|  |  |  |
| --- | --- | --- |
|  | | Moving Picture, Audio and Data Coding by Artificial Intelligence  www.mpai.community  **Public document** |
| **N1378** | 2023/09/29 | |
| **Source** | General Assembly #35 (MPAI-36) | |
| **Title** | MPAI Work Plan | |
| **Target** | MPAI Community | |

Contents

[1 Introduction 1](#_Toc147647027)

[2 MPAI-AIF 3](#_Toc147647028)

[2.1 Version 1 3](#_Toc147647029)

[2.2 Version 2 3](#_Toc147647030)

[3 MPAI-AIH 4](#_Toc147647031)

[4 MPAI-CAE 5](#_Toc147647032)

[4.1 Version 1 5](#_Toc147647033)

[4.2 Version 2 6](#_Toc147647034)

[5 MPAI-CUI 6](#_Toc147647035)

[6 MPAI-CAV 7](#_Toc147647036)

[7 MPAI-EEV 8](#_Toc147647037)

[8 MPAI-EVC 9](#_Toc147647038)

[9 MPAI-GSA 10](#_Toc147647039)

[10 MPAI-MCS 11](#_Toc147647040)

[11 MPAI-MMC 11](#_Toc147647041)

[11.1 Version 1 11](#_Toc147647042)

[11.2 Version 2 11](#_Toc147647043)

[12 MPAI-MMM 13](#_Toc147647044)

[13 MPAI-NNW 13](#_Toc147647045)

[14 MPAI-OSD 14](#_Toc147647046)

[15 MPAI-PAF 14](#_Toc147647047)

[16 MPAI-SPG 15](#_Toc147647048)

[17 MPAI-XRV 15](#_Toc147647049)

# Introduction

MPAI’s standards development is based on projects evolving through a workflow extending on 7 + 1 stages. The 7th stage is split in 4: Technical Specification, Reference Software Specification, Conformance Testing Specification, and Performance Assessment Specification. Technical Reports can also be developed to investigate new fields.

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **Acr** | **Name** | **Description** |
| 0 | IC | Interest Collection | Collection and harmonisation of use cases proposed. |
| 1 | UC | Use cases | Proposals of use cases, their description and merger of compatible use cases. |
| 2 | FR | Functional Requirements | Identification of the functional requirements that the standard in­cluding the Use Case should satisfy. |
| 3 | CR | Commercial Requirements | Development and approval of the framework licence of the stan­dard. |
| 4 | CfT | Call for Technologies | Preparation and publication of a document calling for technologies supporting the functional and commercial requirements. |
| 5 | SD | Standard Development | Development of the standard in a specific Development Com­mit­tee (DC). |
| 6 | CC | Community Comments | When the standard has achieved sufficient maturity it is published with request for comments. |
| 7 | MS | MPAI standard | The standard is approved by the General Assembly comprising 4 documents. |
| 7.1 | TS | Technical Specification | The normative specification to make a conforming implement­ation. |
| 7.2 | RS | Reference Software | The descriptive text and the software implementing the Technical Specification |
| 7,3 | CT | Conformance Testing | The Specification of the steps to be executed to test an implementation for conformance. |
| 7.4 | PA | Conformance Assessment | The Specification of the steps to be executed to assess an implementation for performance. |

A project progresses from one stage to the next by resolution of the General Assembly.

The stages of currently (MPAI-29) active MPAI projects are graphically represented by *Table 1*.

Legend: TS: Technical Specification, RS: Reference Software, CT: Conformance Testing, PA: Performance Assessment; V2: Version 2.

*Table 1 - Snapshot of the MPAI work plan (MPAI-36)*

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **#** | **V** | **Work area** | **IC** | **UC** | **FR** | **CR** | **CfT** | **SD** | **CC** | **TS** | **RS** | **CT** | **PA** | **TR** |
| **AIF** | 1.1 | AI Framework |  |  |  |  |  |  |  | X | X |  |  |  |
| **AIF** | 2.0 | AI Framework |  |  |  |  |  |  |  | X |  |  |  |  |
| **AIH** |  | AI Health Data |  |  |  |  | X |  |  |  |  |  |  |  |
| **PAF** | 1.0 | Portable Avatar Format |  |  |  |  |  |  |  | X |  |  |  |  |
| **CAE** | 1.4 | Context-based Audio Enhancement |  |  |  |  |  |  |  | X | X | X |  |  |
| **CAE** | 2.0 | Context-based Audio Enhancement |  |  |  |  |  |  |  | X |  |  |  |  |
| **CAV** |  | Connected Autonomous Vehicles - Architecture |  |  |  |  |  |  |  | X |  |  |  |  |
| **CUI** | 1.1 | Compression & Understanding of Industrial Data |  |  |  |  |  |  |  | X | X | X | X |  |
| **EEV** |  | AI-based End-to-End Video Coding |  |  | X |  |  |  |  |  |  |  |  |  |
| **EVC** |  | AI-Enhanced Video Coding |  |  | X |  |  |  |  |  |  |  |  |  |
| **GME** | 1.1 | Governance of the MPAI Ecosystem |  |  |  |  |  |  |  | X |  |  |  |  |
| **GSA** |  | Integrated Genomic/Sensor Analysis |  |  | X |  |  |  |  |  |  |  |  |  |
| **MCS** |  | Mixed Reality Collaborative Spaces |  |  | X |  |  |  |  |  |  |  |  |  |
| **MMC** | 1.2 | Multimodal Conversation |  |  |  |  |  |  |  | X |  |  |  |  |
| **MMC** | 2.0 | Multimodal Conversation |  |  |  |  |  |  |  | X |  |  |  |  |
| **MMM** | 1.0 | MPAI Metaverse Model – Functionalities |  |  |  |  |  |  |  |  |  |  |  | X |
| **MMM** | 1.0 | MPAI Metaverse Model – Functionality Profiles |  |  |  |  |  |  |  |  |  |  |  | X |
| **MMM** |  | MPAI Metaverse Model - Architecture |  |  |  |  |  |  |  | X |  |  |  |  |
| **NNW** | 1.0 | Neural Network Watermarking |  |  |  |  |  |  |  | X | X |  |  |  |
| **OSD** |  | Object and Scene Description |  |  |  |  |  | X |  |  |  |  |  |  |
| **SPG** |  | Server-based Predictive Multiplayer Gaming |  |  | X |  |  |  |  |  |  |  |  |  |
| **XRV** |  | XR Venues |  |  |  |  | X |  |  |  |  |  |  |  |

