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

|  |
| --- |
| **Public Document** |
| **N475** | 2021/12/22 |
| **Source** | Requirements (CAV) and MMC-DC |
| **Title** | MMC-HCI Use Case and Functional Requirements WD0.4 |
| **Target** | MPAI-14 |

[Abstract 2](#_Toc90981819)

[1 Introduction 2](#_Toc90981820)

[2 Scope of the Use Cases 3](#_Toc90981821)

[3 Terms and Definitions 4](#_Toc90981822)

[4 References 5](#_Toc90981823)

[4.1 Normative References 5](#_Toc90981824)

[4.2 Informative References 6](#_Toc90981825)

[5 Use Cases 6](#_Toc90981826)

[5.1 Use Case #6 – The Human-CAV Interaction AIW (HCI) 6](#_Toc90981827)

[5.1.1 Use Case description 6](#_Toc90981828)

[5.1.2 Reference architecture 7](#_Toc90981829)

[5.1.3 Input and output data 7](#_Toc90981830)

[5.1.4 AI Modules 8](#_Toc90981831)

[6 Functional Requirements 9](#_Toc90981832)

[6.1 Human-CAV Interaction 9](#_Toc90981833)

[6.1.1 I/O Data summary 9](#_Toc90981834)

[6.1.2 Audio 10](#_Toc90981835)

[6.1.3 Verbal Interaction 10](#_Toc90981836)

[6.1.4 Concept Expression (Face) 11](#_Toc90981837)

[6.1.5 Concept Expression (Speech) 11](#_Toc90981838)

[6.1.6 Emotion 11](#_Toc90981839)

[6.1.7 Face identity 11](#_Toc90981840)

[6.1.8 Face Objects 11](#_Toc90981841)

[6.1.9 Full World Representation 12](#_Toc90981842)

[6.1.10 Full World Representation commands 12](#_Toc90981843)

[6.1.11 Intention 12](#_Toc90981844)

[6.1.12 Meaning 12](#_Toc90981845)

[6.1.13 Object Identifier 12](#_Toc90981846)

[6.1.14 Speaker Identity 12](#_Toc90981847)

[6.1.15 Text 13](#_Toc90981848)

[6.1.16 Video 13](#_Toc90981849)

[7 Data privacy 13](#_Toc90981850)

[7.1 Human-CAV Interaction (HCI) 13](#_Toc90981851)

[Annex 1 – MPAI-wide terms and definitions (Normative) 14](#_Toc90981852)

[Annex 2 - Notices and Disclaimers Concerning MPAI Standards (Informative) 17](#_Toc90981853)

[Annex 3 – The Governance of the MPAI Ecosystem (Informative) 19](#_Toc90981854)

# Abstract

*Technical Specification: Multimodal Conversation* (MPAI-MMC) V1 is a standard comprising the Conversation with Emotion and Multimodal Question Answering Use Cases and 3 Speech Translation Use Cases.

*Connected Autonomous Vehicles* (CAV) is an MPAI standard project. A CAV is subdivided in 5 subsystems For the purpose of standardisation. The Function of one of them called Human-CAV Interaction (HCI) deals with the technology-rich scenario of the ways a human and a CAV interact.

MMC-HCI is the 6th MPAI-MMC Use Case.

# Introduction

In recent years, Artificial Intelligence (AI) and related technologies have been applied to a broad range of applications, have started affecting the life of millions of people and are expected to do so even more in the future. As digital media standards have positively influenced industry and billions of people, so AI-based data coding standards are expected to have a similar positive impact. Indeed, research has shown that data coding with AI-based technologies is generally *more efficient* than with existing technologies for, e.g., compression and feature-based description.

However, some AI technologies may carry inherent risks, e.g., in terms of bias toward some classes of users. Therefore, the need for standardisation is more important and urgent than ever.

The international, unaffiliated, not-for-profit MPAI – Moving Picture, Audio and Data Coding by Artificial Intelligence Standards Developing Organisation has the mission to develop *AI-enabled data coding standards*. MPAI Application Standards enable the development of AI-based products, applications and services.

As a part of its mission, MPAI has developed standards operating procedures to enable users of MPAI implementations to make informed decision about their applicability. Central to this is the notion of Performance, defined as a set of attributes characterising a reliable and trustworthy implementation.

For the aforementioned reasons, to fully achieve the MPAI mission, Technical Specifications must be complemented by an ecosystem designed, created and managed to underpin the life cycle of MPAI standards through the steps of specification, technical testing, assessment of product safety and security, and distribution.

In the following, Terms beginning with a capital letter are defined in *Table 1* if they are specific to this Standard and in *Table 5* if they are common to all MPAI Standards.

The MPAI Ecosystem is specified in [1]. It is composed of:

* MPAI as provider of Technical, Conformance and Performance Specifications.
* Implementers of MPAI standards.
* MPAI-appointed Performance Assessors.
* The MPAI Store which takes care of secure distribution of validated Implementations.

The common infrastructure enabling implementation of MPAI Application Standards and access to the MPAI Store is the AI Framework (AIF) Standard (MPAI-AIF), specified in this document.

