



Moving Picture, Audio and Data Coding
by Artificial Intelligence
www.mpai.community

MPAI Technical Specification

Human and Machine Communication MPAI-HMC

V1.1

WARNING

Use of the technologies described in this Technical Specification may infringe patents, copyrights, or intellectual property rights of MPAI Members or non-members.

MPAI and its Members accept no responsibility whatsoever for damages or liability, direct or consequential, which may result from the use of this Technical Specification.

Readers are invited to review Annex 1 - Notices and Disclaimers.

Technical Specification

Human and Machine Communication

(MPAI-HMC) V1.1

Contents

1	Foreword (Informative).....	3
2	Introduction (Informative).....	4
3	Scope	4
4	Definitions.....	5
5	References	11
5.1	Normative References	11
5.2	Informative References	11
6	Usage Scenarios (informative).....	11
6.1	Introduction	11
6.1.1	Information Service.....	13
6.1.2	Cross-Cultural Information Service	13
6.1.3	Virtual Assistant.....	13
6.1.4	Conversation companion.....	14
6.1.5	Strolling in the metaverse.....	14
6.1.6	Travelling in a Connected Autonomous Vehicle	15
7	AI Workflows.....	15
7.1	AIWs Communicating Entities in Context.....	15
7.1.1	Functions	15
7.1.2	Reference Model	16
7.1.3	Input/Output Data.....	17
7.1.4	Functions of AI Modules.....	18
7.1.5	AIW, AIMS, and JSON Metadata	19
8	AI Modules.....	20
8.1	Entity and Context Understanding	21
8.1.1	Functions	21
8.1.2	Reference Model	21
8.1.3	I/O Data	22
8.1.4	SubAIMs	22
8.1.5	JSON Metadata	23
8.1.6	Profiles	23
8.2	AV Scene Integration and Description.....	23
8.2.1	Functions	23
8.2.2	Reference Architecture.....	24
8.2.3	I/O Data	24
8.2.4	SubAIMs	24
8.2.5	JSON Metadata	24
8.2.6	Profiles	24
9	Data Types.....	25
10	Profiles	25
	Annex 1 - Notices and Disclaimers Concerning MPAI Standards (Informative).....	27
	Annex 2 - Patent Declarations.....	29

1 Foreword (Informative)

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. [Moving Picture, Audio, and Data Coding by Artificial Intelligence](#) (MPAI) has been established to develop standards that promote the efficient use of data especially using Artificial Intelligence technologies.

The accomplishment of the MPAI mission is facilitated/enabled by two foundational Technical Specifications.

[Technical Specification: Artificial Intelligence Framework \(MPAI-AIF\)](#) specifies an environment called AI Framework (AIF) enabling initialisation, dynamic configuration, and control of AI Workflows (AIW) including interconnected AI Modules (AIM) as depicted in Figure 1. AIWs are specified by their functions, input/output data, and AIM topology and AIMs are specified by their functions and input/output data in a technology-agnostic fashion. MPAI-AIF can support small- and large-scale high-performance components and enables solutions with improved explainability.

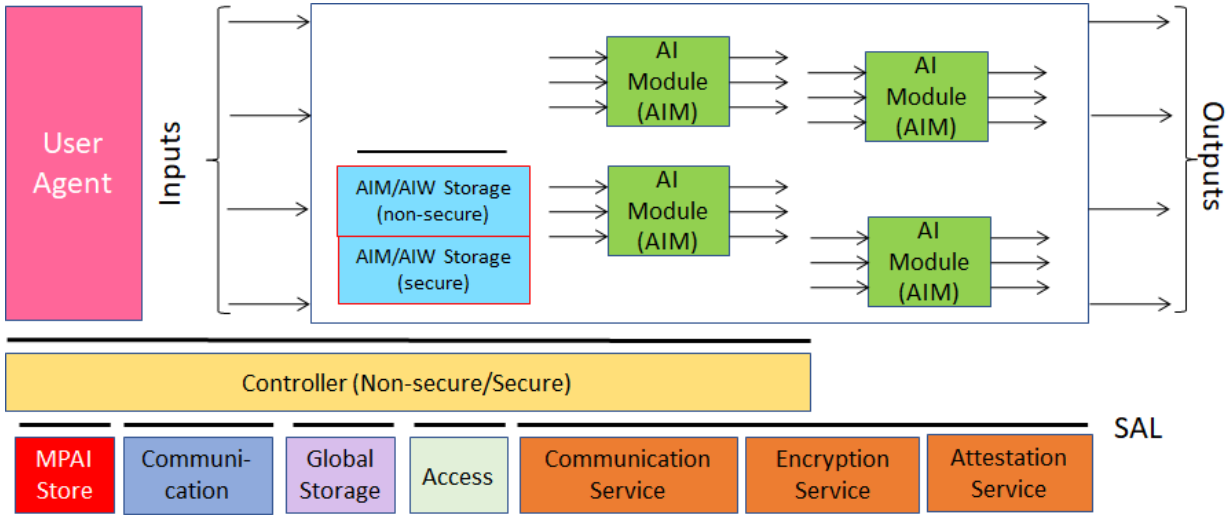


Figure 1 – The AI Framework (MPAI-AIF) V2 Reference Model

[Technical Specification: Governance of the MPAI Ecosystem](#) identifies the following elements:

1. Standards, i.e., the ensemble of Technical Specifications, Reference Software, Conformance Testing, and Performance Assessment.
2. Implementers and Integrators of MPAI Technical Specifications.
3. MPAI Store in charge of making AIMs and AIWs submitted by Implementers available to Integrators and End-Users.
4. Performance Assessors, independent entities assessing the performance of implementations in terms of Reliability, Replicability, Robustness, and Fairness.
5. End Users.