# MPAI-AIF

The MPAI approach to AI standards is based on the belief that by breaking up large AI applications into smaller elements (AI Modules), combined in workflows (AIW), exchanging data with a known semantics to the extent possible. not only improve explainability of AI applications but also promotes a competitive market of components with standard interfaces, possibly with improved performance compared to other implementations.

Artificial Intelligence Framework (MPAI-AIF) enables creation and automation of mixed Artif­icial Intelligence – Machine Learning – Data Processing workflows.

## Version 1

*Figure 1* shows the MPAI-AIF V1 Reference Model.

Diagram

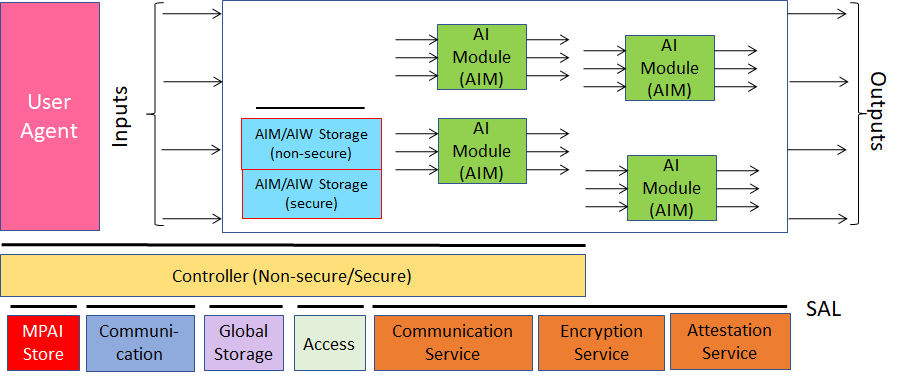
Description automatically generated

*Figure 1 - Reference model of the MPAI AI Framework (MPAI-AIF) V1*

The MPAI-AIF Technical Specification V1 and Reference Software V1 have been approved and are available [here](https://mpai.community/resources/).

## Version 2

MPAI-AIF V1 assumed that the AI Framework was secure but did not provide support to developers wishing to execute an AI application in a secure environment. MPAI-AIF V2 responds to this requirement. As shown in *Figure* ***1***, the standard defines a Security Abstraction Layer (SAL). By accessing the SAL APIs, a developer can indeed create the required level of security with the desired functionalities.



*Figure 2 - Reference model of the MPAI AI Framework (MPAI-AIF) V2*

MPAI-AIF V2 is at the stage of Community Comments.

