“Standardized AI Architectures for Secure TinyML”

Andrea Basso – Research director, Synesthesia

June 20, 2023
Thank you, tinyML Strategic Partners, for committing to take tinyML to the next Level, together
Executive Strategic Partners
EDGE IMPULSE

The Leading Development Platform for Edge ML

edgeimpulse.com
Advancing AI research to make efficient AI ubiquitous

A platform to scale AI across the industry

- **Power efficiency**: Model design, compression, quantization, algorithms, efficient hardware, software tools
- **Personalization**: Continuous learning, contextual, always-on, privacy-preserved, distributed learning
- **Efficient learning**: Robust learning through minimal data, unsupervised learning, on-device learning
- **Perception**: Object detection, speech recognition, contextual fusion
- **Reasoning**: Scene understanding, language understanding, behavior prediction
- **Action**: Reinforcement learning for decision making

IoT/IIoT
Edge cloud
Automotive
Cloud
Mobile

Qualcomm AI Research is an initiative of Qualcomm Technologies, Inc.
Accelerate Your Edge Compute

SYNTIANT

Making Edge AI A Reality

www.syntiant.com
Platinum Strategic Partners
Renesas is enabling the next generation of AI-powered solutions that will revolutionize every industry sector.

- Agriculture
- Building Automation
- City Infrastructure
- Transport & Logistic
- Health Care
- Home Automation
- Industrial Automation
- Consumers

renesas.com
DEPLOY VISION AI AT THE EDGE AT SCALE
Gold Strategic Partners
Witness potential made possible at analog.com.
Easily deploy your tinyML solutions with Arduino Pro

arduino.cc/pro

Made In Italy
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
NEUROMORPHIC INTELLIGENCE FOR THE SENSOR-EDGE

www.innatera.com
The Right Edge AI Tools Can Make or Break Your Next Smart IoT Product

Analytics Toolkit Suite

AutoML

Data Collection

Team Collaboration

Version Control and Model Management

Code Generation

Test & Validation

Data Labeling

sensiml.com/tinyML
STMicroelectronics provides extensive solutions to make tiny Machine Learning easy
ENGINEERING EXCEPTIONAL EXPERIENCES

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

www.synaptics.com
Join Growing tinyML Communities:

tinyML - Enabling ultra-low Power ML at the Edge

The tinyML Community
https://www.linkedin.com/groups/13694488/

15.2k members in 49 Groups in 41 Countries

3.7k members & 12.4k followers
Subscribe to tinyML YouTube Channel for updates and notifications (including this video)
www.youtube.com/tinyML
EMEA 2023
https://www.tinyml.org/event/emea-2023

More sponsorships are available: sponsorships@tinyML.org
Reminders

Slides & Videos will be posted tomorrow

tinyml.org/forums  youtube.com/tinyml

Please use the Q&A window for your questions
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.
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?

<table>
<thead>
<tr>
<th>Definition</th>
<th>TinyML refers to the deployment of ML algorithms on resource-constrained devices (MCUs).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benefits</td>
<td>Enables local, low-latency, and energy-efficient inference on edge devices.</td>
</tr>
<tr>
<td>Examples</td>
<td>Smart wearables, environmental sensors, industrial IoT, etc.</td>
</tr>
</tbody>
</table>
### Advantages of TinyML

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

MOVING PICTURE, AUDIO AND DATA CODING BY ARTIFICIAL INTELLIGENCE

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

<table>
<thead>
<tr>
<th>Technology providers</th>
<th>Application developers</th>
<th>Innovation</th>
<th>Consumers</th>
<th>Society</th>
</tr>
</thead>
<tbody>
<tr>
<td>able to offer their conforming AI technologies to an open market</td>
<td>open market for their applications need</td>
<td>demand for novel and more performing AI components</td>
<td>offered a wider choice of better AI applications by a competitive market</td>
<td>lift the veil of opacity from large, monolithic AI-based applications.</td>
</tr>
</tbody>
</table>
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
## AIF Components