*Figure 1* depicts the MPAI-AIF Reference Model under which Implementations of MPAI Applic­ation Standards and user-defined MPAI-AIF conforming applications operate.

An AIF Implementation allows execution of AI Workflows (AIW), composed of basic processing elements called AI Modules (AIM).



*Figure 1 – The AI Framework (AIF) Reference Model and its Components*

MPAI Application Standards normatively specify Syntax and Semantics of the input and output data and the Function of the AIW and the AIMs, and the Connections between and among the AIMs of an AIW.

In particular, an AIM is defined by its Function and data, but not by its internal architecture, which may be based on AI or data processing, and implemented in software, hardware or hybrid software and hardware technologies.

MPAI defines Interoperability as the ability to replace an AIW or an AIM Implementation with a functionally equivalent Implementation. MPAI also defines 3 Interoperability Levels of an AIW that executes an AIW. The AIW may have 3 Levels:

*Level 1 –* Implementer-specific and satisfying the MPAI-AIF Standard.

*Level 2 –* Specified by an MPAI Application Standard.

*Level 3 –* Specified by an MPAI Application Standard and certified by a Performance Assessor.

MPAI offers Users access to the promised benefits of AI with a guarantee of increased transparency, trust and reliability as the Interoperability Level of an Implementation moves from 1 to 3. Additional information on Interoperability Levels is provided in Annex 3.

# Scope of the Use Cases

A Connected Autonomous Vehicle (CAV) is a mechanical system capable of executing a command to move its body guided by an analysis of data produced by a range of sensors exploring the environment and information transmitted by other sources in range, e.g., CAVs and roadside units (RSU).

MPAI has subdivided a CAV in 5 main subsystems:

*Human-CAV interaction (HCI)*

Recognises the human CAV rights holder

Responds to humans’ commands and queries

Provides extended environment representation (called Full World Repres­entation) for humans to enjoy

Senses human activities during the travel

Activates other subsystems as required by humans or as deemed necessary by the identified conditions.

*Environment Sensing Subsystem (EDS)*

Acquires information from the physical environment via a variety of sensors

Develops the best environment representation (called Basic World Representation).

*Autonomous Motion Subsystem (AMS)*

Computes the Route to destination

Uses dif­ferent sources of information – CAV sensors, other CAVs and transmitting units – to produce a Full World Representation

Gives command that drive the CAV to the intended destination.

*Motion Actuation Subsystem (MAS)*

Provides environment information

Actuates motion commands in the environment.

The 4 subsystems and their interactions are depicted in *Figure 2*.



*Figure 2 – The CAV subsystems*

The following high-level workflow illustrates the CAV operation envisaged by this docum­ent.

A *human* with appropriate credentials requests the CAV, via Human-CAV Interaction, to take the human to a given place.

*Human-CAV Interaction* authenticates the human, interprets the request and passes a command to the *Autonomous Motion Subsystem*.

*Autonomous Motion Subsystem*:

Requests *Environment Sensing Subsystem* to provide the current Pose.

Computes the Route.

Issues the start command.

*Environment Sensing Subsystem* computes and sends Basic World Representations to the *Autonomous Motion Subsystem.*

*Autonomous Motion Subsystem*:

Receives and processes data broadcasted by external sources (CAVs, RSU etc.).

Computes the Full World Representation.

Shares the CAV’s Full World Representation with CAVs in range.

Computes a Path.

Issues commands to the *Motion Actuation Subsystem* to move the CAV accordingly.

While the CAV moves, *Humans*

Hold conversation with *Human-CAV Interaction* and possibly other hum­ans on board.

Issue commands.

Request views of the environment (Full World Representation) etc.

Interact with other CAVs.

# Terms and Definitions

The terms used in this document whose first letter is capital have the meaning defined in *Table 1*.

*Table 1 – Table of terms and definitions*

|  |  |
| --- | --- |
| **Term** | **Definition** |
| Basic World Representation | A description of the Environment based on the CAV sensors and an Offline Map. |
| Command | High-level instructions whose execution allows a CAV to reach a Goal. |
| Connected Autonomous Vehicle | A vehicle capable to autonomously reach an assigned target by understan­ding human utterances, planning a route, sensing and interpreting the environment, exchanging information with other CAVs and acting on the CAV’s motion subsystem. |
| Decision horizon | The estimated time between the current state and the Goal. |
| Emotion | An attribute that indicates an emotion out of a finite set of Emotions |
| Emotion Grade | The intensity of an Emotion |
| Emotion Recognition | An AIM that decides the final Emotion out of Emotions from different sources |
| Environment | The portion of the world within the Decision horizon. |
| Full World Representation | A representation of the Environment using the CAV’s and other CAVs’ Basic World Representations. |
| Goal | The planned State at a future time.  |
| Intention | Intention is the result of a question analysis that denotes information on the input question. |
| Language Understanding | An AIM that analyses natural language as Text to produce its meaning and emotion included in the text |
| Meaning | Information extracted from the input text such as syntactic and semantic information |
| Offline Map | An offline-created map of a location and associated metadata. |
| Question Analysis | An AIM that analyses the meaning of a question sentence and determines its Intention  |
| Question Answering | An AIM that analyses the user’s question and produces a reply based on the user’s Intention  |
| Speech Recognition  | An AIM that converts speech to Text |
| Speech Synthesis | An AIM that converts Text or concept to speech |
| State | Pose, Velocity and Acceleration of a CAV at a given time. |
| Text | A collection of characters drawn from a finite alphabet |
| Translation | An AIM that converts Text in a source language to Text in a target language |
| Video analysis | An AIM that extracts features from video |
| Way Point | A point 𝑤𝑖 given as a coordinate pair (𝑥𝑖, 𝑦𝑖), in an Offline Map |

