|  |  |  |
| --- | --- | --- |
|  | | Moving Picture, Audio and Data Coding by Artificial Intelligence  www.mpai.community  **Public document** |
| **N1278** | 2023/07/12 | |
| **Source** | General Assembly #34 (MPAI-34) | |
| **Title** | MPAI Work Plan | |
| **Target** | MPAI Community | |

Contents

[1 Introduction 1](#_Toc127551594)

[2 MPAI-AIF 3](#_Toc127551595)

[2.1 Version 1 3](#_Toc127551596)

[2.2 Version 2 3](#_Toc127551597)

[3 MPAI-AIH 4](#_Toc127551598)

[4 MPAI-ARA 5](#_Toc127551599)

[5 MPAI-CAE 5](#_Toc127551600)

[5.1 Version 1 5](#_Toc127551601)

[5.2 Version 2 6](#_Toc127551602)

[6 MPAI-CUI 6](#_Toc127551603)

[7 MPAI-CAV 7](#_Toc127551604)

[8 MPAI-EEV 8](#_Toc127551605)

[9 MPAI-EVC 9](#_Toc127551606)

[10 MPAI-GSA 10](#_Toc127551607)

[11 MPAI-MCS 11](#_Toc127551608)

[12 MPAI-MMC 11](#_Toc127551609)

[12.1 Version 1 11](#_Toc127551610)

[12.2 Version 2 11](#_Toc127551611)

[13 MPAI-MMM 12](#_Toc127551612)

[14 MPAI-NNW 13](#_Toc127551613)

[15 MPAI-OSD 13](#_Toc127551614)

[16 MPAI-SPG 13](#_Toc127551615)

[17 MPAI-XRV 14](#_Toc127551616)

# Introduction

MPAI’s standards development is based on projects evolving through a workflow extending on 7 + 1 stages.

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **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-31)*

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **#** | **V** | **Work area** | **IC** | **UC** | **FR** | **CR** | **CfT** | **SD** | **CC** | **TS** | **RS** | **CT** | **PA** | **TR** |
| AIF | 1.1 | AI Framework |  |  |  |  |  |  |  | X | X | X |  |  |
| AIF | 2.0 | AI Framework |  |  |  |  |  | X |  |  |  |  |  |  |
| AIH |  | AI Health Data |  |  | X |  |  |  |  |  |  |  |  |  |
| ARA | 1.0 | Avatar Representation & Animation |  |  |  |  |  | X |  |  |  |  |  |  |
| CAE | 1.4 | Context-based Audio Enhancement |  |  |  |  |  |  |  | X |  |  |  |  |
| CAE | 2.0 | Context-based Audio Enhancement |  |  |  |  |  |  |  | X |  |  |  |  |
| CAV |  | Connected Autonomous Vehicles |  |  | X |  |  |  |  |  |  |  |  |  |
| CUI | 1.1 | Compression & Under­stan­ding of Ind. 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 |
| NNW | 1.0 | Neural Network Watermarking |  |  |  |  |  |  | X | X |  |  |  |  |
| OSD |  | Visual Object and Scene Description |  | X |  |  |  |  |  |  |  |  |  |  |
| SPG |  | Server-based Predictive Multiplayer Gaming |  |  | X |  |  |  |  |  |  |  |  |  |
| XRV |  | XR Venues |  |  | X |  |  |  |  |  |  |  |  |  |

# MPAI-AIF

Artificial Intelligence Framework (MPAI-AIF) enables creation and automation of mixed Artif­icial Intelligence – Machine Learning – Data Processing workflows for the application areas cur­rently considered by the MPAI work plan.

## Version 1

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

Diagram

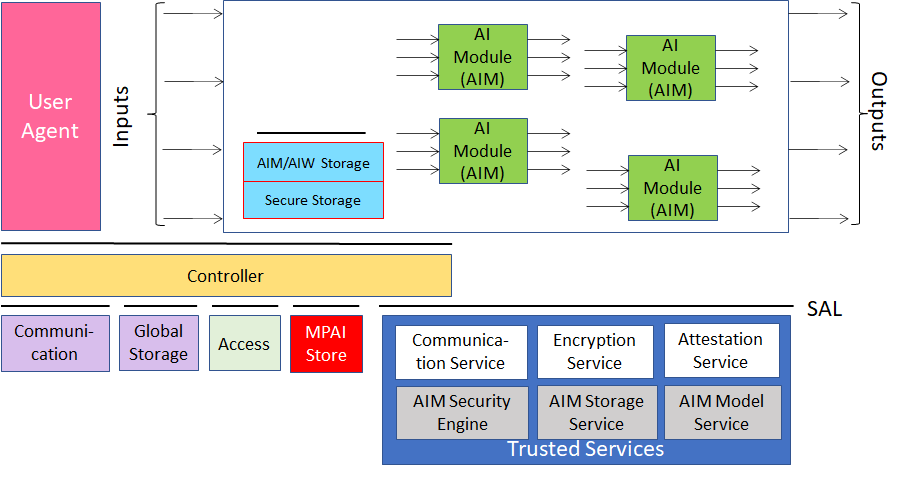
Description automatically generated

*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 is available [here](https://mpai.community/resources/).