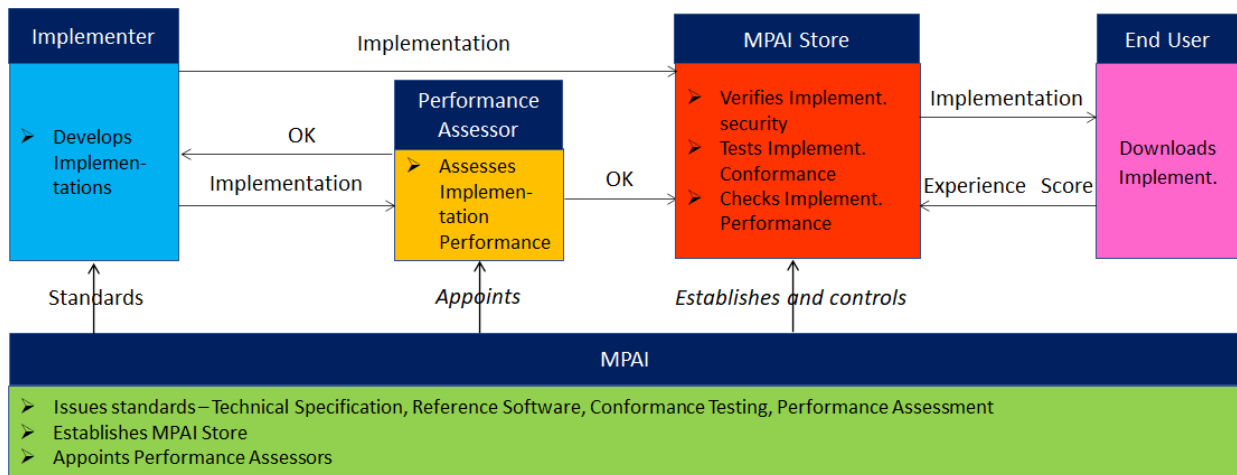


Figure 2 - The MPAI Ecosystem

2 Introduction (Informative)

This *Technical Specification: Human and Machine Communication (MPAI-HMC)* leverages five MPAI Technical Specifications: Context-based Audio Enhancement [3], MPAI Metaverse Model – Architecture [5], Multimodal Conversation [4], Object and Scene Description [6], and Portable Avatar Format [7]. All these Technical Specifications deal with technologies enabling communication of real and digital humans in real or virtual environments. MPAI-HMC adopts the normative elements from the five Technical Specifications that are required by this Technical Specification.

Capitalised Terms are defined in Table 1 of [Definitions](#) if they are MPAI-HMC-specific or in Table 2 if their use extends across MPAI Technical Specifications. A term beginning with a small letter has the commonly intended meaning.

This Chapter is Informative. Chapters, Sections, and Annexes are Normative unless they are explicitly identified as Informative.

3 Scope

Technical Specification: Human and Machine Communication (MPAI-HMC) – referred to in the following as MPAI-HMC – specifies technologies that enable advanced forms of communication involving humans in different forms, e.g., as humans in a real space or their digital representations in a Virtual Space, and Machines also in different forms, e.g., as digital instantiations in a Virtual Space or as humanoids rendered in a real space.

In this specification, a word beginning with a small letter represents an object in the real world. A word beginning with a capital letter represents an Object in the Virtual World.

MPAI-HMC Version 1.1 specifies the *Communicating Entities in Context* Use Case as a first example of how effective forms of human-Machine communication can be implemented in an interoperable fashion.

MPAI-HMC has been developed by CAE-DC, MMC-DC, and PAF-DC.

MPAI may develop future versions of this Technical Specification or new Technical Specifications addressing the Human and Machine Communication area.

4 Definitions

Terms beginning with a capital letter have the meaning defined in *Table 1*. Terms beginning with a small letter have the meaning commonly defined for the context in which they are used. For instance, *Table 1* defines *Object* and *Scene* but does not define *object* and *scene*.

A dash “-” preceding a Term in *Table 1* indicates the following readings according to the font:

1. Normal font: the Term in the table without a dash and preceding the one with a dash should be read before that Term. For example, “Avatar” and “- Model” will yield “Avatar Model.”
2. *Italic* font: the Term in the table without a dash and preceding the one with a dash should be read after that Term. For example, “Avatar” and “- Portable” will yield “Portable Avatar.”

The full set of Terms and Definitions relevant to all MPAI Technical Specifications, including MPAI-HMC, can be [accessed](#) online.

Table 1 - General MPAI-HMC terms

Terms	Definitions
Attitude	
- <i>Social</i>	The coded representation of the internal state related to the way a human or avatar intends to position vis-à-vis the Environment or subsets of it, e.g., “Respectful”, “Confrontational”, “Soothing”.
- <i>Spatial</i>	Position and Orientation and their velocities and accelerations of an Object in a Real or Virtual Environment.
Audio	Digital representation of an analogue audio signal sampled at a frequency between 8-192 kHz with a number of bits/sample between 8 and 32, and non-linear and linear quantisation. Data with characteristics of Audio may be synthetically produced.
Audio Block	A set of consecutive Audio samples.
Audio Channel	A sequence of Audio Blocks.
Avatar	An Object rendered to represent a Human of a Machine in a virtual space.
- Model	An inanimate Avatar exposing animation interfaces.
- <i>Portable</i>	A Data Type including Avatar ID, Time, Visual Environment, Spatial Attitude, Avatar Model, Body Descriptors, Face Descriptors, Language Preference, Speech Coding, Speech Data, Text, and Personal Status [6].
Body	A digital representation of a human body, head included, face excluded.
Centre Point	The point of an Object selected to have Local Coordinates (0,0,0).
Cognitive State	The coded representation of the internal state reflecting the way a human or avatar understands the Environment, such as “Confused”, “Dubious”, “Convinced”.
Communication Item	An element generated by a Machine communicating with an Entity expressed with a Portable Avatar.
Context	The semantics of the information emitted by an Entity or included in its surrounding Scene.
Coordinate System	A coordinate system where the position of a point is specified by three numbers.
- <i>Cartesian</i>	A coordinate system where the three numbers are the signed distances from the point to three mutually perpendicular planes.

