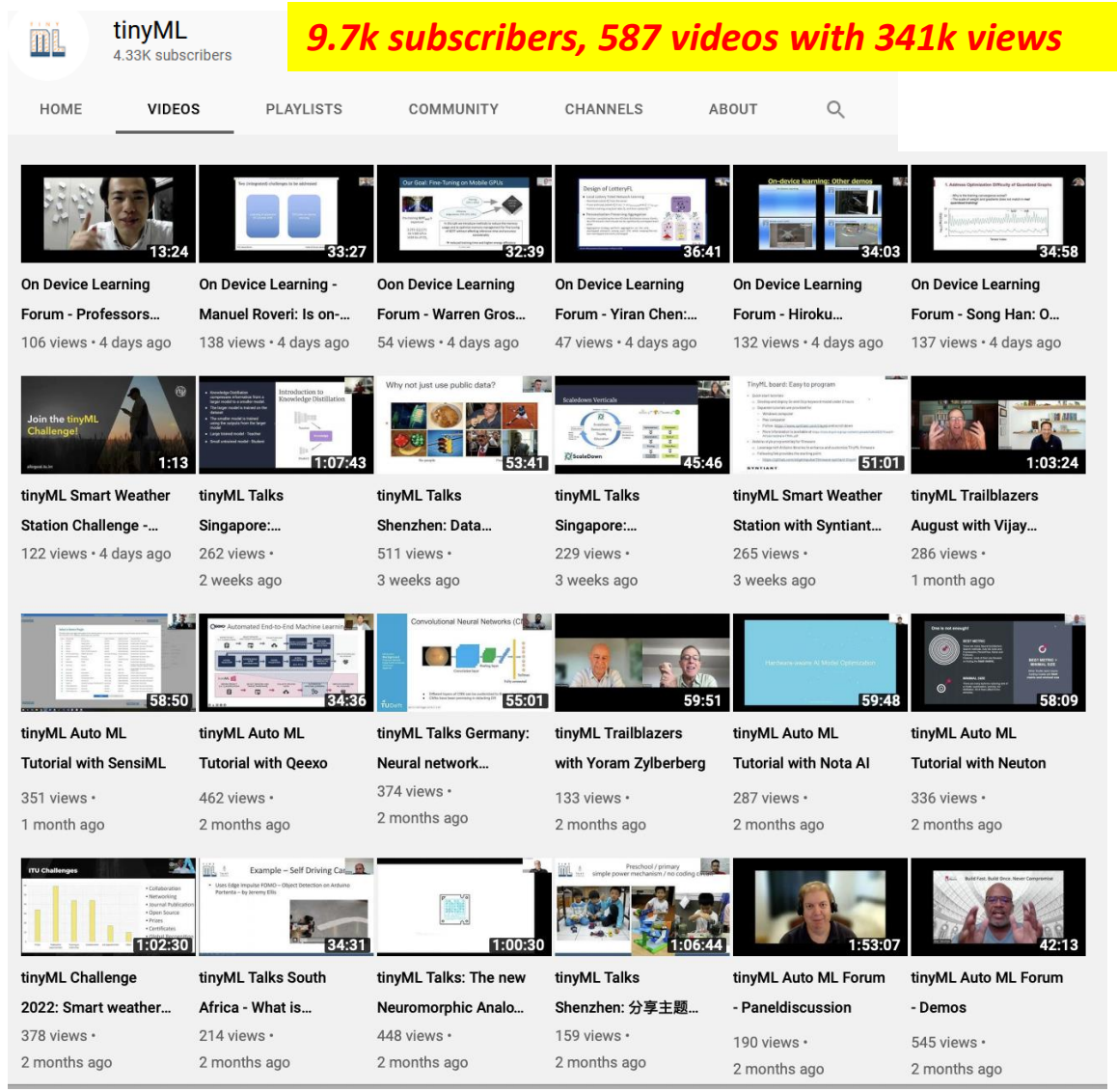
3.7k members
&
12.4k followers

The tinyML Community

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

TINY ML TALKS webcast

YouTube



Subscribe to tinyML YouTube Channel for updates and notifications

(including this video)

www.youtube.com/tinyML

TINY ML TALKS webcast

TINY ML FOUNDATION

EMEA 2023

https://www.tinyml.org/event/emea-2023

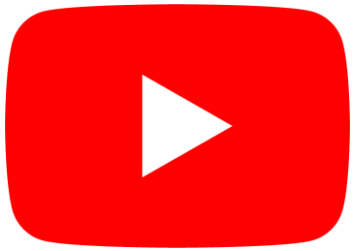
More sponsorships are available: sponsorships@tinyML.org