## Version 2

MPAI has developed Use Cases and Requirements for MPAI-AIF V2 adding security support to MPAI-AIF V1.



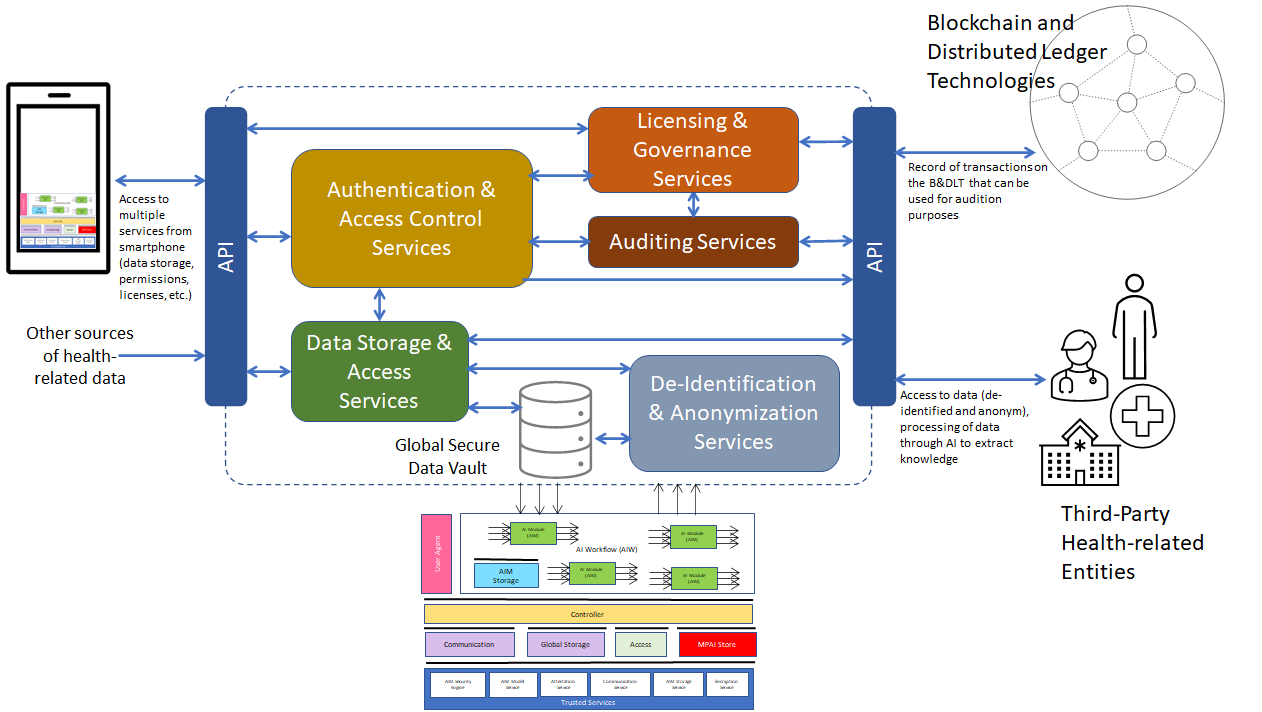
*Figure 2 - Reference model of the MPAI AI Framework (MPAI-AIF) V2*

The Call for Technologies has been published, responses received and the development of V2 is under way.

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 addressing the secure collection, AI-based processing and secure access to Health data (*Figure 3*).