# References

## Normative References

This document references the following normative documents:

1. Technical Specification: The Governance of the MPAI Ecosystem V1.
2. Technical Specification: AI Framework (MPAI-AIF), MPAI document N359.
3. Technical Specification: Technical Specification: Multimodal Conversation (MPAI-MMC) V1; published at https://mpai.community/standards/resources/1.
4. Draft Technical Specification: Context-based Audio Enhancement (MPAI-CAE) VI, to be published at https://mpai.community/standards/resources/.
5. Universal Coded Character Set (UCS): ISO/IEC 10646; December 2020
6. ISO/IEC 14496-10; Information technology – Coding of audio-visual objects – Part 10: Advanced Video Coding.
7. ISO/IEC 23008-2; Information technology – High efficiency coding and media delivery in heterogeneous environments – Part 2: High Efficiency Video Coding.
8. ISO/IEC 23094-1; Information technology – General video coding – Part 1: Essential Video Coding.

## Informative References

MPAI-CAV Use Cases and Functional Requirements WD0.7, N479

# Use Cases

## Use Case #6 – The Human-CAV Interaction AIW (HCI)

### Use Case description

Human-CAV Interaction operated based on the principle that the CAV is impersonated by an avatar, selected/produced by the CAV rights-holder. The visible features of the avatar are head face and torso, and the audible feature is speech that embeds as much as possible the sentiment, e.g., emotion, that would be displayed by a human driver.

The CAV’s avatar is reactive to:

1. The Environment, e.g., it can show an altered face because a human driver has done what it considers an improper action.
2. A human, e.g., it shows an appropriate face to a human in the cabins who has made a joke.
3. Etc.

Other forms of interaction are:

1. CAV authenticates human interacting with it.
2. A human issues commands to a CAV, e.g.:
	1. Commands to Autonomous Motion Subsystem, e.g.: go to a Waypoint or display Full World Representation (see 5.3), etc.
	2. Other commands, e.g.: turn off air conditioning, turn on radio, call a person, open window or door, search for information etc.
3. A human entertains a dialogue with a CAV, e.g.:
	1. CAV offers a selection of offers to human (e.g., long but safe way, short but likely to have interruptions).
	2. Human requests information, e.g.: time to destination, route conditions, weather at destination etc.
	3. Human entertains a casual conversation.
4. A CAV monitors the passenger cabin, e.g.:
	1. Physical conditions, e.g.: temperature level, media being played, sound level, noise level, anomalous noise, etc.
	2. Passenger data, e.g.: number of passengers, ID, estimated age, destination of passengers.
	3. Passenger activity, e.g.: level of passenger activity, level of passenger-generated sound, level of passenger movement, emotion on face of passengers.
	4. Passenger-to-passenger dialogue, two passengers shake hands, or passengers hold everyday conversation.

It is important to point out that, regardless of the fact that vehicles can exhibit different levels of autonomy, the exhibited autonomy should always be adjustable [1]. The system should recognise people as intelligent agents it should inform and be informed by. A CAV should be able to change its level of autonomy to one of several levels during its operation. Such an adjustment may be initiated by a human, another system, or the CAV itself. An important benefit of adjustable, user-centered autonomy is increased user acceptance of the system.

### Reference architecture

*Figure 3* represents the Human-CAV Interaction (HCI) Reference Model. The following is noted:

1. A combination of Conversation with Emotion and Multimodal Question Answering AIMs with gesture recognition capabilities covers most Human-CAV Interaction needs.
2. Additional AIMs can be added should new HCI interactions be required.



*Figure 3 – Human-CAV Interaction Reference Model*

The speech of the human is separated from the audio captured from the environment; speaker and speech are recognised; meaning and emotion are extracted from speech; gesture and object information are extracted from video; human is identified, and emotion and meaning extracted from video of the face; emotions are fused; meanings are fused and intention derived to produce speech and face of the avatar interacting with humans; position of humans is computed to provide a realistic gazing. Additionally, human commands and responses from Autonomous Motion Subsystem are processed; Full World Representation is presented to let humans get a complete view of the Environment.

Depending on the technology used (data processing or AI), the AIMs in *Figure 3* may need to access external information, such as Knowledge Bases, to perform their functions. While not represented in *Figure 3*, they will be identified, if required, in the AI Modules subsection.

### Input and output data

*Table 2* gives the input/output data of Human-CAV Interaction.