TINY ML TALKS webcast

Reminders

Slides & Videos will be posted tomorrow

tinyml.org/forums



youtube.com/tinyml

Please use the Q&A window for your questions

TINY ML TALKS webcast

Andrea Basso

Andrea Basso is currently serving as research director in Synesthesia Italy where oversees the research and development activities in AI and IoT areas. He is the chair of the AIF-DC in MPAI. He is also advisor at the PROGRESS TECH TRANSFER investment fund. He serves also as CTO of MITO Technology (Italy), senior expert for the European Commission and for WIPO. In previous positions served as CEO of Sisvel Technology and CTO of the Sisvel group where he oversaw evolution of Sisvel strategic technology areas and worked on business strategy and new market development. He has 182 granted patents mainly in the area of multimedia indexing and video coding. While in Bell Labs and AT&T Labs – Research USA, as Research Manager has developed 22 years of research experience he led research on multimedia technologies and he has developed innovative services and architectures for IPTV and Over The Top (OTT). Andrea has been involved in the development of standards in several international bodies including IETF, ISO/MPEG, 3GPP, ITU-T and IMTC. He has published 60 papers, several books and book chapters. He is a frequent speaker in international conferences and events.

TINY ML TALKS webcast

Outline

- TinyML (Tiny Machine Learning) and its applications.
- MPAI and MPAI-AIF platform and their significance.
- potential benefits of combining TinyML and MPAI-AIF
- Use cases and implementations
- AIF 2.0 : focus on security

What is TinyML?

TINY ML TALKS webcast

Definition

TinyML refers to the deployment of ML algorithms on resource-constrained devices (MCUs).

Benefits

Enables local, low-latency, and energy-efficient inference on edge devices.

Examples

Smart wearables, environmental sensors, industrial IoT, etc.

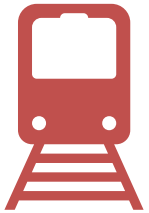
TINY ML TALKS webcast

Advantages of TinyML

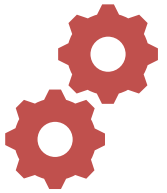
Edge computing	Enables processing and decision-making directly on the edge device, reducing latency and dependence on the cloud.
Privacy and security	Data remains on the device, minimizing privacy concerns and reducing reliance on network connectivity.
Energy efficiency	Local inference reduces the need for data transmission, saving energy and extending battery life.
Scalability	TinyML enables deploying machine learning capabilities across a vast number of edge devices, creating a network of intelligent endpoints.

TinyML Workflow

Data collection: Gather labeled or annotated data for training TinyML models.



Model development: Train or design machine learning models tailored for resource-constrained devices.



Model optimization: Compress, quantize, or prune the model to reduce its size and computational requirements.

Deployment: Load the optimized TinyML model onto the target device for local inference.

TINY ML TALKS webcast

What is MPAI

https://mpai.community/

Book:
[Towards Pervasive and Trustworthy Artificial Intelligence]

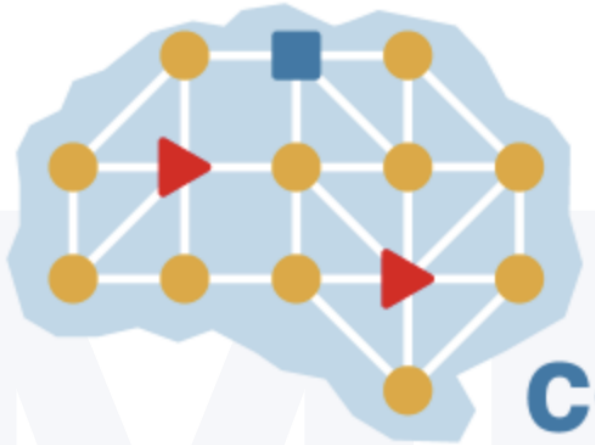
Towards Pervasive and Trustworthy Artificial Intelligence

How standards can put a great technology at the service of humankind

By: Alessandro Artusi, Andrea Basso, Marina Bosi, Sergio Canazza, Leonardo Chiariglione, Miran Choi, Fabiano Columbano, Mert Burkay Çöteli, Nadir Dalla Pozza, Roberto Dini, Michelangelo Guarise, Hüseyin Hacıhabiboğlu, Roberto Iacoviello, Chuanmin Jia, Jisu Kang, Panos Kudumakis, Valeria Lazzaroli, Marco Mazzaglia, Guido Perboli, Niccolò Pretto, Paolo Ribeca, Mariangela Rosano, Mark Seligman

MOVING PICTURE, AUDIO AND DATA CODING BY ARTIFICIAL INTELLIGENCE

Home

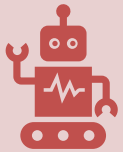


MPAI.
community

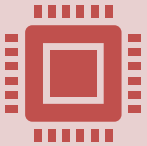
The international, unaffiliated, no-profit organisation developing standards for AI-based data coding with clear Intellectual Property Rights licensing frameworks.

What is MPAI-AIF?

Definition: MPAI-AIF stands for MPAI AI Framework.

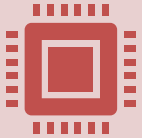


Significance: MPAI-AIF provides a standardized way to interface and integrate AI modules into applications.

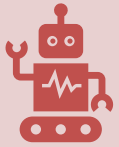


Purpose: Enables interoperability, composability, and scalability of AI solutions across different devices and platforms.