- <i>Spherical</i>	A coordinate system where the three numbers are: <ul style="list-style-type: none"> - the radial distance of that point from a fixed origin. - the polar angle measured from a fixed zenith direction. - the azimuthal angle of its orthogonal projection on a reference plane.
Culture	The collection of language and customs governing the way a human, or a group of humans employ to express their internal statuses.
Data	Information in digital form.
- Format	The standard digital representation of Data.
- Type	An instance of Data with a specific Data Format.
Descriptor	The Digital Representation of a feature of an Object.
- <i>Body</i>	A Data Type including the digital representation of the features of the body of a real or digital human.
- <i>Face</i>	A Data Type including the digital representation of a feature of the face of a real or digital human.
- <i>Speech</i>	A Data Type representing a variety of information elements incorporated in a Speech Segment, e.g., personal identity, Personal Status, additional factors such as vocal tension, creakiness, whispery quality, etc.
- <i>Text</i>	A Data Type including the digital representation of a feature of text.
Digital Representation	Data corresponding to and representing a physical entity.
Emotion	The coded representation of the internal state resulting from the interaction of a human or avatar with the Environment or subsets of it, such as “Angry”, “Sad”, “Determined”.
Entity	A Human of a Machine communicating with a Machine.
Environment	A Virtual Space that may be null or may include an Audio-Visual Scene.
Experience	The state of an Entity whose senses/sensors are continuously affected for a meaningful period.
Face	A digital representation of a human face.
Factor	One of Emotion, Cognitive State, and Attitude.
Gesture	A movement of a Digital Human or part of it, such as the head, arm, hand, and finger, often a complement to a vocal utterance.
Human	A human being in a real space.
- <i>Digital</i>	A Digitised or a Virtual Human in a Virtual Space.
- <i>Digitised</i>	An Object in a Virtual Space that has the appearance of a specific human when rendered.
- <i>Virtual</i>	An Object in a Virtual Space created by a computer that has a human appearance when rendered but is not a Digitised Human.
Identifier	The label uniquely associated with a human or an Object.
Instance	An element of a set of entities – Objects, Digital Humans etc. – belonging to some levels in a hierarchical classification (taxonomy).
- <i>Audio</i>	The instance of an Audio Object.
- <i>Visual</i>	The instance of a Visual Object.
Machine	An Implementation of MPAI-MMC.
Meaning	Information extracted from Text such as syntactic and semantic information, Personal Status, and other information, such as an Object Identifier.
Microphone Array Geometry	A Data Type representing the position of each microphone comprising a microphone array and characteristics such as microphone type, look directions, and array type.
Modality	One of Text, Speech, Face, or Gesture.

Object	A data structure that can be rendered to cause an Experience.
- <i>Audio</i>	An Object described by Audio Descriptors.
- <i>Audio-Visual</i>	An Object described by Audio-Visual Descriptors.
- <i>Body</i>	A digital representation of the body of a Human or a Machine.
- <i>Descriptor</i>	The digital representation of the feature of an Object.
- <i>Digital</i>	A Digitised or a Virtual Object.
- <i>Digitised</i>	The digital representation of a real object.
- <i>Face</i>	The digital representation of the face of a Human or a Machine.
- <i>Speech</i>	An Object described by Speech Descriptors.
- <i>Text</i>	A string of Text.
- <i>Virtual</i>	An Object not representing an object in the real environment.
- <i>Visual</i>	An Object described by Visual Descriptors.
Orientation	The 3 Euler angles of an Object in a Virtual Space.
Personal Status	A Data Type including three Factors – Cognitive State, Emotion and Social Attitude – conveyed by four Modalities – Text, Speech, Face, and Gesture and providing standard extensible labels for the three Factors [4].
- <i>Face</i>	The Cognitive State, Emotion, and Social Attitude conveyed by a Face Object.
- <i>Gesture</i>	The Cognitive State, Emotion, and Social Attitude conveyed by the Gesture of a Body Object.
- <i>Speech</i>	The Cognitive State, Emotion, and Social Attitude conveyed by a Speech Object.
- <i>Text</i>	The Cognitive State, Emotion, and Social Attitude conveyed by a Text Object.
Position	The coordinates of a representative point for an object in a Virtual Space with respect to a set of coordinate axes.
Principal Axis	The x axis of an Object.
Rendering	The process of instantiating a Virtual Space as a human-perceptible entity.
Scene	A composition of Objects located according to a Scene Geometry.
- <i>Audio</i>	A Scene composed of Audio Objects.
- <i>Audio-Visual</i>	A Scene composed of Audio Objects, Visual Objects and co-located Audio-Visual Objects.
- <i>Multichannel</i>	A data structure containing at least 2 time-aligned interleaved Audio Channels.
- <i>Visual</i>	A Scene composed of Visual Objects.
Scene Descriptors	The digital representation of a feature of a scene.
- <i>Audio</i>	A Data Type including the digital representation of the audio features of a real or digital scene.
- <i>Audio-Visual</i>	A Data Type combining the Audio or Visual Scene Descriptors.
- <i>Visual</i>	A Data Type including the digital representation of the visual features of a real or digital scene.
Scene Geometry	The digital representation of the Object arrangement of a Scene.
- <i>Audio</i>	A Data Type describing the spatial arrangement of the Visual Objects of a Scene.
- <i>Audio-Visual</i>	A Data Type describing the spatial arrangement of the Audio, Visual, and Audio-Visual Objects of a Scene.
- <i>Visual</i>	A Data Type describing the spatial arrangement of the Visual Objects of a Scene.