*Table 2 – I/O data of* *Human-CAV Interaction*

|  |  |  |
| --- | --- | --- |
| **Input data** | **From** | **Comment** |
| Audio | User Outdoor | User authenticationUser command |
| Text | User Outdoor | User authenticationUser command |
| Text | Passenger Cabin | Social life of userCommands or interaction with CAV |
| Audio | Passenger Cabin | User’s social life Commands or interaction with CAV |
| Video | Passenger Cabin | Social life of userCommands or interaction with CAV |
| Full World Representation | Autonomous Motion SS | For processing by FWR Viewer |
| **Output data** | **To** | **Comments** |
| Text | Autonomous Motion SS | Commands to be executed |
| Synthetic Speech | Passenger Cabin | CAV’s response to passengers |
| Synthetic Face | Passenger Cabin | CAV’s response to passengers |
| Full World Representation | Passenger Cabin | For passengers to view external world  |

### AI Modules

*Table 3* gives the AI Modules of the Human-CAV Interaction depicted in *Figure 3*.

*Table 3 – AI Modules of* *Human-CAV interaction*

|  |  |
| --- | --- |
| **AIM** | **Function** |
| **Speech detection and separation** | 1. Separates relevant speech vs non-speech signals
2. Detects request for dialogue.
 |
| **Speaker identification** | Recognises speaker. |
| **Speech recognition** | 1. Analyses the speech input.
2. Generates text and emotion output.
 |
| **Object and gesture analysis**  | 1. Analyses video to identify object.
2. Produces the ID of the object in focus.
3. Analyses video.
4. Produces motion and mean­ing of gesture.
 |
| **Face recognition** | 1. Analyses the video of the face of a human.
2. Recognises the human’s identity.
 |
| **Face analysis** | 1. Analyses the video of the face of a human.
2. Extracts emotion and meaning.
 |
| **Language understanding** | 1. Analyses natural language expressed as text using a language model (embedded in AIM).
2. Produces the meaning of the text.
3. Identifies Object ID.
 |
| **Emotion recognition** | Produces Final Emotion by fusing Emotions from Speech, Face and Gesture. |
| **Question analysis** | 1. Fuses Meanings of Speech, Face and Gesture
2. Analyses the meaning of the sentence.
3. Determines the Intention.
4. Outputs Final Meaning.
 |
| **Question & dialog processing** | 1. Receives Speaker ID and Face ID.
2. If speaker ID and face ID match, then:
	1. Produces a command to Autonomous Motion SS
	2. Analyses user’s emotion, intention, meaning and/or ques­tion, text.
	3. Produces Concept (speech) and Concept (face).
3. Else, responds appropriately.
 |
| **Speech synthesis** | Converts Concept (Speech) to Output Speech. |
| **Face animation** | Converts Concept (Face) to Output Video. |
| **Full World Representation Viewer** | 1. Receives Full World Representation (FWR)
2. Presents a FWR view as instructed by human via FWR Com­mands.
 |

# Functional Requirements

The Functional Requirements refer to the individual technologies identified as necessary to implement MPAI-CAV Use Cases belonging using AIMs operating in an MPAI-AIF AI Framework. The Functional Requirements developed in this document adhere to the following guidelines:

AIMs are defined to allow implementations by multiple technologies (AI, ML, DP)

DP-based AIMs need interfaces, e.g., to a Know­ledge Base. AI-based AIM will typically require a learning process, however, support for this process is not included in the document. MPAI may develop further requirements covering that process in a future document.

AIMs can be aggregated in larger AIMs. Consequently, some data flows of aggregated AIMs may no longer be accessible.

## Human-CAV Interaction

### I/O Data summary

For each AIM (1st column), *Table 4* gives the input (2nd column) and the output data (3rd column).

*Table 4 – I/O data of Human-CAV Interaction AIMs*

|  |  |  |
| --- | --- | --- |
| **AIM** | **Input Data** | **Output Data** |
| **Speech Separation** | Input Audio | Separated Speech |
| **Internal AV Scene** | Input Video | Face Objects |
| **Speaker Recognition** | Separated Speech | Speaker ID |
| **Speech Recognition** | Separated Speech | Emotion (Speech)Text (Speech) |
| **Object and Gesture Analysis** | Input Video | Object IDEmotion (Gesture)Meaning (Gesture) |
| **Face Analysis** | Face Objects | Emotion (Face)Meaning (Face) |
| **Face Identification** | Face Objects | Face ID |
| **Full World Representation Viewer** | Full World RepresentationViewer Command  | FWRV AudioFWRV Video |
| **Emotion Fusion** | Emotion (Speech)Emotion (Face)Emotion (Gesture) | Fused Emotion |
| **Language Understanding** | Text (Speech)Input TextObject ID | Text (Language Understanding)Meaning (Text) |
| **Question analysis** | Meaning (Text)Meaning (Gesture)Meaning (Face) | Fused MeaningIntention |
| **Question and dialogue processing** | Input TextSpeaker IDFused emotionText (Speech)Fused MeaningIntentionFace IDFace Objects | Command/RequestFeedback/ResponseConcept (Speech)Output TextConcept (Face) |
| **Speech synthesis** | Concept (Speech) | Output Speech |
| **Face animation** | Concept (face) | Output Video |

### Audio

Monochannel Audio is the digital representation of an analogue audio signal sampled at a frequency between 8-192 kHz with a number of bits between 8 bits/sample and 32 bits/sample and a quantisation that is linear or companded.