Key Features of MPAI AIF



Interoperability: Enables seamless integration of AI modules from different vendors and platforms.



Composability: Facilitates the combination and reusability of AI modules to create complex and customized AI workflows.

Scalability: Supports the deployment of AI solutions across various devices and networked environments.

An AI Framework (AIF) is needed

AI framework enables creation, execution, composition and update of AIM-based workflows for AI solutions interconnecting multi-vendor AIMs, operating in the standard AI framework and exchanging data in standard formats via standard interfaces.

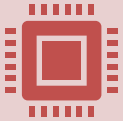
It will benefit various actors

Technology providers	Application developers	Innovation	Consumers	Society
able to offer their conforming AI technologies to an open market	open market for their applications need	demand for novel and more performing AI components	offered a wider choice of better AI applications by a competitive market	lift the veil of opacity from large, monolithic AI-based applications.

TINY ML TALKS webcast

MPAI. community

Benefits of MPAI AIF

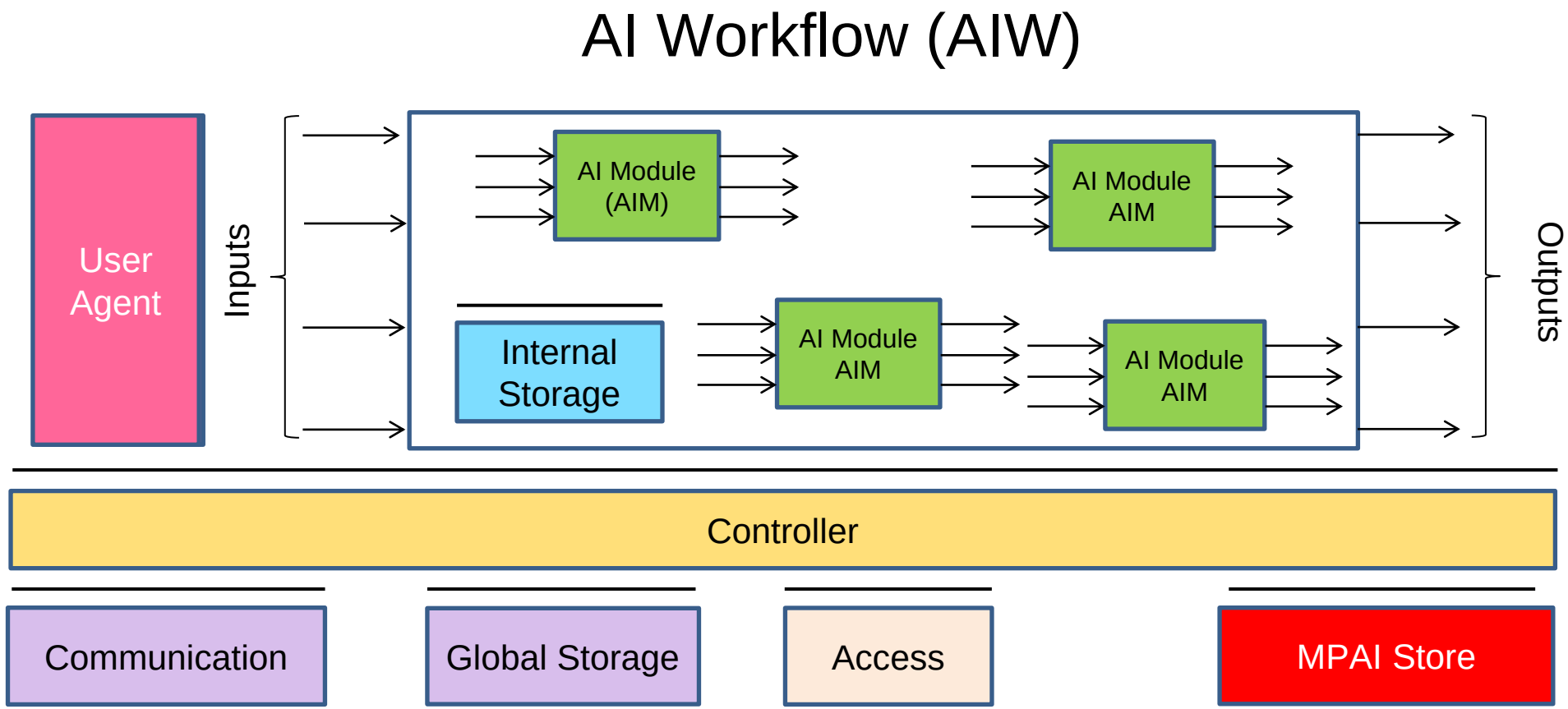


Standardization Facilitates the development of interoperable and reusable AI modules across different platforms and vendors.

Flexibility Enables the creation of customized AI workflows by combining and reconfiguring AI modules.

Ecosystem Growth Encourages collaboration and innovation among AI module developers, leading to a vibrant marketplace for AI solutions.