<table>
<thead>
<tr>
<th>Component</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access</td>
<td>access to static or slowly changing data</td>
</tr>
<tr>
<td>AI Module (AIM)</td>
<td>data processing element receiving Inputs and producing Outputs according to its Function. An AIM may be an aggregation of AIMS.</td>
</tr>
<tr>
<td>AI Workflow (AIW)</td>
<td>an organised aggregation of AIMS implementing a Use Case</td>
</tr>
<tr>
<td>Communication</td>
<td>connects the Components of an AIF.</td>
</tr>
</tbody>
</table>
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

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

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 AIF 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.
- Interfaces with the MPAI Store enables access to validated Components.
- Component can be implemented as a 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 interwork with other implementations operating in proximity.
- Direct support to Machine Learning functionalities.

WG OFFICERS
Chair
Stone Dukas, Dreamwire
Vice Chair
Andrew Basos, Synthesia
Secretary
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

Management & Control

Activity Detection PM
HAR

Audio Scene Detection PM
AED

Conference Audio Sensors

ST Bluecoin

Beamforming PM

AUD Output

ST SensorTile

ST SensorTile
STEVAL-STLKT01V1

ST Bluecoin

ST Bluecoin
STEVAL-BCNKT01V1

Audio Output

PM = Processing Module
AIF = Artificial Intelligence Framework
HAR = Human Activity Recognition *
AED = Audio Event Detection *
Use Case 2

Matching a specific movement with specific sounds

Entertainment  Rehabilitation

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

- **Mic Receiving sensor AIM** (data_mic) reads the data from the microphone, performs filtering and analysis, and 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 before sending 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, sending out to LEDs to indicate correct/wrong exercise.

---

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Title</th>
<th>Conference/Platform</th>
</tr>
</thead>
<tbody>
<tr>
<td>Andrea Basso et al.</td>
<td>Implementation of an IoT Wearable Prototype on a Standard AI Architecture</td>
<td>COMMON-WEARS 2022</td>
</tr>
<tr>
<td>Andrea Basso et al.</td>
<td>Architecture standardization for AI deployment on tiny micro-controllers on a Standard AI Architecture</td>
<td>12th IEEE Int. Conf. on CT 2022</td>
</tr>
<tr>
<td>Andrea Basso et al.</td>
<td>AI-Based Media Coding Standards</td>
<td>SMPTE Motion Imaging Journal</td>
</tr>
<tr>
<td>Andrea Basso et al.</td>
<td>Architecture standardization for AI deployment on tiny micro-controllers</td>
<td>IEEE – ICCE 2022</td>
</tr>
</tbody>
</table>
Security in TinyML and MPAI-AIF

Importance of Security

critical role of security in TinyML and MPAI-AIF

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

Protecting user data in applications.

Need for secure data transmission and storage.

Encryption and anonymization to safeguard data privacy.
Model Protection

Protecting models from unauthorized access or tampering

Model encryption, obfuscation, and hardware-based security measures

Secure model updates and version control
Threat Mitigation

- secure boot, secure communication protocols, and access controls.
- continuous monitoring, vulnerability assessments, and threat intelligence.
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.
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:
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:
Features of MPAI-AIF (V2)

Controller
Conclusion

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

A 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.
Questions?

Contact information andrea.basso@synesthesia.it

https://mpai.community
Thank you, tinyML Strategic Partners, for committing to take tinyML to the next Level, together
Copyright Notice

This multimedia file is copyright © 2023 by tinyML Foundation. All rights reserved. It may not be duplicated or distributed in any form without prior written approval.

tinyML® is a registered trademark of the tinyML Foundation.

www.tinymce.org
Copyright Notice

This presentation in this publication was presented as a tinyML® Talks webcast. The content reflects the opinion of the author(s) and their respective companies. The inclusion of presentations in this publication does not constitute an endorsement by tinyML Foundation or the sponsors.

There is no copyright protection claimed by this publication. However, each presentation is the work of the authors and their respective companies and may contain copyrighted material. As such, it is strongly encouraged that any use reflect proper acknowledgement to the appropriate source. Any questions regarding the use of any materials presented should be directed to the author(s) or their companies.

tinyML is a registered trademark of the tinyML Foundation.

www.tinyml.org