**To respondents**

Respondents are invited to comment on these definitions and/or provide specific restrictions suitable to CAV-HCI.

**To respondents**

Respondents are requested to propose a coded representation of the above commands coordinated with the requirements of the with the Autonomous Motion Subsystem Responses. Proposals of coded representation of additional commands are welcome.

### Verbal Interaction

Some commands given to the Autonomous Motion Subsystem are:

1. Go to a Waypoint.
2. How long does it take to get there.
3. Park close to a Waypoint.
4. Drive faster.
5. Drive slowly.
6. Display Full World Representation.

Some of the responses of the Autonomous Motion Subsystem are:

1. Enumeration of possible routes with major features of each route.
2. Enumeration of possible parking places with major features of each place.
3. Announcement of obstacles preventing the expeditious accomplishment of the Command.
4. Announcement that the desired Waypoint has been reached.

**To respondents**

Respondents are requested to propose a coded representation of the above Commands/Responses. Proposals of coded representation of additional responses are welcome.

### Concept Expression (Face)

MPAI-MMC [3] specifies a Lips Animation format.

**To Respondents**

In this call, MPAI is looking for a technology that can animate head and face of the avatar with the purpose to represent:

1. Motion of head when speaking.
2. Motion of face muscles and eyeballs.
3. Turning of gaze to a particular person.
4. Emotion of the associated spoken sentence.
5. Meaning of the associated spoken sentence.

### Concept Expression (Speech)

MPAI-MMC [3] specifies Text With Emotion as Reply (speech) format.

**To Respondents**

Respondents are requested to propose a “Concept to Speech” format with the following requir­ements:

1. Capability to represent Emotions varying in time in the synthesised Speech.
2. Capability to represent Meanings varying in time in the CAV reply.

### Emotion

MPAI-MMC [3] specifies an extensible 3-level Basic Emotion Set.

**To respondents**

Respondents are requested to comment on the suitability of the technology standardised in [3] for the purpose of supporting human dialogue with a CAV. In case this is considered unsuitable, respondents are requested to motivate their assessment and provide an extension of the MPAI Basic Emotion Set or a new solution.

### Face identity

The Face Identity AIM shall be able to represent the identity of a limited number of faces.

**To respondents**

Respondents are requested to propose a face identification system suitable for a limited number of faces.

Proposals of a face identification usable in the context of a company renting CAVs to customers are welcome.

### Face Objects

In order for the HCI Subsystem to have a full understanding of what is happening in the passenger cabin (e.g., to have a more natural audio-visual interaction with the passengers, recording of what happened in the cabin etc.), the HCI Subsystem needs to represent the data acqu­ired from the cabin. The current use is:

1. To extract the face of a passenger for the purpose of extracting Emotion and Identity.
2. To determine the exact location of a passenger in the cabin in order to animate the CAV’s Avatar Face in such a way that the Avatar gazes into the eyes of the passenger it is talking to.

**To respondents**

Respondents are invited to propose a Face Objects format satisfying the above requirements to be used as input to Face Analysis, Face Identification and Question and Dialogue Processing.

### Full World Representation

The Full World Representation requirements are developed in the context of Autonomous Motion Subsystem requirements.

**To respondents**

Respondents are invited to comment.

### Full World Representation commands

The requirements of FWR interaction will be developed once the FWR requirements are defined.

**To respondents**

Respondents are invited to comment.

### Intention

MPAI-MMC [3] specifies a digital representation format for Intention.

**To respondents**

Respondents are requested to comment on the suitability of the technology standardised in [3] for CAV purposes.

### Meaning

MPAI-MMC [3] specifies a digital representation format for Meaning.

**To respondents**

Respondents are requested to comment on the suitability of the technology standardised in [3] for CAV purposes.

### Object Identifier

MPAI-MMC [3] specifies a digital representation format for Object Identifier to be used to identify objects held in the hand of a person.

**To respondents**

Respondents are requested to comment on the suitability of the technology standardised in [3] for CAV purposes.

### Speaker Identity

The current Speaker Identity requirements demand the ability to identify a limited number of Speakers.

**To respondents**

Respondents are requested to propose a Speaker Identification methods suitable for a limited number of speakers.

Proposals of a Speaker Identification method usable in a content of a company renting CAVs to customers are welcome.

### Text

MPAI-MMC [3] specifies ISO/IEC 10646, Information technology – Universal Coded Character Set (UCS) [5] as digital Text representation to support most languages in use.

**To respondents**

Respondents are invited to comment on this choice.

### Video

Video is intended for use in the passenger cabin. MPAI-MMC [3] specifies Video as:

1. Pixel shape: square
2. Bit depth: 8 or 10 bits/pixel
3. Aspect ratio: 4/3 or 16/9
4. 640 < # of horizontal pixels < 1920
5. 480 < # of vertical pixels < 1080
6. Frame frequency 50-120 Hz
7. Scanning: progressive
8. Colorimetry: ITU-R BT709 or BT2020
9. Colour format: RGB or YUV
10. Compression, either:
	1. Uncompressed;
	2. Compressed according to one of the following standards: MPEG-4 AVC [6], MPEG-H HEVC [7], MPEG-5 EVC [8]

**To respondents**

Respondents are invited to comment on MPAI’s choice for 2D Video.