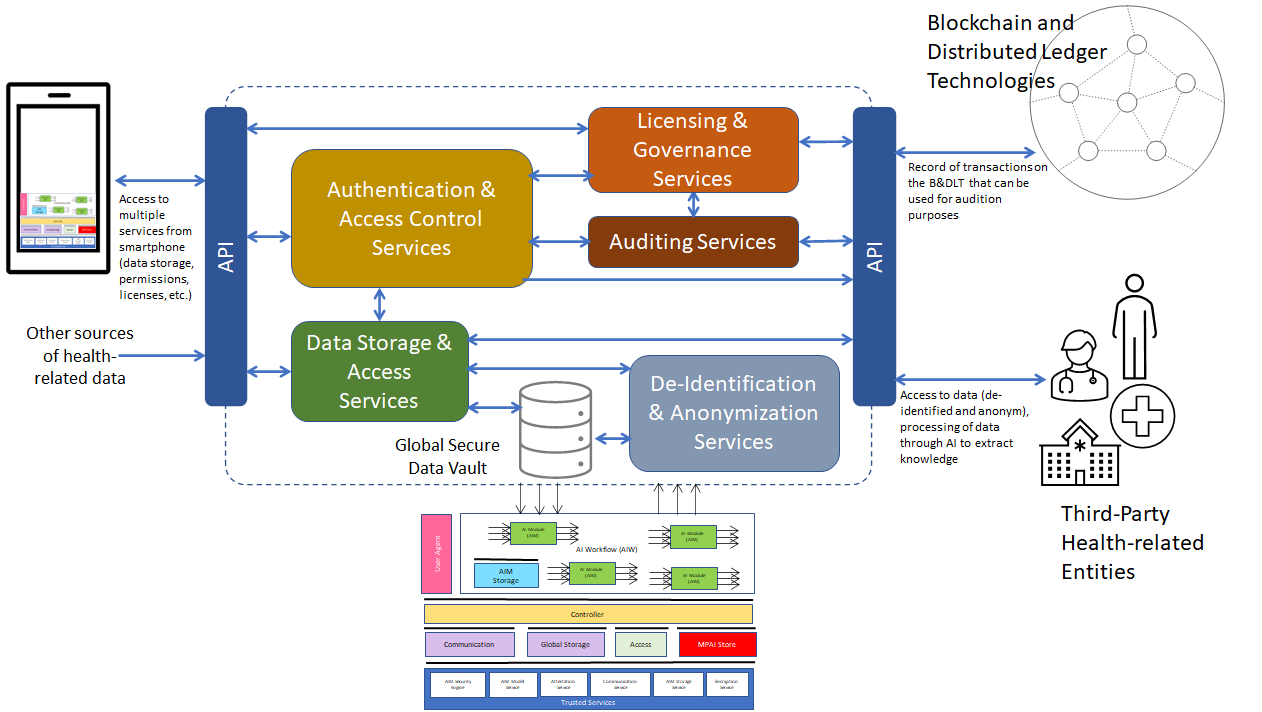
The collection of public documents is available [here](https://mpai.community/standards/mpai-aif/).

# MPAI-AIH

*Artificial Intelligence for Health data* (MPAI-AIH) is an MPAI project aiming to specify the interfaces and the relevant data formats of a system called AI Health Platform (AIH Platform) where:

1. End Users use handsets with an MPAI AI Framework (AI Health Frontends) to acquire and process health data.
2. An AIH Backend collects processed health data delivered by AIH Frontends with associated Smart Contracts specifying the rights granted by End Users.
3. Smart Contracts are stored on a blockchain.
4. Third Party Users can process their own and End User-provided data based on the relevant smart contracts.
5. The AIH Backend periodically collects the AI Models trained by the AIH Frontends while processing the health data, updates its AI Model and distributes it to AI Health Platform Frontends (Federated Learning).

This is depicted in *Figure 3* (for simplicity the security part of the AI Framework is not included).



*Figure 3 - MPAI-AIH Reference Model*

MPAI-AIH is at the Call for Technologies stage with responses due by October 19 at 23:59 UTC.

The collection of public documents is available [here](https://mpai.community/standards/mpai-aih/).

# MPAI-CAE

Context-based Audio Enhancement (MPAI-CAE) improves the user experience for several audio-related applications including entertainment, communication, teleconferencing, gaming, post-production, restoration etc. in a variety of contexts such as in the home, in the car, on-the-go, in the studio etc. using context information to act on the input audio content using AI.

## Version 1

*Figure 4* is the reference model of Unidirectional Speech Translation, a Use Case developed by Version 1.

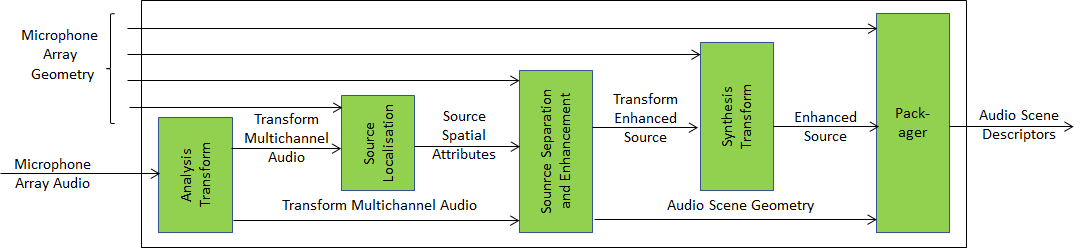


*Figure 4 - An MPAI-CAE Use Case: Emotion-Enhanced Speech*

The MPAI-AIF Technical Specification has been approved and is available [here](https://mpai.community/resources/).

## Version 2

MPAI has developed the specification of the Audio Scene Description Composite AIM as part of the MPAI-CAE V2 standard.



*Figure 5 - Audio Scene Description Composite AIM*

MPAI-CAE V2 is at the Technical Specification stage.

The collection of public documents is available [here](https://mpai.community/standards/mpai-cae/).

# MPAI-CUI

Compression and understanding of industrial data (MPAI-CUI) aims to enable AI-based filtering and extraction of key information to predict company performance by applying Artificial Intellig­ence to governance, financial and risk data. This is depicted in *Figure 6*.

A picture containing chart

Description automatically generated

*Figure 6 - The MPAI-CUI Use Case*

The collection of publicly available MPAI-CUI documents is [here](https://mpai.community/standards/mpai-cui/). The set of specifications composing the MPAI-CUI standard is available [here](https://mpai.community/resources/).

# MPAI-CAV

Connected Autonomous Vehicles (CAV) is a Use Case addressing the Connected Autonomous Vehicle (CAV) domain and the 5 main operating instances of a CAV:

*Human-CAV interaction (HCI)*, i.e., the CAV subsystem that responds to humans’ com­mands and queries, senses human activities in the CAV passenger compartment and activates other subsystems as required by humans or as deemed necessary by the identified conditions.