*Figure 3 - MPAI-AIH Reference Model*

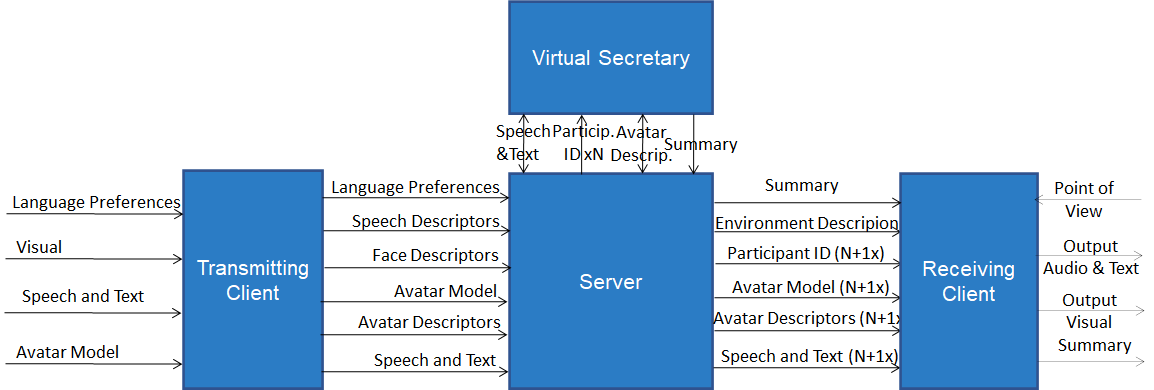
The AIH System includes Front Ends (mobile apps) and a Back End. Date is processed by the Front Ends and uploaded to the Back End with an associate smart contract that determines the terms and conditions of use. The Back End processes the data and can make it available to Third Parties based on the terms and conditions of the smart contract. The neural network models of the Front Ends are collected, and a new model created by Federated Learning technicques and redistributed to the Front Ends.

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

# MPAI-ARA

Avatar Representation and Animation (MPAI-ARA) specifies the technologies enabling the implementation of the Avatar-Based Videoconference Use Case specified in the Avatar-Based Videoconference Use Case (*Figure 4*), specifically:

1. A Digital Environment populated with non-human audio-visual objects.
2. A Digital Human Model.
3. The Representation of the motion a human.
4. The Animation of a Digital Human Model.
5. The Representation of the features of a human.



*Figure 4 - End-to-End block diagram of Avatar-Based Videoconference*

The development of the MPAI-ARA standard is under way.

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

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



*Figure 5 - 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 Use Cases and Requirements for Version 2 as part of the MPAI-MMC V2 standard. Responses have been received and the development of V2 is under way.

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.

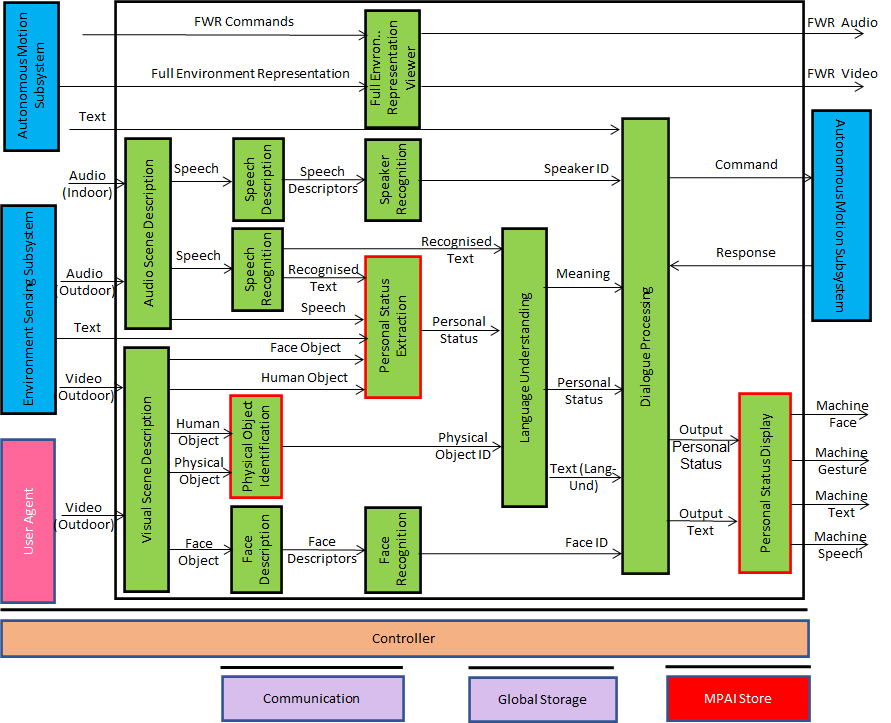
The interaction of the 5 subsystems in depicted in *Figure 7*.



*Figure 7 -– The CAV subsystems*

Requirements for the Human-CAV Interaction subsystem (*Figure 8*) have been developed and used in the MPAI-MMC V2 Call for Technologies.

The MPAI-CAV Use Cases and Functional Requirements have been developed.