Respondents are also requested to propose a data format for an array of cameras having video+depth as the baseline format or other 3D Video data formats.

# Data privacy

A CAV can generate or acquire data for which privacy is an important characteristic. Here are some of the functions potentially affected by data privacy or that are liable to become accessible to authorities, e.g., police and judiciary.

## Human-CAV Interaction (HCI)

By having interactions with humans, HCI becomes aware of potentially sensitive information, e.g.:

1. Result of monitoring the passenger cabin.
2. Minute requests from humans, e.g., go to a way point, display Full World Representation, turn off air conditioning, etc.
3. Dialogue with human

# Annex 1 – MPAI-wide terms and definitions (Normative)

The Terms used in this standard whose first letter is capital and are not already included in *Table 1* are defined in *Table 5.*

*Table 5 – MPAI-wide Terms*

|  |  |
| --- | --- |
| **Term** | **Definition** |
| Access | Static or slowly changing data that are required by an application such as domain knowledge data, data models, etc. |
| AIF Metadata | The data set describing the capabilities of an AIF set by the AIF Implem­enter. |
| AIM Metadata | The data set describing the capabilities of an AIM set by the AIM Implem­enter. |
| AI Framework (AIF) | The environment where AIWs are executed. |
| AI Module (AIM) | A processing element receiving AIM-specific Inputs and producing AIM-specific Outputs according to according to its Function. An AIM may be an aggregation of AIMs. |
| AI Workflow (AIW) | A structured aggregation of AIMs implementing a Use Case receiving AIM-spec­ific inputs and producing AIM-specific inputs according to its Function. |
| Application Programming Interface (API) | A software interface that allows two applications to talk to each other |
| Application Standard  | An MPAI Standard specifying AIWs, AIMs, Topologies and Formats suitable for a particular application domain. |
| Channel | A physical or logical connection between an output Port of an AIM and an input Port of an AIM. The term “connection” is also used as a synonym. |
| Communication | The infrastructure that implements message passing between AIMs. |
| Component | One of the 9 AIF elements: Access, AI Module, AI Workflow, Commun­ication, Controller, Internal Storage, Global Storage, MPAI Store, and User Agent. |
| Conformance | The attribute of an Implementation of being a correct technical Implem­entation of a Technical Specification. |
| Conformance Tester | An entity authorised by MPAI to Test the Conformance of an Implem­entation. |
| Conformance Testing | The normative document specifying the Means to Test the Conformance of an Implem­entation. |
| Conformance Testing Means | Procedures, tools, data sets and/or data set characteristics to Test the Conformance of an Implem­en­tation. |
| Connection | A channel connecting an output port of an AIM and an input port of an AIM. |
| Controller | A Component that manages and controls the AIMs in the AIF, so that they execute in the correct order and at the time when they are needed. |
| Data | Information in digital form. |
| Data Format | The standard digital representation of Data. |
| Data Semantics | The meaning of Data. |
| Device | A hardware and/or software entity running at least one instance of an AIF. |
| Ecosystem | The ensemble of the following actors: MPAI, MPAI Store, Implementers, Conformance Testers, Performance Testers and Users of MPAI-AIF Im­plem­en­tations as needed to enable an Interoperability Level. |
| Event | An occurrence acted on by an Implementation. |
| Explainability | The ability to trace the output of an Implementation back to the inputs that have produced it. |
| Fairness | The attribute of an Implementation whose extent of applicability can be assessed by making the training set and/or network open to testing for bias and unanticipated results. |
| Function | The operations effected by an AIW or an AIM on input data. |
| Global Storage | A Component to store data shared by AIMs. |
| Internal Storage | A Component to store data of the individual AIMs. |
| Identifier | A name that uniquely identifies an Implementation. |
| Implementation | 1. An embodiment of the MPAI-AIF Technical Specification, or
2. An AIW or AIM of a particular Level (1-2-3).
 |
| Interoperability | The ability to functionally replace an AIM/AIW with another AIM/AIW having the same Interoperability Level |
| Interoperability Level | The attribute of an AIW and its AIMs to be executable in an AIF Implem­en­tati­on and to be: 1. Implementer-specific and satisfying the MPAI-AIF Standard *(Level 1)*.
2. Specified by an MPAI Application Standard (*Level 2)*.
3. Specified by an MPAI Application Standard and certified by a Performance Assessor (*Level 3)*.
 |
| Knowledge Base | Structured and/or unstructured information made accessible to AIMs via MPAI-specified interfaces |
| Message | A sequence of Records. |
| Normativity | The set of attributes of a technology or a set of technologies specified by the applicable parts of an MPAI standard. |
| Performance | The attribute of an Implementation of being Reliable, Robust, Fair and Replicable. |
| Performance Assessment | The normative document specifying the procedures, the tools, the data sets and/or the data set characteristics to Assess the Grade of Performance of an Implementation. |
| Performance Assessment Means | Procedures, tools, data sets and/or data set characteristics to Assess the Performance of an Implem­en­tation. |
| Performance Assessor | An entity authorised by MPAI to Assess the Performance of an Implementation in a given Application domain |
| Port | A physical or logical communication interface of an AIM. |
| Profile | A particular subset of the technologies used in MPAI-AIF or an AIW of an Application Standard and, where applicable, the classes, other subsets, options and parameters relevant to that subset. |
| Record | Data with a specified structure. |
| Reference Model | The AIMs and theirs Connections in an AIW. |
| Reference Software | A technically correct software implementation of a Technical Specific­ation containing source code, or source and compiled code.  |
| Reliability | The attribute of an Implementation that performs as specified by the Application Standard, profile and version the Implementation refers to, e.g., within the application scope, stated limitations, and for the period of time specified by the Implementer. |
| Replicability | The attribute of an Implementation whose Performance, as Assessed by a Performance Assessor, can be replicated, within an agreed level, by another Performance Assessor. |
| Robustness | The attribute of an Implementation that copes with data outside of the stated application scope with an estimated degree of confidence. |
| Scope | The domain of applicability of an MPAI Application Standard |
| Service Provider | An entrepreneur who offers an Implementation as a service (e.g., a recommendation service) to Users. |
| Specification | A collection of normative clauses. |
| Standard | The ensemble of Technical Specification, Reference Software, Confor­man­ce Testing and Performance Assessment of an MPAI application Standard.  |
| Technical Specification | (Framework) the normative specification of the AIF.(Application) the normative specification of the set of AIWs belon­ging to an application domain along with the AIMs required to Im­plem­ent the AIWs that includes:1. The formats of the Input/Output data of the AIWs implementing the AIWs.
2. The Connections of the AIMs of the AIW.
3. The formats of the Input/Output data of the AIMs belonging to the AIW.
 |
| Testing Laboratory | A laboratory accredited by MPAI to Assess the Grade of Performance of Implementations.  |
| Time Base | The protocol specifying how Components can access timing information |
| Topology | The set of AIM Connections of an AIW. |
| Use Case | A particular instance of the Application domain target of an Application Standard. |
| User | A user of an Implementation. |
| User Agent | The Component interfacing the user with an AIF through the Controller |
| Version | A revision or extension of a Standard or of one of its elements. |
| Zero Trust | A cybersecurity model primarily focused on data and service protection that assumes no implicit trust. |