Selector	Input Data having the goal to set a parameter (e.g., use of Text vs Speech or Language Preference) or an operating mode of a Machine.
Speech	Digital representation of analogue speech sampled at a frequency between 8 kHz and 96 kHz with a number of bits/sample of 8, 16 or 24, and non-linear and linear quantisation or compressed. Data with characteristics of Speech may be synthetically produced.
Text	A sequence of characters represented according to [10].
- <i>Recognised</i>	The Text at the output of an Automatic Speech Recognition AIM.
- <i>Refined Text</i>	The Text at the output of a Natural Language Understanding AIM.
- <i>Translated Text</i>	The Text at the output of a Natural Language Translation AIM.
Virtual Space	A space generated and maintained by a computing platform that can be rendered.

The capitalised Terms used in this standard that are not already included in *Table 1* are defined in *Table 2*.

NOTE: A hyphenated entry for e.g., “- Testing” should be read as adding that word to the closest non-hyphenated entry above it – in this case, “Conformance,” giving “Conformance Testing” as the complete entry name.

Table 2 – Terms used across several MPAI Technical Specifications

Term	Definition
Access	Static or slowly changing data that are required by an application such as domain knowledge data, data models, etc.
AI Framework (AIF)	The environment where AIWs are executed.
AI Model (AIM)	A data processing element receiving AIM-specific Inputs and producing AIM-specific Outputs 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 AIW-specific inputs and producing AIW-specific outputs according to the AIW Function.
Application Standard	An MPAI Standard designed to enable a particular application domain.
Assessment Laboratory	A laboratory accredited to Assess the Grade of Performance of Implementations.
Channel	A connection between an output port of an AIM and an input port of an AIM. The term “connection” is also used as synonymous.
Communication	The infrastructure that implements message passing between AIMs.
Component	One of the 7 AIF elements: Access, Communication, Controller, Internal Storage, Global Storage, Store, and User Agent
Composite AIM	An AIM aggregating more than one AIM.
Component	One of the 7 AIF elements: Access, Communication, Controller, Internal Storage, Global Storage, Store, and User Agent
Conformance	The attribute of an Implementation of being a correct technical Implementation of a Technical Specification.

- Testing	The normative document specifying the Means to Test the Conformance of an Implementation.
- Testing Dataset	A dataset used to Test the Conformance of an implementation to a Technical Specification.
- Testing Means	Procedures, tools, data sets and/or data set characteristics to Test the Conformance of an Implementation.
- Testing Procedure	The sequence of steps to be performed to Test the Conformance of an implementation.
- Testing Tools	Devices and/or software used 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.
- Format	The standard digital representation of Data.
- Type	An instance of Data with a specific Data Format.
- Semantics	The meaning of Data.
Descriptor	Coded representation of a text, audio, speech, or visual feature.
Digital Representation	Data corresponding to and representing a physical entity.
Ecosystem	The ensemble of actors making it possible for a User to execute an application composed of an AIF, one or more AIWs, each with one or more AIMs potentially sourced from independent implementers.
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.
AIM/AIW 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) conforming with a Use Case of an MPAI Application Standard.
Implementer	A legal entity implementing MPAI Technical Specifications.
ImplementerID (IID)	A unique name assigned by the ImplementerID Registration Authority to an Implementer.
ImplementerID Registration Authority (IIDRA)	The entity appointed by MPAI to assign ImplementerID's to Implementers.
Instance ID	Instance of a class of Objects and the Group of Objects the Instance belongs to.
Interoperability	The ability to functionally replace an AIM with another AIW having the same Interoperability Level
- Level	The attribute of an AIW and its AIMs to be executable in an AIF Implementation and to: 1. Be proprietary (Level 1)

	<p>2. Pass the Conformance Testing (Level 2) of an Application Standard</p> <p>3. Pass the Performance Testing (Level 3) of an Application Standard.</p>
Knowledge Base	Structured and/or unstructured information made accessible to AIMs via MPAI-specified interfaces
Message	A sequence of Records transported by Communication through Channels.
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.
- Assessment	The normative document specifying the Means to Assess the Grade of Performance of an Implementation.
- Assessment Means	Procedures, tools, data sets and/or data set characteristics to Assess the Performance of an Implementation.
- Assessor	An entity Assessing the Performance of an Implementation.
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	A data structure with a specified structure
Reference Model	The AIMs and their 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.
Standard	A set of Technical Specification, Reference Software, Conformance Testing, Performance Assessment, and Technical Report of an MPAI application Standard.
Technical Specification	<p>(Framework) the normative specification of the AIF.</p> <p>(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:</p> <ol style="list-style-type: none"> 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.
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.

5 References

This page provides normative and information references. The full set of normative references can be [accessed](#) online.

5.1 Normative References

1. Technical Specification; [MPAI Ecosystem Governance](#) (MPAI-GME) V1.1.
2. Technical Specification; [AI Framework](#) (MPAI-AIF) V2.0.
3. Technical Specification: [Context-based Audio Enhancement](#) (MPAI-CAE) V2.2.
4. Technical Specification: [Multimodal Conversation](#) (MPAI-MMC) V2.2.
5. Technical Specification: [MPAI Metaverse Model](#) (MPAI-MMM) – Architecture V1.1.
6. Technical Specification: [Object and Scene Description](#) (MPAI-OSD) V1.1.
7. Technical Specification: [Portable Avatar Format](#) (MPAI-PAF) V1.2.
8. ITU-R; [Long-form file format for the international exchange of audio programme materials with metadata](#); BS.2088 (10/2019)
<https://www.loc.gov/preservation/digital/formats/fdd/fdd000001.shtml>.
9. ISO 639; Codes for the Representation of Names of Languages – Part 1: Alpha-2 Code.
10. ISO/IEC 10646; Information technology – Universal Coded Character Set.

