|  |  |
| --- | --- |
|  | Moving Picture, Audio and Data Coding by Artificial Intelligencewww.mpai.community |

|  |
| --- |
| **Public document** |
| **N353** | 2021/09/30 |
| **Source** | General Assembly #12 (MPAI-12) |
| **Title** | MPAI workplan  |
| **Target** | MPAI Community |

# Introduction

MPAI’s standards development is based on projects evolving through a workflow extending on 6 + 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 Reqs | Identification of the functional requirements that the standard in­cluding the Use Case should satisfy. |
| 3 | CR | Commercial Reqs | 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. |

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

The stages of currently (MPAI-12) active MPAI projects are graphically represented by *Figure 1*.

TS: Technical Specification, RS: Reference Software, CT: Conformance Testing, PA: Performance Assessment



*Figure 1 – Snapshot of the MPAI work plan*

# Approved standards

## MPAI-CUI

## 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 2*.

**

*Figure 2* – *The MPAI-CUI Use Case*

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

# Areas at stage 6 (SD)

## 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.

So far, 5 Use Cases have been identified for MPAI-MMC: Conversation with emotion, Multimodal Question Answering (QA) and 3 Automatic Speech Translation Use Cases.

*Figure 3* addresses the Conversation with Emotion Use Case.

**

*Figure 3* – *An MPAI-MMC Use Case: Conversation with emotion*

MPAI has initiated work on Audio on the Go (AOG), a new use case for MPAI-CUI V2.

The collection of public documents is available [here](https://mpai.community/standards/mpai-cui/). The MPAI-MMC Technical Specification has been approved and is available [here](https://mpai.community/resources/).

## 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. MPAI-AIF will be extended to support new applic­ations areas, if the need will arise. *Figure 4* shows the general MPAI-AIF Reference Model.



*Figure 4 – Reference model of the MPAI AI Framework*

Stage 6 of the Technical Specification is expected to be reached in November 2021. The collection of public documents is available [here](https://mpai.community/standards/mpai-aif/).

## 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.

*Figure 5* is the reference model of Unidirectional Speech Translation.



*Figure 5* – *An MPAI-CAE Use Case: Emotion-Enhanced Speech*

Stage 6 of the Technical Specification is expected to be reached in November 2021. The collection of public documents is available [here](https://mpai.community/standards/mpai-cae/).

# Areas at stage 2 (FR)

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



*Figure 6* – *The MPAI-CUI Use Case*

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

## 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 7*.



*Figure 7* – *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 8* where green circles represent tools that can be replaced or enhanced with their AI-based equivalent.

**

*Figure 8* – *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-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 9*



*Figure 9 – The CAV subsystems*

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

## Mixed-Reality Collaborative Spaces

New technologies are emerging which equip developers to deliver mixed-reality collaborative space (MCS) scenarios where biomedical, scientific, and industrial sensor streams and recordings are to be viewed. Artificial Intelligence can be utilized throughout MCS systems for immersive presence, spatial maps (e.g. Lidar scans, inside-out tracking) rendering, and multiuser synchronis­ation etc.

*Figure 10* depicts one Use Case being considered “Avatar videoconference in a local 3D audio-visual space”



*Figure 10 – An avatar-based videoconference model (participant)*

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

## 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 11* addresses the Smart Farming Use Case.



*Figure 11* – *An MPAI-GSA Use Case: Smart Framing*

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

# Areas at stage 1 (UC)

## 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 satifaction of the requirements.

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

# Areas at stage 0 (IC)

## 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.

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