*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 envigaes 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

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*

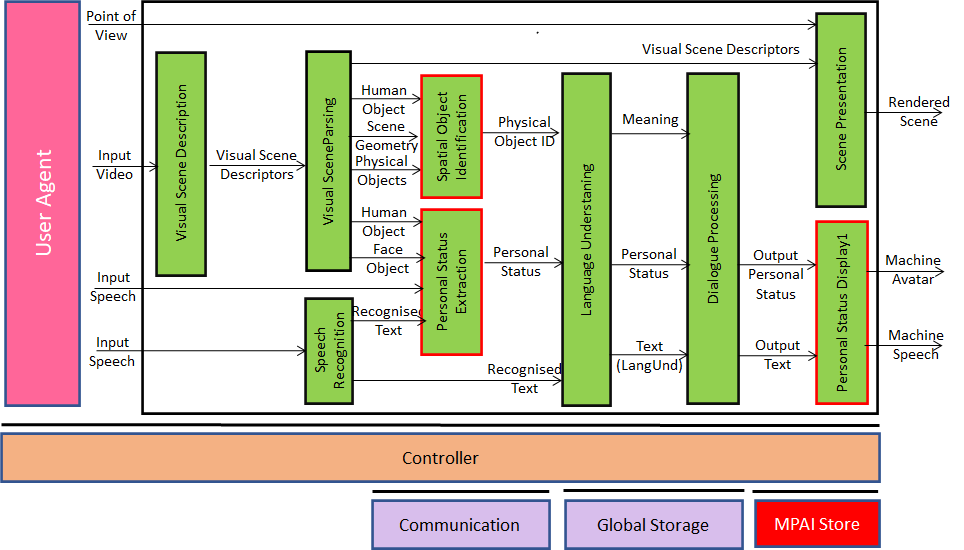
The MPAI-MMC Technical Specification V1.2 has been approved and is available [here](https://mpai.community/resources/).

## Version 2

Five new Use Cases have been identified for Multi-modal conversation V2 (MPAI-MMC V2).

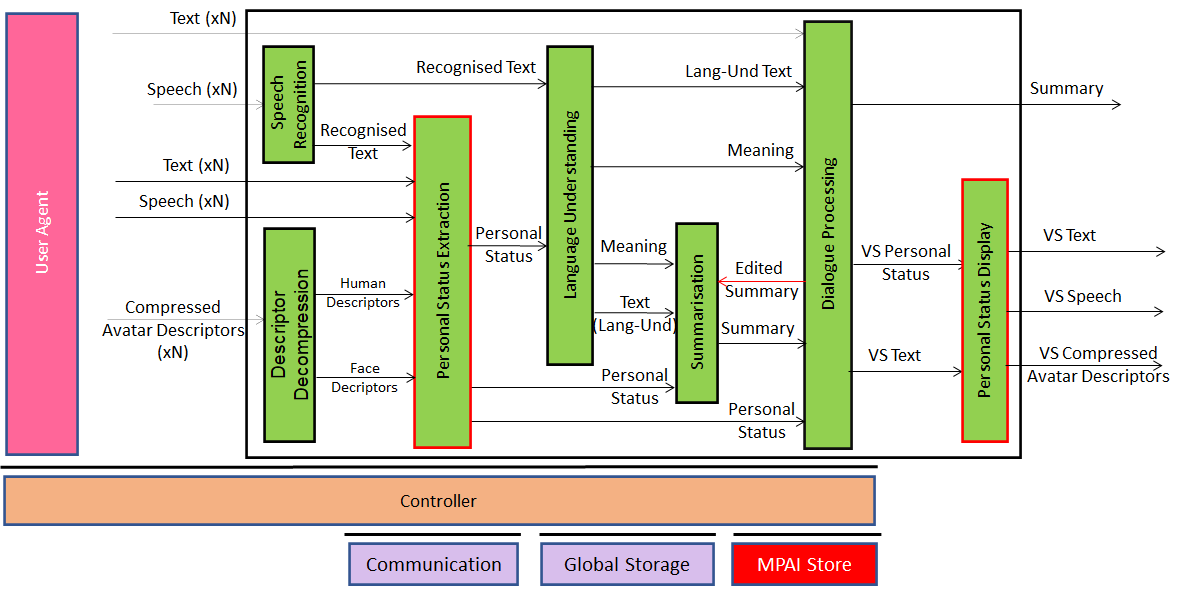
Conversation About a Scene (CAS) and Avatar-Based Videoconference (ABV).

*Figure 14* is the reference model of the Conversation About a Scene (CAS) Use Case.