5.2 Informative References

11. MPAI; [The MPAI Statutes](#).
12. MPAI; [The MPAI Patent Policy](#).
13. MPAI; [Framework Licence: Human and Machine Communication \(MPAI-HMC\)](#).
14. Technical Specification: [Connected Autonomous Vehicle](#) (MPAI-CAV) – Architecture V1.
15. Technical Specification: [Compression and Understanding of Industrial Data](#) (MPAI-CUI) V1.1.
16. Technical Specification: [MPAI Metaverse Model](#) (MPAI-MMM) – Architecture V1.1.
17. Technical Specification: [Neural Network Watermarking](#) (MPAI-NNW) V1.0.
18. Ekman, Paul (1999), "Basic Emotions", in Dalgleish, T; Power, M (eds.), Handbook of Cognition and Emotion (PDF), Sussex, UK: John Wiley & Sons.

6 Usage Scenarios (informative)

6.1 Introduction

This Chapter includes five usage scenarios mostly described as particular cases of the combined usage scenarios of Figure 3 that combines some of the communication settings between Humans and Machines targeted by HMC-CEC. In Figure 3, the term Machine followed by a number indicates an HMC-CEC Implementation. A Machine can be an application, a device, or a function of a larger system. For the sake of simplicity, the Text component is not included in Figure 3, but is supported by HMC-CEC.

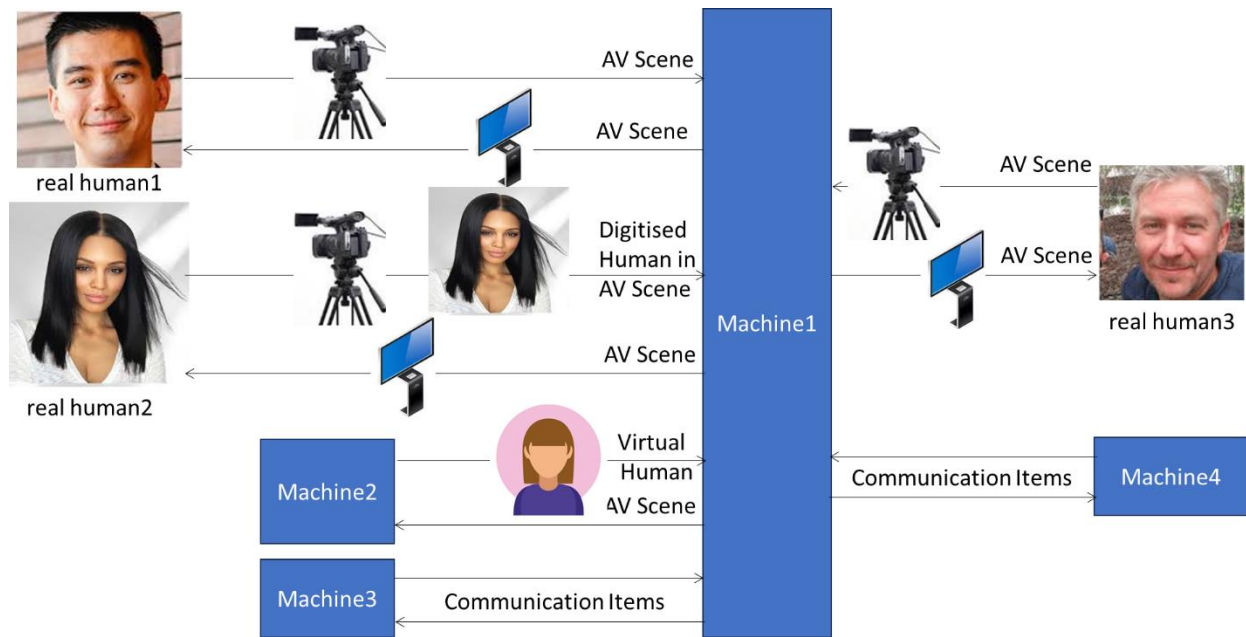


Figure 3 – Combined usage scenarios of HMC-CEC communication.

Figure 3 describes the following usage scenarios in which:

1. real human1 in his real environment and Machine1 communicate if:
 - 1.1. real human1 emits audio-visual signals in an audio-visual scene that the sensors of Machine1 convert to Audio-Visual Scenes.
 - 1.2. Machine1 generates Audio-Visual Scenes that its actuators convert to audio-visual signals.
2. real human1 and real human3 – belonging to different cultural environments – communicate if:
 - 2.1. Both real humans emit audio-visual signals in audio-visual scenes that the sensors of Machine1 convert to Audio-Visual Scenes.
 - 2.2. Machine1 converts (e.g., translates) the semantics of the Audio-Visual Scenes sensed from the audio-visual scenes where real human1 or real human 3 reside to those of the cultural environment of real human3 or real human 1, and generates Audio-Visual Scenes that its actuators convert to audio-visual signals.
3. real human1 and Machine4 communicate if:
 - 3.1. real human1 emits audio-visual signals that the Sensors of Machine1 convert to Audio-Visual Scenes.
 - 3.2. Machine1 converts the semantics of the Audio-Visual Scenes to Machine4’s cultural environment and generates either Audio-Visual Scenes or Communication Items – called Communication Items – formatted according to the Portable Avatar Format [6].
 - 3.3. Machine4 generates and emits either Visual Scenes or Communication Items in its own cultural environment.
 - 3.4. Machine1 converts (e.g., translates) the semantics of Audio-Visual Scenes or Communication Items to the semantics of Audio-Visual Scenes in real human1’s cultural environment, and emits Audio-Visual Scenes that real human1 can perceive.
4. real human2 in her real environment communicates with Machine 1 if:
 - 4.1. real human2 locates her Digitised Human in a Virtual Environment, such as the one specified by the MPAI Metaverse Model – Architecture [5].
 - 4.2. Machine1 perceives the Digitised Human in the Virtual Environment and generates a Virtual Human that real human2 can perceive. The Virtual Environment may use various means to enable real human2 to perceive the Virtual Environment.