MPAI-AIF: AI Framework

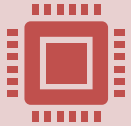


TINY ML TALKS webcast

AIF Components

Access	access to static or slowly changing data
AI Module (AIM)	data processing element receiving Inputs and producing Outputs according to its Function. An AIM may be an aggregation of AIMs.
AI Workflow (AIW)	an organised aggregation of AIMs implementing a Use Case
Communication	connects the Components of an AIF.

AIF Interfaces



Define the communication and data exchange protocols between AI modules and AIF Components.

interfaces for data input/output, control signals, and metadata exchange.

Ensure compatibility and interoperability among different AI modules.

AI Framework Features

- Event-based and port and channel-based (unicast)
- High-Priority Messages and Normal-Priority Messages
- Controller may run on a different computing platform than the AIW.
- The AIMs of a AIW may run on different computing platforms
- The Controller will always be present even if the AIF is a lightweight Implementation.
- AIMs may be hot-pluggable and register themselves on the fly.

MPAI-AIF is IEEE P3301

IEEE SA STANDARDS ASSOCIATION | IEEE P3301 Artificial Intelligence Framework Working Group

IEEE Advancing Technology for Humanity

IEEE 3301 2022 Edition, December 3, 2022

LOOK INSIDE

Complete Document

Adoption of Moving Picture, Audio and Data Coding by Artificial Intelligence (MPAI) Technical Specification Artificial Intelligence Framework (AIF) 1.1

VIEW ABSTRACT

PRODUCT DETAILS

Detail Summary View all details

Active, Most Current

EN

Format	Details	Price (USD)
PDF	Single User	$83.00
Print	In Stock	$103.00

ENHANCED BY Google

Home | Meetings | Members | Meeting Agenda & Minutes

IEEE P3301 - ARTIFICIAL INTELLIGENCE FRAMEWORK WORKING GROUP

Title: Adoption of Moving Picture, Audio and Data Coding by Artificial Intelligence (MPAI) Technical Specification Artificial Intelligence Framework (AIF) Version 1

Scope: The MPAI AI Framework (MPAI-AIF) Technical Specification specifies architecture, interfaces, protocols and Application Programming Interfaces (API) of an AI Framework (AIF), especially designed for execution of AI-based implementations, but also suitable for mixed AI and traditional data processing workflows.

MPAI-AIF possesses the following main features:

- Operating System-independent.
- Component-based modular architecture with standard interfaces.
- Interfaces encapsulate Components to abstract them from the development environment.
- Interface with the MPAI Store enables access to validated Components.
- Component can be implemented as: software only (from Micro-Controller Units to High-Performance Computing), hardware only, and hybrid hardware-software.
- Component system features are:
- Execution in local and distributed Zero-Trust architectures.
- Possibility to interact with other Implementations operating in proximity.
- Direct support to Machine Learning functionalities.

WG OFFICERS

Chair
Steve Dukes, Dreamerse

Vice Chair

Andrea Basso, Synesthesia

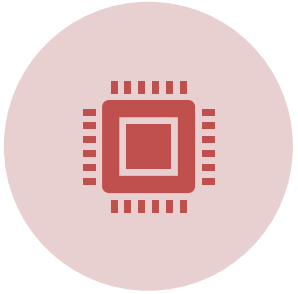
Secretary

Program Manager
Jonathan Goldberg, IEEE Standards Association

Benefits of joint TinyML and AIF

Efficient and low-latency machine learning inference on resource-constrained devices.

Standardized interoperability and composability through MPAI-AIF.

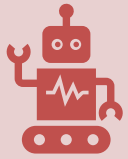


Scalability across a wide range of applications and devices.

Improved user experience and decision-making capabilities.