**

*Figure 14 - An MPAI-MMC V2 Use Case: Conversation About a Scene*

*Figure 15* gives the Reference Model of the Virtual Secretary of the MPAI-MCS Avatar-Based Videoconference (MCS-ABV).



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

MPAI-MMC V2 is under development.

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

# MPAI-MMM

MPAI Metaverse Model (MMM) is an MPAI p;roject targeting a series of Technical Reports and Specifications promoting Metaverse Interoperability. This is the planned list of documents:

1. Functionalities
2. Functionality Profiles
3. Metaverse Architecture
4. Data Type Functional Requirements
5. Common Metaverse Specifications - Table of Contents
6. Initial Technology mapping to the Common Metaverse Specifications

Document #1 has been https://mpai.community/standards/mpai-mmm/mpai-metaverse-model/. Document #2 is under development.

# 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 published.

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

# MPAI-OSD

Visual object and scene description is a collection of Use Cases sharing the goal of describe visual object and locate them in the space. Scene description includes the usual des­cription of objects and their attributes in a scene and the semantic description of the objects.

Unlike proprietary solutions that address the needs of the use cases but lack interoperability or force all users to adopt a single technology or application, a standard representation of the ob­jects in a scene allows for better satisfaction of the requirements.

MPAI has developed MPAI-OSD related requirements for

1. MMC-PSE
2. MMC-PSD
3. MMC-CAS
4. CAV-HCI
5. ARA-ABV

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

# 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 16* depicts the MPAI-SPG reference model including the cloud gaming model.

Graphical user interface

Description automatically generated

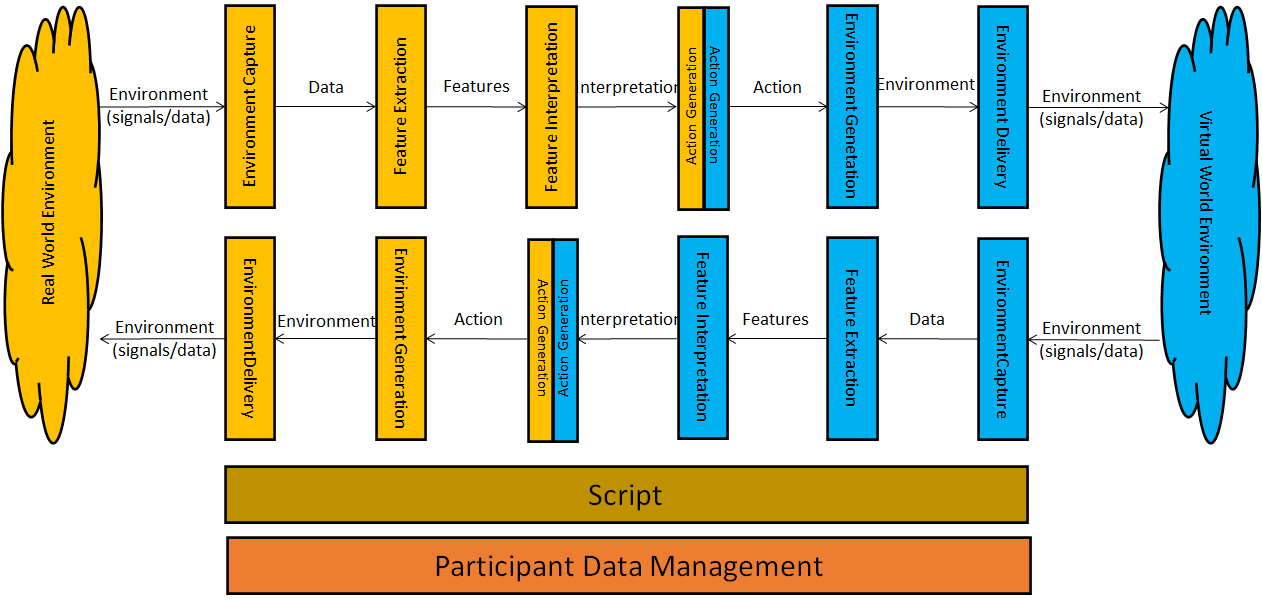
*Figure 16 - The MPAI-SPG Use Case*

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

# MPAI-XRV

XR Venues (MPAI-XRV) is an MPAI project addressing a multiplicity of use cases enabled by AR/VR/MR (XR) and enhanced by Artificial Intelligence technologies. The word venue is used as a synonym to Environment, and can be both real and virtual.

MPAI-XRV has developed a reference model that describes the components of the Real-to-Virtual-to-Real scenario depicted in *Figure 17.*



*Figure 17 - General Reference Model of the Real-to-Virtual-to-Real Interaction*

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