5. Machine2 communicates with Machine1 if both Machines generate Virtual Humans in a Virtual Environment. Both Machines may communicate with the Digitised Human of point 2. above if all participants are in the same Virtual Environment.
6. Machine3 communicates with:
 - 6.1. real human3 by using the same process as in point 2. above.
 - 6.2. Machine4 by exchanging Audio-Visual Scenes or Communication Items.

Note that Communication Items may include a multimodal message (Text, Speech, Face, and Gesture), an associated Personal Status specifying Emotion, Cognitive State, and Social Attitude [4], language, and information about a Virtual Space [6].

6.1.1 Information Service

A human in a public space wants to access an information service implemented as a kiosk equipped with audio-visual sensors able to capture the space containing the human and processing functions to extract the human as an audio (speech) and visual object, while ignoring other humans and other audio and visual objects. The human may request information on an object present in the real space that the human indicates with a forefinger (see Conversation About a Scene in MPAI-MMC V2 [4]). The kiosk responds with a speaking avatar displayed by its actuators.

Figure 4 depicts the usage scenario using an appropriate subset of Figure 3.



Figure 4 - Information Service

6.1.2 Cross-Cultural Information Service

A frequently travelling human uses his portable HMC-enabled device (Machine1) designed and trained to capture the subtleties of that human's Culture. The human interacts with a local Information Service (Machine4) using Machine1 that acts as an interpreter between the human and Machine4 by exchanging Communication Items that may include the human's avatar.

Figure 5 depicts the usage scenario using the appropriate subset of Figure 3.

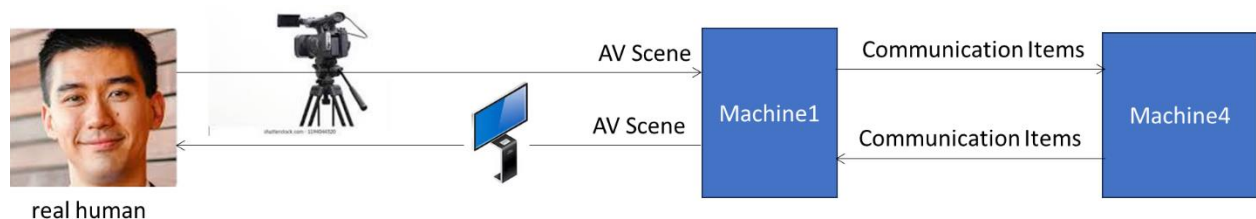


Figure 5 – Cross-Cultural Information Service

6.1.3 Virtual Assistant

This usage scenario has already been developed by MPAI-MMC V2 [4] and is used by the MPAI-PAF Avatar-Based Videoconference (PAF-ABV) [6], a videoconference whose participants are speaking avatars realistically impersonating the human participants. A speaking avatar not representing a human participant is the Virtual Secretary (generated by Machine2) which plays the role of note-taker and summariser by:

1. Listening to all Avatars' Speech.

2. Monitoring their Personal Statuses [4].
3. Drafting a Summary using the Avatars' Personal Status and Text, which may be obtained via Face and Body analysis, Speech Recognition, or Text input.

The Portable Avatar of the Virtual Secretary is distributed to all participants, who can then place it around the meeting table.

Figure 6 depicts two Digitised Human participants and one Virtual Human participant (Virtual Secretary). Machine 1 acts as cultural mediator between real human2 and real human3.

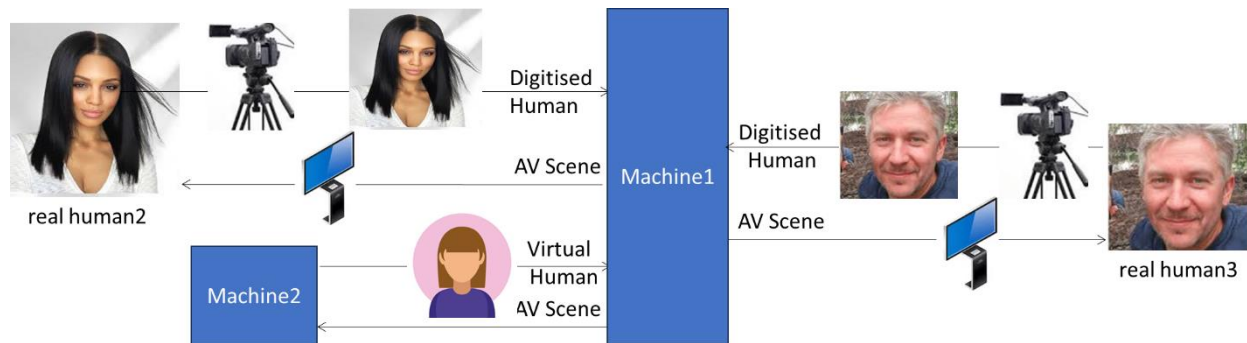


Figure 6 - Virtual Assistant

6.1.4 Conversation companion

A human is sitting in her living room wishing to converse about a topic with a Machine, represented and displayed as a speaking avatar. The human asks questions, and the Machine responds. The human displays pleasure, dissatisfaction, or other indications of Personal Status (including Emotion, Cognitive State, and Social Attitude). The Machine responds appropriately, with appropriate vocal and facial expressions.

Figure 7 illustrates the usage scenario.