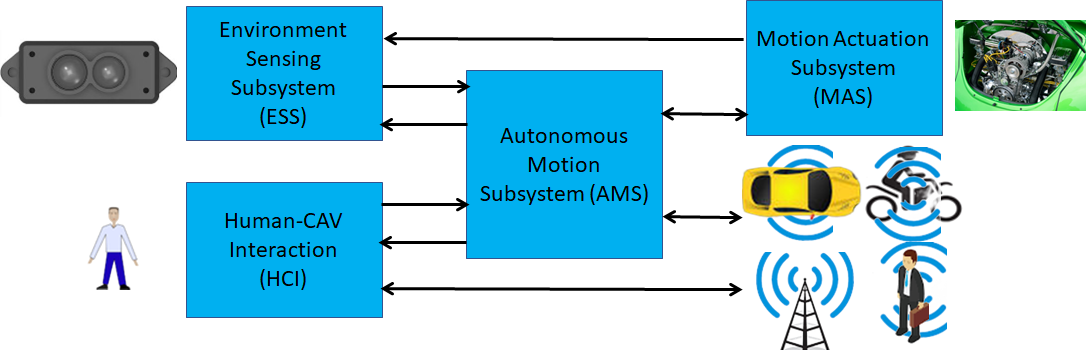
*CAV-Environment interaction*, i.e., the subsystem that acquires information from the physical environment via a variety of sensors.

*Autonomous Motion Subsystem (AMS)*, i.e., the CAV subsystem that uses different sources of information to instructs the CAV to reach the intended destination.

*CAV-Device Interaction (CDI)*, i.e., the subsystem that communicates with sources of external information, including other CAVs, Roadside Units (RSU), other vehicles etc.

*Motion Actuation Subsystem (MAS)*, i.e., the subsystem that operates and actuates the motion instructions in the physical world.

The interaction of the 5 subsystems in depicted in *Figure 7*.

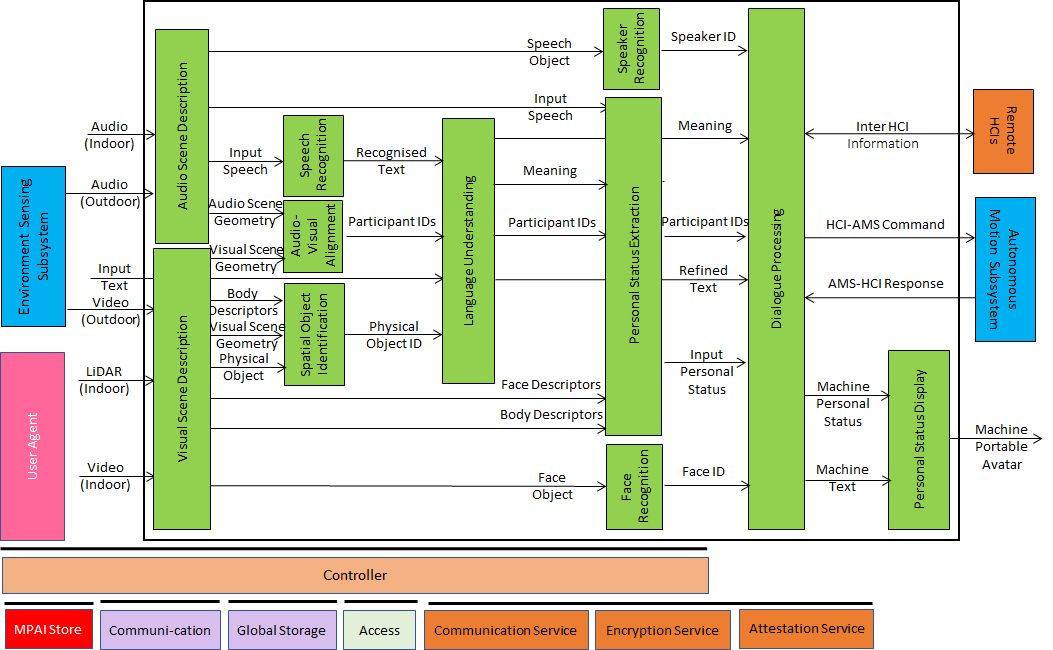


*Figure 7 -– The CAV subsystems*

MPAI-CAV – Architecture is published as *Technical Specification: Connected Autonomous Vehicle (MPAI-CAV) – Architecture* specifying the Architecture of a Connected Autonomous Vehicle (CAV) based on a Reference Model comprising:

1. A CAV broken down into Subsystems for each of which the following is specified:
   1. The Functions
   2. The input/output Data
   3. The Topology of Components
2. Each Subsystem broken down into Components of which the following is specified:
   1. The Functions
   2. The input/output Data.

*Figure 8* depicts the Reference Model of the Human-CAV Interaction.



*Figure 8 - Reference Model of the Human-CAV Interaction Subsystem*

The collection of public documents is available [here](https://mpai.community/standards/mpai-cav/).

# MPAI-EEV

There is consensus in the video coding research community that the so-called End-to-End (E2E) video coding schemes can yield significantly higher performance than those target, e.g., by MPAI-EVC. AI-based End-to-End Video Coding intends to address this promising area.