Use Cases

Use Case 1

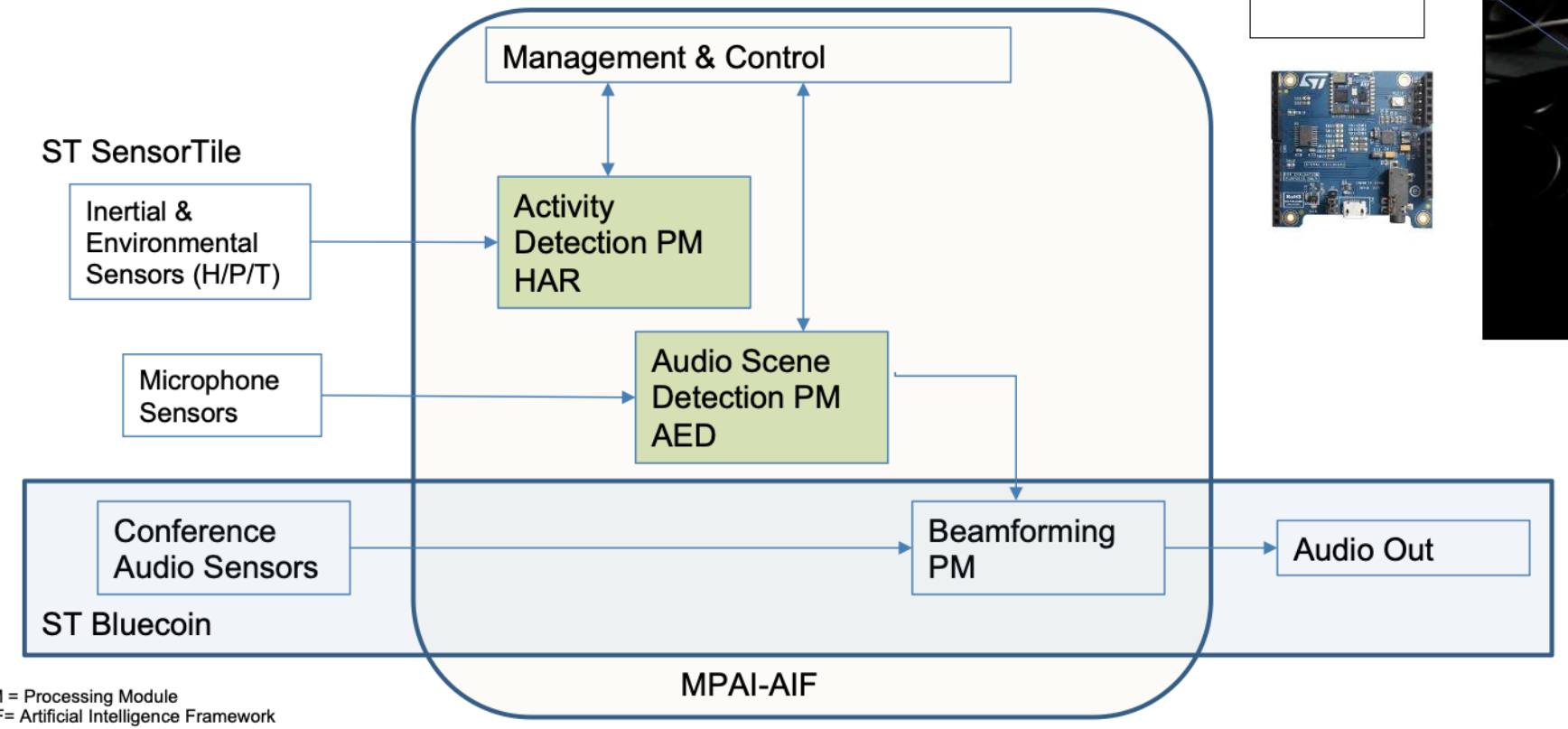


AIM1: AI based and performs Audio Scene Classification (ASC)

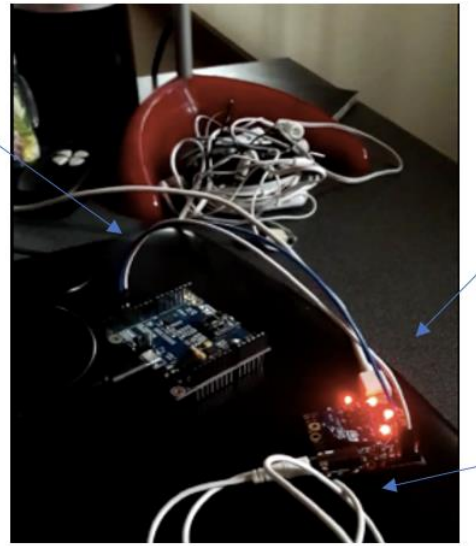
AIM2: beamforming and source localization on signals coming from a microphone array (4 microphones).

On the basis acoustic environment classification done by AIM1, AIM2 switches from direct signal passthrough to beamforming based on source localization.