Figure 7 - Conversation Companion

6.1.5 Strolling in the metaverse

User_A – a Process representing a human in an M-Instance (a metaverse instantiation) rendered as a speaking Avatar – is in a public area in the M-Instance. She is approached by User_B, a Process rendered as an animated speaking Avatar representing personnel of a company promoting a particular product. User_A does not reject the encounter. User_B captures all relevant information from the speech, face, and body of User_A's Avatar, and expresses itself by uttering relevant speech and appropriately moving its face and body. Eventually, User_A gets annoyed and calls a security entity (Machine3 in Figure 8) that deals with the complaints of User_A, using audio and visual information as if it were representing a real human.

Figure 8 illustrates the usage scenario. Note that Machine1 includes the function that enables hosting of Digitised and Virtual Humans.

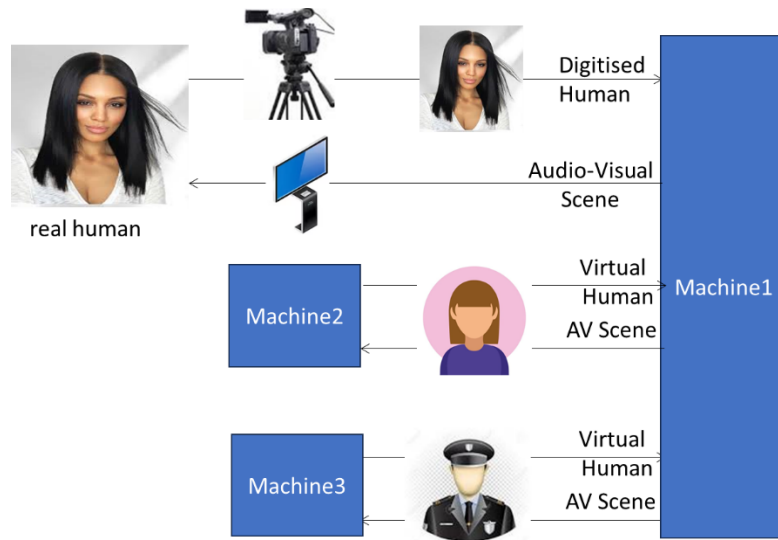


Figure 8 - Strolling in the metaverse

6.1.6 Travelling in a Connected Autonomous Vehicle

Two humans travel in a Connected Autonomous Vehicle (CAV) conversing with a Machine that performs some of the functions of the CAV's Human-CAV Interaction Subsystem [11]. The Machine is aware of the position of the human it is talking to at a particular time and directs its Avatar's gaze accordingly.

7 AI Workflows

MPAI-HMC specifies the [Communicating Entities in Context](#) AI Workflow.

7.1 AIWs Communicating Entities in Context

Communicating Entities in Context (HMC-CEC) AI Workflow (AIW) specifies [Functions](#), [Reference Model](#), and [Input and Output Data](#) of the AIW, and the [Functions](#) and the [Input and Output Data](#) of its AI Modules (AIM). Each Input and Output Data of the HMC-CEC AIW and its AIMs is linked to its online specification.

7.1.1 Functions

The *Communicating Entities in Context* AI Workflow enables Machines to communicate with Entities in different Contexts where:

1. **Machine** is software embedded in a device that implements the HMC-CEC specification.
2. **Entity** refers to one of:
 1. A human in a real audio-visual scene.
 2. A human in a real scene represented as a Digitised Human in an Audio-Visual Scene.
 3. A Machine represented as a Virtual Human in an Audio-Visual Scene.
 4. A Digital Human in an Audio-Visual Scene rendered as a real audio-visual scene.
3. **Context** is information describing the Attributes of an Entity, such as language, culture etc.
4. **Digital Human** is either a Digitised Human, i.e., the digital representation of a human, or a Virtual Human, i.e., a Machine that can be rendered for human perception as humanoids.
5. A word beginning with a small letter represents an object in the **real world**. A word beginning with a capital letter represents an Object in the **Virtual World**.

Entities communicate in one of the following ways:

1. When communicating to **humans**:
 1. Use the Entities' body, speech, Context, and the audio-visual scene that the Entities are immersed in.
 2. Use HMC-CEC-enabled Machines emitting Communication Items.
2. When communicating to **Machines**:
 1. Render Entities as speaking humanoids in audio-visual scenes, as appropriate.
 2. Emit Communication Items.

Communication Items are implementations of Portable Avatar, a Data Type providing information on an Avatar and its context to enable a receiver to render an Avatar as intended by the sender.

HMC-CEC assumes that:

1. *Input Audio* and *Input Visual* are Audio Object and Visual Object, respectively.
2. *Output Audio* and *Output Visual* convey audio and visual information rendered by the Audio-Visual Rendering AIM.
3. The *real space* is digitally represented as an Audio-Visual Scene that includes the communicating human and may include other humans and generic objects.
4. The *Virtual Space* contains a Digital Humans and/or its Audio components and may include other Digital Humans and generic Objects in an Audio-Visual Scene.
5. The *Machine* can:
 1. Understand the semantics of the Communication Item at different layers of depth.
 2. Produce a multimodal response expected to be congruent with the received information.
 3. Render the response as a speaking Virtual Human in an Audio-Visual Scene.
 4. Convert the semantics of the information produced by an Entity to a form that is compatible with the Context of another Entity.

An AI Module is specified only by its Functions and Interfaces. Implementers are free to use their preferred technologies to achieve the Functions providing the features while respecting the constraints of the interfaces. An implementation may subdivide a given AI Module into more than one AI Module, provided that the combined AI Module exposes the interfaces of the corresponding AI Modules of the HMC-CEC Specification. An implementation may combine AI Modules into one, provided that the resulting AI Module exposes the interfaces of the corresponding AI Modules of the HMC-CEC Specification.

Usage Scenarios offer a collection of example applications enabled by HMC-CEC.