MPAI is extending the OpenDVC model [*Figure 9*]

Diagram

Description automatically generated

*Figure 9 - MPAI-EEV Reference Model*

The collection of public documents is available [here](https://mpai.community/standards/mpai-eev/).

# MPAI-EVC

AI-Enhanced Video Coding (MPAI-EVC) is a video compression stan­dard that substantially en­hances the performance of a traditional video codec by improving or replacing traditional tools with AI-based tools. Two approaches – Horizontal Hybrid and Vertical Hybrid – are envisaged. The **Vertical Hybrid** approach envisages an AVC/HEVC/EVC/VVC base layer plus an enhanced machine learning-based layer. This case can be represented by *Figure 10*.

Diagram

Description automatically generated

*Figure 10 - A reference diagram for the Vertical Hybrid approach*

The **Horizontal Hybrid** approach introduces AI based algorithms combined with trad­itional image/video codec, trying to replace one block of the traditional schema with a machine learn­ing-based one. This case can be described by *Figure 11* where green circles represent tools that can be replaced or enhanced with their AI-based equivalent.

Diagram

Description automatically generated

*Figure 11 - A reference diagram for the Horizontal Hybrid approach*

MPAI is engaged in the MPAI-EVC Evidence Project seeking to find evidence that AI-based technologies provide sufficient improvement to the Horizontal Hybrid approach. A second project on the Vertical Hybrid approach is being considered.

The collection of public documents is available [here](https://mpai.community/standards/mpai-evc/).

# MPAI-GSA

Integrative Genomic/Sensor Analysis (MPAI-GSA) uses AI to understand and compress the res­ult of high-throughput experiments combining genomic/proteomic and other data, e.g., from video, motion, location, weather, medical sensors.

*Figure 12* addresses the Smart Farming Use Case.

Diagram

Description automatically generated

*Figure 12 - An MPAI-GSA Use Case: Smart Framing*

The collection of public documents is available [here](https://mpai.community/standards/mpai-gsa/).

# MPAI-MCS

Mixed-Reality Collaborative (MPAI-MCS) Spaces is a project riding on the opportunities offered by emerging technologies enabling developers to deliver mixed-reality collaborative space (MCS) applications where biomedical, scientific, and industrial sensor streams and recordings are to be viewed. MCS systems use AI to achieve immersive presence, spatial maps (e.g., Lidar scans, inside-out tracking) rendering, and multiuser synchronis­ation etc.

The collection of public documents is available [here](https://mpai.community/standards/mpai-mcs/).

# MPAI-MMC

Multi-modal conversation (MPAI-MMC) aims to enable human-machine conversation that emul­ates human-human conversation in completeness and intensity by using AI.

## Version 1

The MPAI mission is to develop *AI-enabled data coding standards*. MPAI believes that its standards should enable humans to select machines whose internal operation they understand to some degree, rather than machines that are “black boxes” resulting from unknown training with unknown data. Thus, an implemented MPAI standard breaks up monolithic AI applications, yielding a set of interacting components with identified data whose semantics is known, as far as possible.

Technical Specification: Multimodal Conversation (MPAI-MMC) is an implementation of this vision for human-machine conversation. Five Use Cases have been developed for MPAI-MMC V1: Conversation with emotion, Multimodal Question Answering (QA) and 3 Automatic Speech Translation Use Cases.

*Figure 13* depicts the Reference Model of the Conversation with Emotion Use Case.

Graphical user interface

Description automatically generated with medium confidence

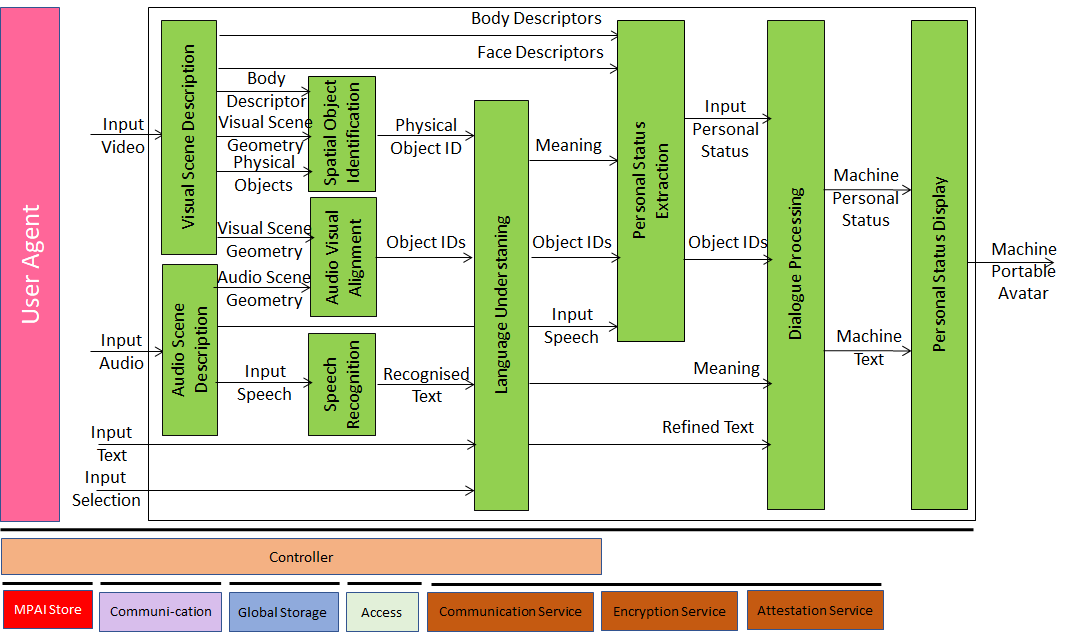
*Figure 13 - An MPAI-MMC V1 Use Case: Conversation with Emotion*

The MPAI-MMC Technical Specification V1.2 has been approved and is available [here](https://mpai.community/resources/).