# Annex 2 - Notices and Disclaimers Concerning MPAI Standards (Informative)

The notices and legal disclaimers given below shall be borne in mind when [downloading](https://www.mpai.community/resources/) and using approved MPAI Standards.

In the following, “Standard” means the collection of four MPAI-approved and [published](https://www.mpai.community/resources/) documents: “Technical Specification”, “Reference Software” and “Conformance Testing” and, where applicable, “Performance Testing”.

Life cycle of MPAI Standards

MPAI Standards are developed in accordance with the [MPAI Statutes](https://mpai.community/statutes/). An MPAI Standard may only be developed when a Framework Licence has been adopted. MPAI Standards are developed by especially established MPAI Development Committees who operate on the basis of consensus, as specified in Annex 1 of the [MPAI Statutes](https://mpai.community/statutes/). While the MPAI General Assembly and the Board of Directors administer the process of the said Annex 1, MPAI does not independently evaluate, test, or verify the accuracy of any of the information or the suitability of any of the technology choices made in its Standards.

MPAI Standards may be modified at any time by corrigenda or new editions. A new edition, however, may not necessarily replace an existing MPAI standard. Visit the [web page](https://mpai.community/resources/) to determine the status of any given published MPAI Standard.

Comments on MPAI Standards are welcome from any interested parties, whether MPAI members or not. Comments shall mandatorily include the name and the version of the MPAI Standard and, if applicable, the specific page or line the comment applies to. Comments should be sent to the MPAI Secretariat. Comments will be reviewed by the appropriate committee for their technical relevance. However, MPAI does not provide interpretation, consulting information, or advice on MPAI Standards. Interested parties are invited to join MPAI so that they can attend the relevant Development Committees.

Coverage and Applicability of MPAI Standards

MPAI makes no warranties or representations of any kind concerning its Standards, and expressly disclaims all warranties, expressed or implied, concerning any of its Standards, including but not limited to the warranties of merchantability, fitness for a particular purpose, non-infringement etc. MPAI Standards are supplied “AS IS”.

The existence of an MPAI Standard does not imply that there are no other ways to produce and distribute products and services in the scope of the Standard. Technical progress may render the technologies included in the MPAI Standard obsolete by the time the Standard is used, especially in a field as dynamic as AI. Therefore, those looking for standards in the Data Compression by Artificial Intelligence area should carefully assess the suitability of MPAI Standards for their needs.

IN NO EVENT SHALL MPAI BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO: THE NEED TO PROCURE SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE PUBLICATION, USE OF, OR RELIANCE UPON ANY STANDARD, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE AND REGARDLESS OF WHETHER SUCH DAMAGE WAS FORESEEABLE.

MPAI alerts users that practicing its Standards may infringe patents and other rights of third parties. Submitters of technologies to this standard have agreed to licence their Intellectual Property according to their respective Framework Licences.