Use Case 1



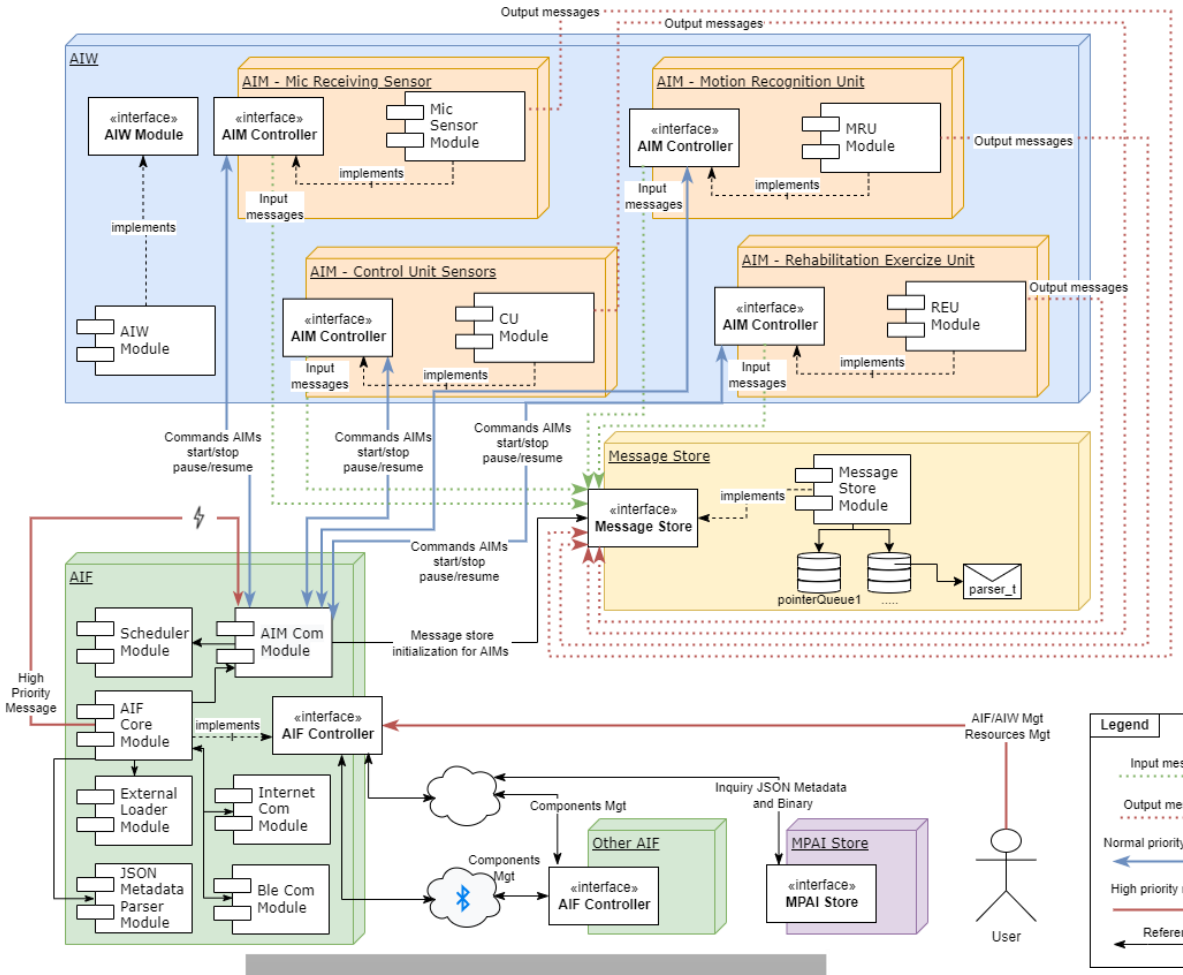
ST Sensortile STEVAL-STLKT01V1

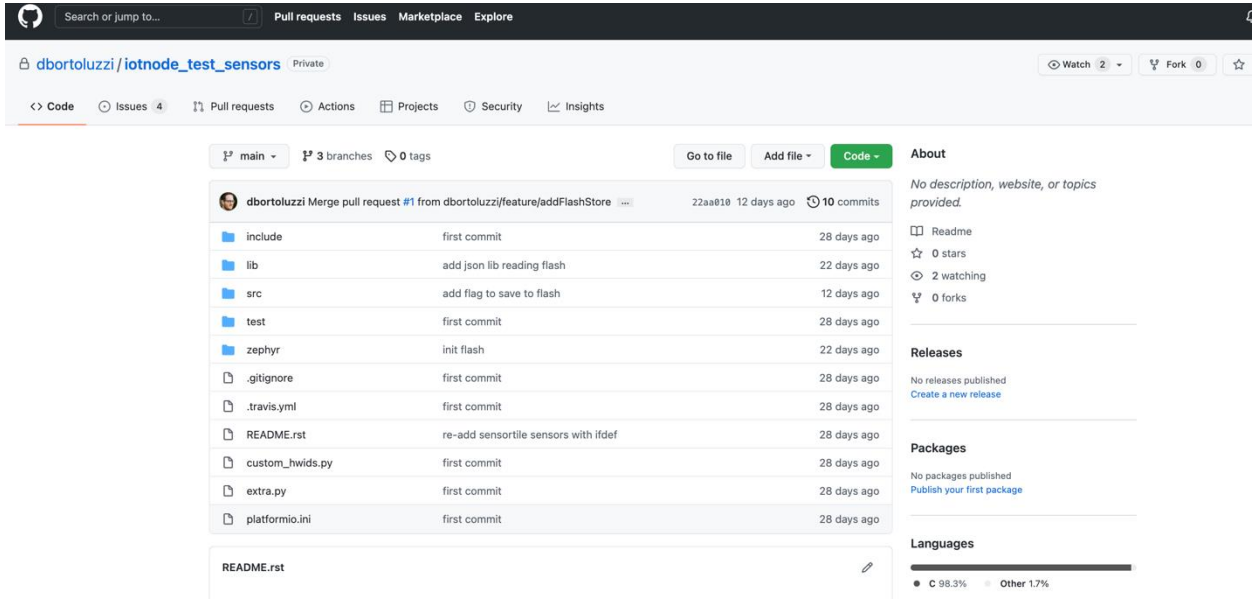


ST Bluecoin STEVAL-BCNKT01V1

Audio Output

PM = Processing Module
AIF= Artificial Intelligence Framework
HAR = Human Activity Recognition *
AED = Audio Event Detection *
* https://www.st.com/en/embedded-software/fp-ai-sensing1.html