## Version 2

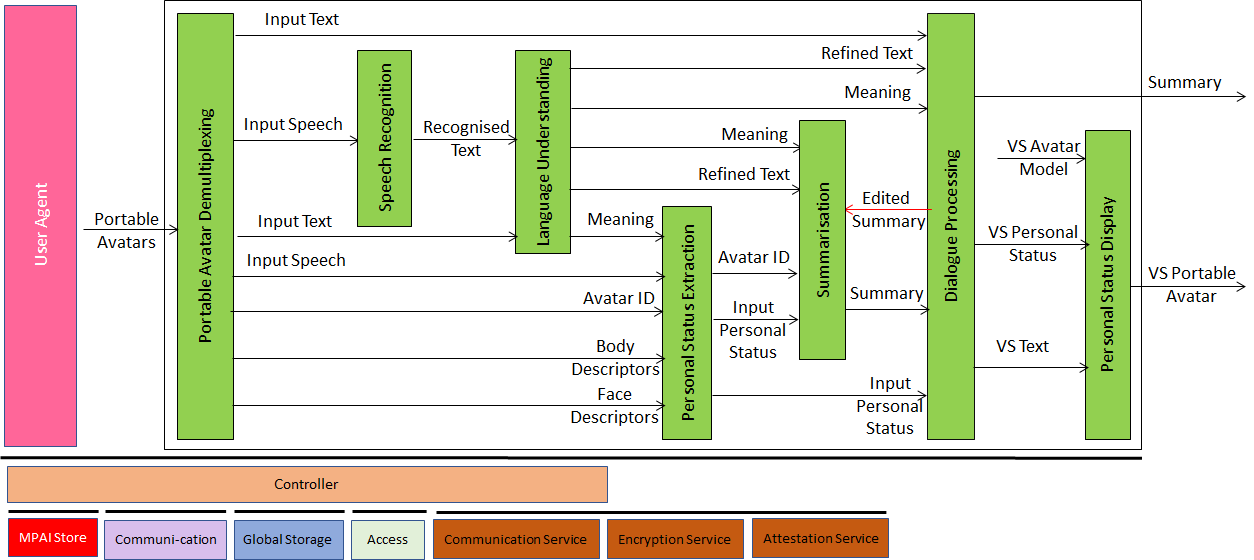
Extending the role of emotion as introduced in Version 1 of the standard, MPAI-MMC V2 introduces Personal Status, an internal status of humans that a machine needs to estimate and that it artificially creates for itself with the goal of improving its conversation with the human or, even with another machine. Personal Status is applied to MPAI-MMC specific Use Cases, such as Conversation about a Scene, Virtual Secretary for Videoconference, and Human-Connected Autonomous Vehicle Interaction.

Several new Use Cases have been specified for Multi-modal conversation V2 (MPAI-MMC V2). One of them is Conversation About a Scene (CAS). *Figure 14* is the reference model of the Conversation About a Scene (CAS) Use Case.

**

*Figure 14 - An MPAI-MMC V2 Use Case: Conversation with Personal Status*

*Figure 15* gives the Reference Model of a second use case: Virtual Secretary (used by the Avatar-Based Videoconference use case).



*Figure 15 - Reference Model of Avatar-Based Videoconference*

MPAI-MMC V2 is at the Technical Specification stage.

The collection of public documents is available [here](https://mpai.community/standards/mpai-cui/).

# MPAI-MMM

Metaverse is a word conveying different meaning to different persons. In this document the word metaverse is characterised as a system that captures data from the real world, processes it, and combines it with internally generated data to create virtual environments that users can interact with. System developers have made technology decisions that best responded to their needs, often without considering the choices that other developers might have made for similar purposes.

Recently, however, there have been mounting concerns that such metaverse “walled gardens” do not fully exploit the opportunities offered by current and expected technologies. Calls have been made to make metaverse instances “Interoperable”.

MPAI Metaverse Model (MMM) is an MPAI project targeting a series of Technical Reports and Specifications promoting Metaverse Interoperability. Two MPAI Technical Reports on Metaverse Functionalities and Functionality Profiles have laid down the groundwork. With the Technical Specification – MPAI Metaverse Model – Architecture, MPAI provides initial Interoperability tools by specifying the Functional Requirements of Processes, Items, Actions, and Data Types that allow two or more metaverse instances to Interoperate via a Conversion Service if they implement the Operation Model and produce Data whose Format complies with the Specification’s Functional Requirements.

*Figure* ***16*** depicts one aspect of the Specification where a Process in a metaverse instance requests a Process in another metaverse instance to perform an Action by relying on the instances’ Resolution Service.