Users of MPAI Standards should consider all applicable laws and regulations when using an MPAI Standard. The validity of Conformance Testing is strictly technical and refers to the correct implementation of the MPAI Standard. Moreover, positive Performance Assessment of an implementation applies exclusively in the context of the [MPAI Governance](https://mpai.community/governance/) and does not imply compliance with any regulatory requirements in the context of any jurisdiction. Therefore, it is the responsibility of the MPAI Standard implementer to observe or refer to the applicable regulatory requirements. By publishing an MPAI Standard, MPAI does not intend to promote actions that are not in compliance with applicable laws, and the Standard shall not be construed as doing so. In particular, users should evaluate MPAI Standards from the viewpoint of data privacy and data ownership in the context of their jurisdictions.

Implementers and users of MPAI Standards documents are responsible for determining and complying with all appropriate safety, security, environmental and health and all applicable laws and regulations.

Copyright

MPAI draft and approved standards, whether they are in the form of documents or as web pages or otherwise, are copyrighted by MPAI under Swiss and international copyright laws. MPAI Standards are made available and may be used for a wide variety of public and private uses, e.g., implementation, use and reference, in laws and regulations and standardisation. By making these documents available for these and other uses, however, MPAI does not waive any rights in copyright to its Standards. For inquiries regarding the copyright of MPAI standards, please contact the MPAI Secretariat.

The Reference Software of an MPAI Standard is released with the [MPAI Modified Berkeley Software Distribution licence](https://mpai.community/about/licence/). However, implementers should be aware that the Reference Software of an MPAI Standard may reference some third party software that may have a different licence.

# Annex 3 – The Governance of the MPAI Ecosystem (Informative)

**Level 1 Interoperability**

With reference to *Figure 1*, MPAI issues and maintains a standard – called MPAI-AIF – whose components are:

1. An environment called AI Framework (AIF) running AI Workflows (AIW) composed of inter­connected AI Modules (AIM) exposing standard interfaces.
2. A distribution system of AIW and AIM Implementation called MPAI Store from which an AIF Implementation can download AIWs and AIMs.

|  |  |
| --- | --- |
| Implementers’ benefits | Upload to the MPAI Store and have globally distributed Implementations of* AIFs conforming to MPAI-AIF.
* AIWs and AIMs performing prop­rietary functions executable in AIF.
 |
| Users’ benefits | Rely on Implementations that have been tested for security. |
| MPAI Store’s role | * Tests the Conformance of Implementations to MPAI-AIF.
* Verifies Implementations’ security, e.g., absence of malware.
* Indicates unambiguously that Implementations are Level 1.
 |

**Level 2 Interoperability**

In a Level 2 Implem­entation, the AIW must be an Implementation of an MPAI Use Case and the AIMs must con­form with an MPAI Applicati­on Standard.

|  |  |
| --- | --- |
| Implementers’ benefits | Upload to the MPAI Store and have globally distributed Implementations of* AIFs conforming to MPAI-AIF.
* AIWs and AIMs conforming to MPAI Application Standards.
 |
| Users’ benefits | * Rely on Implementations of AIWs and AIMs whose Functions have been reviewed during standardisation.
* Have a degree of Explainability of the AIW operation because the AIM Func­tions and the data Formats are known.
 |
| Market’s benefits | * Open AIW and AIM markets foster competition leading to better products.
* Competition of AIW and AIM Implementations fosters AI innovation.
 |
| MPAI Store’s role | * Tests Conformance of Implementations with the relevant MPAI Standard.
* Verifies Implementations’ security.
* Indicates unambiguously that Implementations are Level 2.
 |

**Level 3 Interoperability**

MPAI does not generally set standards on how and with what data an AIM should be trained. This is an important differentiator that promotes competition leading to better solutions. However, the performance of an AIM is typically higher if the data used for training are in greater quantity and more in tune with the scope. Training data that have large variety and cover the spec­trum of all cases of interest in breadth and depth typically lead to Implementations of higher “quality”.

For Level 3, MPAI normatively specifies the process, the tools and the data or the characteristics of the data to be used to Assess the Grade of Performance of an AIM or an AIW.

|  |  |
| --- | --- |
| Implementers’ benefits | May claim their Implementations have passed Performance Assessment. |
| Users’ benefits | Get assurance that the Implementation being used performs correctly, e.g., it has been properly trained. |
| Market’s benefits | Implementations’ Performance Grades stimulate the development of more Performing AIM and AIW Implementations. |
| MPAI Store’s role | * Verifies the Implementations’ security
* Indicates unambiguously that Implementations are Level 3.
 |

**The MPAI ecosystem**

The following *Figure 4* is a high-level description of the MPAI ecosystem operation applicable to fully conforming MPAI implementations:

1. MPAI establishes and controls the not-for-profit MPAI Store.
2. MPAI appoints Performance Assessors.
3. MPAI publishes Standards.
4. Implementers submit Implementations to Performance Assessors.
5. If the Implementation Performance is acceptable, Performance Assessors inform Implementers (step 5a) and MPAI Store.
6. Implementers submit Implementations to the MPAI Store tested for Confor­mance and security.
7. Users download and use Implementations, and submit experience scores.

*Figure 4 – The MPAI ecosystem operation*