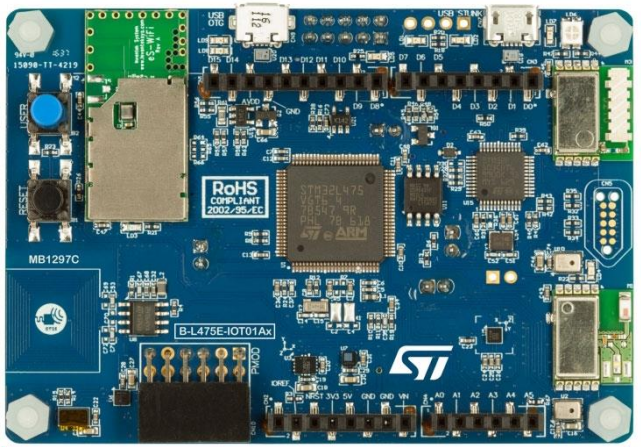


Matching a specific movement with specific sounds

Entertainment | Rehabilitation

A given sound is emitted and the user has to perform a predefined movement

Use Case 2



The board's serial number is indicated on a sticker under the MB1297 reference on the bottom of the board. If this number is lower than 182404896 for B-L475EIOT01A1C boards, or than 184906074 for B-L475EIOT01A2C boards, the default firmware connecting to AWS Cloud needs updating.
Download the latest version available at http://www.st.com/x-cube-aws

STM32

aws qualified device FreeRTOS

arm MBED Enabled

TINY ML TALKS webcast

Implementation

- **Mic Receiving sensor AIM** (data_mic) reads the data from the microphone performs filtering and analysis sends the data to the message_store on channel MIC_PEAK_DATA_CHANNEL
- **Sensors AIM** (sensor_aim) reads the data from the sensors, the inertial unit, and processes them sends data to the message_store on channel SENSORS_DATA_CHANNEL)
- **Human Activity AIM** motion_aim reads data from sensors_aim, and detects the movement pattern. Sends event to the message_store on channel MOTION_DATA_CHANNEL
- **Rehabilitation_goal AIM** reads and crosses the data coming from the message_store from the MIC_PEAK_DATA_CHANNEL and MOTION_DATA_CHANNEL channels Sends out to LEDs to indicate correct/wrong exercise

Basso, A., Bortoluzzi, D., Torta, G.: Implementation of an iot wearable prototype on a standard ai architecture. In: 2022 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech), pp. 1–5. IEEE (2022)

TINY ML TALKS webcast

References

Andrea Basso et al.	Implementation of an IoT Wearable Prototype on a Standard AI Architecture	COMMON-WEARS 2022
Andrea Basso et al.	Architecture standardization for AI deployment on tiny micro-controllers on a Standard AI Architecture	12th IEEE Int. Conf. on CT 2022
Andrea Basso et al.	AI-Based Media Coding Standards	SMPTE Motion Imaging Journal
Andrea Basso et al.	Architecture standardization for AI deployment on tiny micro-controllers	IEEE – ICCE 2022

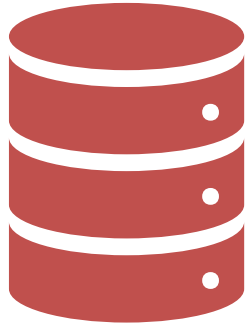
TINY ML TALKS webcast

Security in TinyML and MPAI-AIF

Muhammad Yasir Shabir, Gianluca Torta, Andrea Basso, Ferruccio Damiani "Towards Secure TinyML on a Standardized AI Architecture" to appear in Device-Edge-Cloud Continuum - Paradigms, Architectures and Applications" Springer-Verglag (to appear)

Importance of Security

critical role of security in TinyML and MPAI-AIF



Need to protect sensitive data and ensure the integrity of AI models

TINY ML TALKS webcast

Data Privacy

PROTECTING USER DATA IN APPLICATIONS.

NEED FOR SECURE DATA TRANSMISSION AND STORAGE.

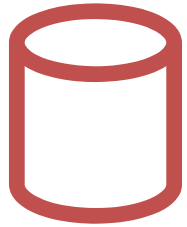
ENCRYPTION AND ANONYMIZATION TO SAFEGUARD DATA PRIVACY.

TINY ML TALKS webcast

Model Protection

Protecting models from unauthorized access or tampering

Model encryption, obfuscation, and hardware-based security measures



Secure model updates and version control

TINY ML TALKS webcast

Threat Mitigation

secure boot, secure communication protocols, and access controls.

continuous monitoring, vulnerability assessments, and threat intelligence.

TINY ML TALKS webcast

Certification and Compliance

Adhering to security standards and regulations.

Common Criteria and industry-specific security guidelines.

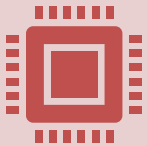
Security audits and assessments throughout the development and deployment lifecycle.