A picture containing text, diagram, screenshot, font

Description automatically generated

*Figure 16 - Resolution and Conversion Services*

MPAI-MMM – Architecture is at the Technical Specification level.

The collection of public documents is available [here](https://mpai.community/standards/mpai-mmm/).

# MPAI-NNW

Neural Network Watermarking is a standard whose purpose is to enable watermarking technology providers to qualify their products by providing the means to measure, for a given size of the watermarking payload, the ability of:

1. The watermark inserter to inject a payload without deteriorating the NN performance.
2. The watermark detector to recognise the presence of the inserted watermark when applied to
   1. A watermarked network that has been modified (e.g., by transfer learning or pruning)
   2. An inference of the modified model.
3. The watermark decoder to successfully retrieve the payload when applied to
   1. A watermarked network that has been modified (e.g., by transfer learning or pruning)
   2. An inference of the modified model.
4. The watermark inserter to inject a payload at a measures computational cost on a given processing environment.
5. The watermark detector/decoder to detect/decode a payload from a watermarked model or from any of its inferences, at a low computational cost, e.g., execution time on a given processing environment.

The Neural Network Watermarking Technical Specification is at the Technology Specification level and the Reference Software Specification is published.

MPAI-NNW is being implemented in the MPAI-AIF as shown in

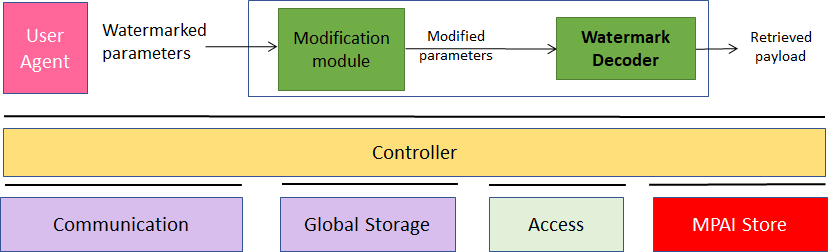


Figure 17 - MPAI-NNW implemented in MPAI-AIF

The collection of public documents is available [here](https://mpai.community/standards/mpai-nnw/).

# MPAI-OSD

Visual object and scene description is an MPAI project seeking to define a set of technologies for coordinated use in many use cases target of MPAI projects and standards. Examples are: Spatial Attitude, Point of View, Audio Scene Descriptors, Visual Scene Descriptors, Audio-Visual Scene Description.

MPAI-OSD is at the Technical Specification development stage.

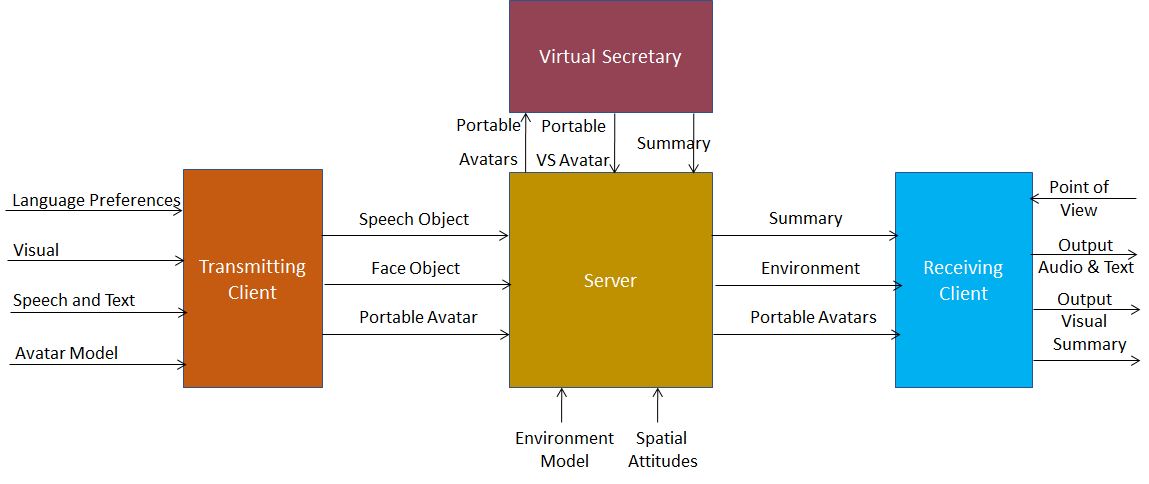
The collection of public documents is available [here](https://mpai.community/standards/mpai-osd/).

# MPAI-PAF

There is a long history of computer-created objects called “digital humans”, i.e., digital objects that can be rendered to show a human appearance. In most cases the underlying assumption of these objects has been that creation, animation, and rendering is done in a closed environment. Such digital humans had little or no need for standards. However, in a communication and even more in a metaverse context, there are many cases where a digital human is not constrained within a closed environment. For instance, a client sends may data to a remote client that should be able to unambiguously interpret and use the data to reproduce a digital human as intended by the transmitting client.

These new usage scenarios require forms of standardisation. Technical Specification: Portable Avatar Format (MPAI-PAF) is a first response to the need of a user wishing to enable their transmitting client to send data that a remote client can interpret to render a digital human, having the body movement and the facial expression representing their own movements and expression.

