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

|  |
| --- |
| **Public document** |
| N85 | 2020/12/16 |
| Source | General Assembly (MPAI-3) |
| Title | MPAI workplan  |
| Target | Public document |

# Introduction

MPAI develops its standard using a 6-stage workflow

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **Acr** | **Name** | **Description** |
| 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 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 requirements |
| 5 | SD | Standard development | Development of the standard in a specific Development Committee (DC) |
| 6 | MS | MPAI standard | The standard has been successfully completed and all Members have made the appropriate declarations |

# Areas at stage 4 (CT)

## MPAI-AIF

Artificial Intelligence Framework (MPAI-AIF) enables creation and automation of mixed ML-AI-DP processing and inference workflows for the areas of work currently considered at stages 1, 2 and 3 of the MPAI work plan.

The said areas of work share the notion of an environment (the Framework) that includes 6 com­ponents – Management and Control, Execution, AI Modules (AIM), Communication, Storage and Access. AIMs are connected in a variety of topologies and executed under the super­vision of Management and Control. AIMs expose standard interfaces that make them re-usable in different applications. *Figure 1* shows the general MPAI-AIF Reference Model.

Approved MPAI documents supporting the MPAI-AIF project at the current stage are [1], [2], [3], [4].



*Figure 1 – Reference model of the MPAI AI Framework*

# Areas at stage 3 (CR)

## 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, processing such content via AIMs, and may deliver the processed output via the most appropriate protocol.

So far, MPAI-CAE has been found applicable to 11 usage examples, for 4 of which the definition of AIM interfaces is at an advanced stage: Emotion enhanced speech, Audio Recording Preservation, Enhanced Audioconference Experience and Audio-on-the-go. *Figure 2* addresses the Emotion enhanced speech Use Case.

Approved MPAI documents supporting the MPAI-CAE work area are [5] and [6].



*Figure 2* – *An MPAI-CAE Use Case: Emotion-enhanced speech*

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

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

Approved MPAI documents supporting the MPAI-GSA work area are [7] and [8].

**

*Figure 3* – *An MPAI-CAE usage example: Conversation with emotion*

# Areas at stage 2 (FR)

## MPAI-CUI

Compression and understanding of industrial data (MPAI-CUI) aims to enable AI-based filtering and extraction of key information from the flow of data that combines data produced by companies and external data (e.g., data on vertical risks such as seismic, cyber etc.)

MPAI-CUI requires standardisation of all data formats to be fed into an AI machine to extract information that is relevant to the intended use. Converted data undergo a further conversion and are then fed to specific neural networks. This is depicted in *Figure 9*.

Approved MPAI document supporting the MPAI-CUI work area are [9] and [10].

**

*Figure 4* – *The MPAI-CUI Use Case*

## MPAI-GSA

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

So far, MPAI-GSA has been found applicable to 7 usage examples ranging from personalised medicine to smart farming, for 2 of which the definition of AIM interfaces is under way: Personalised and Integrative Genomics and Automated Analysis of Animal Behaviour.

Approved MPAI documents supporting the MPAI-GSA work area are [5] and [6].

*Figure 5* addresses the Use Case Smart Farming.

Approved MPAI document supporting the MPAI-GSA work area are [11] and [12].



*Figure 5* – *An MPAI-GSA Use Case: Smart Framing*

## 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 **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 6* where green circles represent tools that can be replaced or enhanced with their AI-based equivalent.

**

*Figure 6* – *A reference diagram for the Horizontal Hybrid approach*

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*

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.

Approved MPAI documents supporting the MPAI-EVC work area are [13], [14], [15] and [16].

## 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.

Approved MPAI document supporting the MPAI-EVC work area is [17].



*Figure 8* – *Identification of MPAI-SPG standardisation area*

# Areas at stage 1 (UC)

## MPAI-OSD

Visual object and scene description addresses the “scene description” components of several use cases (Multiplayer online gaming ME.MP-09, Person matching ME.MP-11, Tracking game player’s movements ME.MP-12, AI-assisted driving TP.MP-01, Correct Posture HC.MP-02, Integrative genomic/video experiments ST.OD-06). 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.

Approved MPAI document supporting a potential MPAI-OSD work area is [18].

# Other possible areas

Several potential areas for standardisation are likely to emerge from [19]. A selection of such promising area has been derived from [19]

## Anomalous service access

A machine that has learnt "typical" service access values for a particular service provider can detect attempts beyond "typical" values.

## Anomalous vibrations

A machine learns from the data generated by inertial sensors (accelerometer with gyroscope) to distinguish between regular and anomalous vibrations.

## Vision-to-Sound Transformation

It is possible to give a spatial representation of an image that visually impaired people can hear with two headphones as a localization and description medium. It is a conversion (compression) technique from one space to a different interpretation space.

# References

[1] MPAI Application Note #4 - MPAI-AIF, N37

[2] Use Cases & Functional Requirements of MPAI-AIF, N74

[3] MPAI-AIF Call for Technologies, N100

[4] MPAI-AIF Framework Licence, N101

[5] MPAI Application Note #1 Rev. 1 - MPAI-CAE, N34

[6] Use Case, Requirements and candidate technologies for MPAI-CAE CfT, N96

[7] MPAI Application Note #6 - MPAI-CAE, N36

[8] Use Case, Requirements and candidate technologies for MPAI-MMC CfT, N97

[9] MPAI Application Note #7 - MPAI-CUI, N63

[10] MPAI-CUI Use Cases and Functional Requirements work programme, N95

[11] MPAI Application Note #2 - MPAI-GSA, N35

[12] MPAI-GSA Use Cases and Functional Requirements work programme, N72

[13] MPAI Application Note #3 R1 - MPAI-EVC, N61

[14] MPAI-EVC Use Cases and Requirements, N92

[15] Collaborative Evidence Conditions for MPAI-EVC Evidence Project Rev.1, N69

[16] Operational Guidelines for MPAI-EVC Evidence Project, N70

[17] MPAI Application Note #5 R1 - MPAI-SPG, N62

[18] MPAI Application Note #8 - MPAI-OSD, N93

[19] MPAI Use Case Rev2.0, N46