TINY ML TALKS webcast

MPAI AIF and Security

multiple technology providers with potentially different security requirements

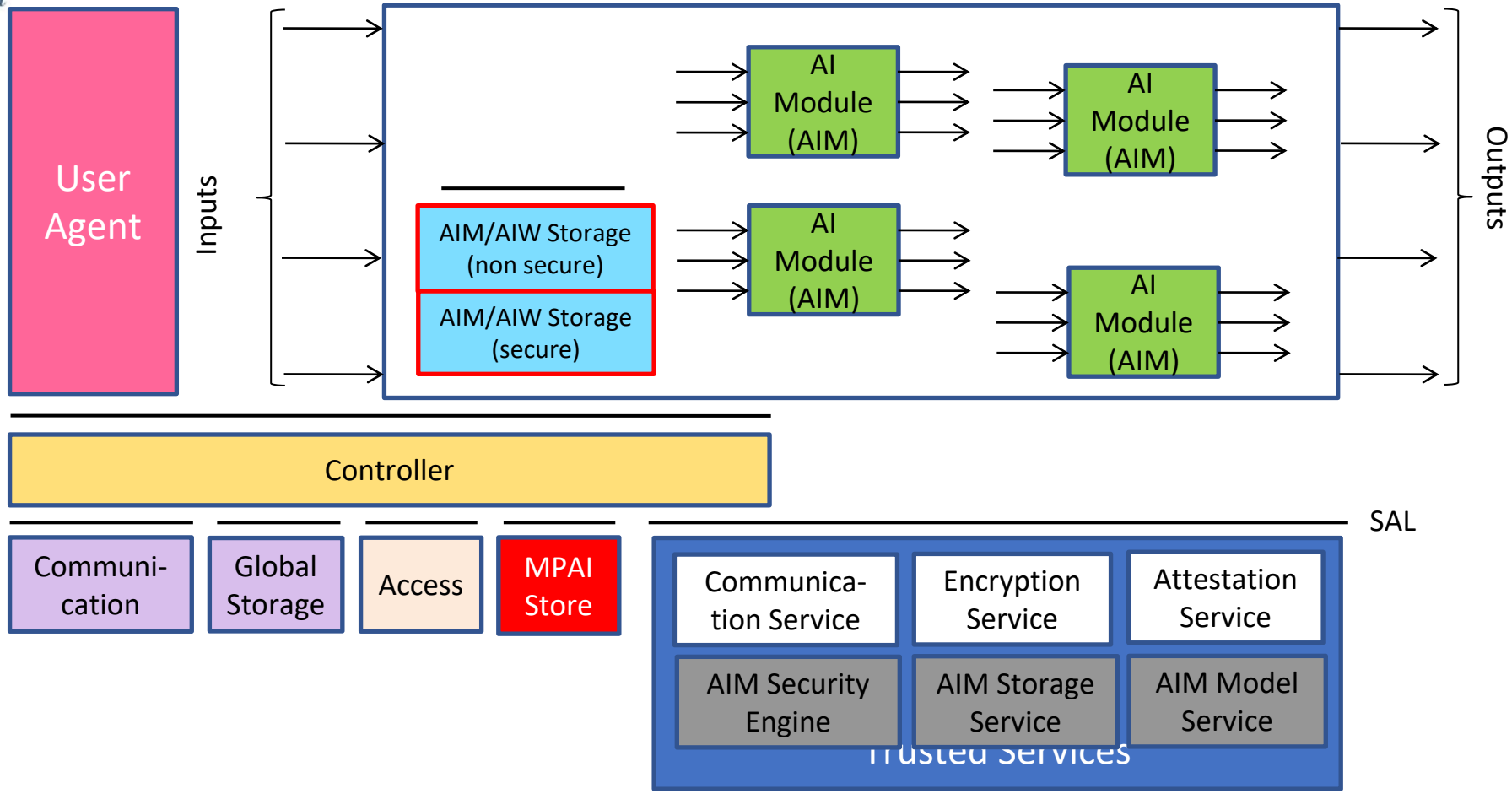
multiple users with potentially different security requirements



Application developers shall be able to select the application's security either or both by:

TINY ML TALKS webcast

MPAI AIF V2



TINY ML TALKS webcast

Features of MPAI-AIF (V2)

High-level implementation-independent Trusted Services API shall be able to use hardware and OS security features already existing in the hardware and software of the environment in which the AIF is implemented.

Security supported at AIF level, in the following configuration:

TINY ML TALKS webcast

Features of MPAI-AIF (V2)

Controller

TINY ML TALKS webcast

Conclusion

Security is a vital aspect of integrating TinyML and MPAI-AIF.

holistic security approach encompassing data privacy, model protection, threat mitigation, and compliance.

By prioritizing security, we can build trust and unlock the full potential of TinyML and MPAI-AIF.

TINY ML TALKS webcast

Questions?

Contact information andrea.basso@synesthesia.it

https://mpai.community

TINY ML TALKS webcast

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

Copyright Notice

www.tinyml.org

Copyright Notice

www.tinyml.org