*Figure 18* is the system diagram of the Avatar-Based Videoconference Use Case enabled by MPAI-PAF.



*Figure 18 - End-to-End block diagram of Avatar-Based Videoconference*

MPAI-PAF is at Technical Specification stage.

The collection of public documents is available [here](https://mpai.community/standards/mpai-ara/).

# MPAI-SPG

Server-based Predictive Multiplayer Gaming (MPAI-SPG) aims to minimise the audio-visual and gameplay discontinuities caused by high latency or packet losses during an online real-time game. In case information from a client is missing, the data collected from the clients involved in a particular game are fed to an AI-based system that predicts the moves of the client whose data are missing. The same technologies provide a response to the need to detect who amongst the players is cheating.

*Figure 19* depicts the MPAI-SPG reference model including the cloud gaming model.

Graphical user interface

Description automatically generated

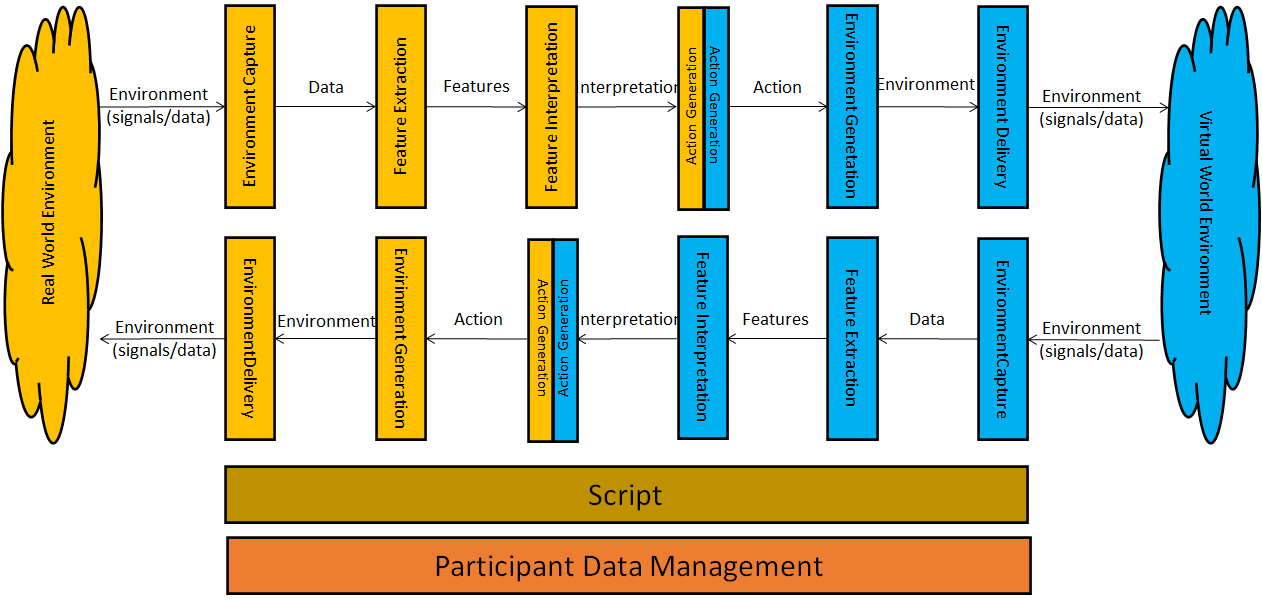
*Figure 19 - The MPAI-SPG Use Case*

The collection of public documents is available [here](https://mpai.community/standards/mpai-spg/).

# MPAI-XRV

XR Venues (MPAI-XRV) Theatrical stage performances such as Broadway theatres, musicals, dramas, operas, and other performing arts increasingly use video scrims, backdrops, and projection mapping to create digital sets rather than constructing physical stage sets. This allows animated backdrops and reduces the cost of mounting shows. The use of immersion domes – especially LED volumes – promises to surround audiences with virtual environments that the live performers can inhabit and interact with.

MPAI-XRV has developed a reference model that describes the components of the Real-to-Virtual-to-Real scenario depicted in *Figure 20.*



*Figure 20 - General Reference Model of the Real-to-Virtual-to-Real Interaction*

The MPAI XR Venues (XRV) – Live Theatrical Stage Performance project, a use case of MPAI-XRV intends to define AI Modules that facilitate setting up live multisensory immersive stage performances which ordinarily require extensive on-site show control staff to operate. With XRV it will be possible to have more direct, precise yet spontaneous show implementation and control that achieve the show director’s vision but free staff from repetitive and technical tasks letting them amplify their artistic and creative skills.

An XRV Live Theatrical Stage Performance can extend into the metaverse as a digital twin. In this case, elements of the Virtual Environment experience can be projected in the Real Environment and elements of the Real Environment experience can be rendered in the Virtual Environment (metaverse).

The figure shows how the XRV system captures the Real (stage and audience) and Virtual (metaverse) Environment, AI-processes the captured data, injects new components into the Real and Virtual Environments.

A screenshot of a computer

Description automatically generated

*Figure 21 - Reference Model of MPAI-XRV – Live Theatrical Stage Performance*

MPAI-XRV – Live Theatrical Stage Performance is at the Call for Technologies stage.

The collection of public documents is available [here](https://mpai.community/standards/mpai-xrv/).