7.1.2 Reference Model

Figure 1 depicts the Reference Model of the *Communicating Entities in Context (HMC-CEC)* Use Case implemented as an AI Workflow (AIW) that includes *AI Modules (AIM)* per *Technical Specification: AI Framework (MPAI-AIF)*. Three out of six AIMs in Figure 1 (*Audio-Visual Scene Description, Entity Context Understanding, and Personal Status Display*) are Composite AIMs, i.e., they include interconnected AIMs. An introduction to MPAI-AIF is provided [here](#).

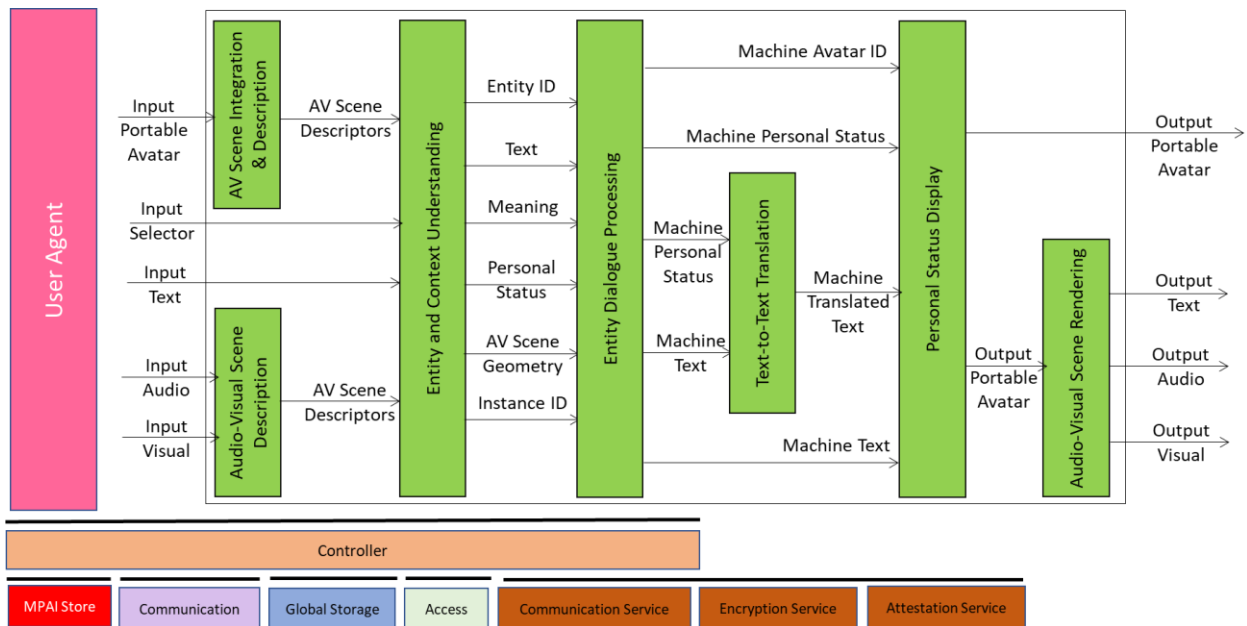


Figure 9 – Human-Machine Communication AIW

Note that:

1. Words beginning with a capital are defined in Definitions, Words beginning with a small letter have the commonly understood meaning.
2. The *Input Selector* enables the Entity to inform the Machine through the Entity and Context Understanding AIM about use of Text vs. Speech in the communication, Language Preferences, and Selected Language in translation.
3. The Machine captures the information emitted by the Entity and its Context through *Input Text*, *Input Speech*, *Input Audio* and *Input Visual*.
4. The *Input Portable Avatar* is the Communication Item emitted by a communicating Machine.
5. The *Audio-Visual Scene Descriptors* are digital representations of a real audio-visual scene or a Virtual Audio-Visual Scene produced either by the *Audio-Visual Scene Description* AIM or the *Audio-Visual Scene Integration and Description* AIM.
6. To facilitate identification, AIMs are labelled with three letters indicating the Technical Specification that specifies it, followed by a hyphen “-”, followed by three letters uniquely identifying the AIM defined by that Technical Specification. For instance, Portable Avatar Demultiplexing is indicated as PAF-PDX where PAF refers to *Technical Specification: Portable Avatar Format (MPAI-PAF)* and PDX refers to the [Portable Avatar Demultiplexing](#) AIM specified by MPAI-PAF.

7.1.3 Input/Output Data

Table 3 gives the Input/Output Data of the MPAI-HMC AIW.

Table 3 - Input/Output Data of the MPAI-HMC AIW

Input	Description
Portable Avatar	A Communication Item emitted by the Entity communicating with the ego Entity.
Input Selector	Selector containing data specifying the media and the language used in the communication.

Input Text	Text Object generated by the communicating Entity as information additional to or in lieu of Speech Object.
Input Audio	The audio scene captured by the Machine.
Input Visual	The visual scene captured by the Machine.
Output	Description
Portable Avatar	The Communication Item produced by the Machine.
Output Audio	The rendered audio corresponding to the Audio in the Communication Item.
Output Visual	The rendered visual corresponding to the visual in the Communication Item.
Output Text	The Text contained in a Communication Item or associated with Output Audio and Output Visual.

7.1.4 Functions of AI Modules

Table 4 gives the functions of HMC-CEC AIMs.

Table 4 – Functions of AI Modules

AIM	Functions
Audio-Visual Scene Integration and Description	Adds Avatar to Audio-Visual Scene in Portable Avatar providing Audio-Visual Scene Descriptors.
Audio-Visual Scene Description	Provides Audio-Visual Scene Descriptors.
Entity Context Understanding	Understands the information emitted by the Entity and its Context.
Entity Dialogue Processing	Produces Text and Personal Status of Machine in response to inputs.
Text-to-Text Translation	Produces Machine Translated Text from Machine Text and Personal Status.
Personal Status Display	Produces Portable Avatar.
Audio-Visual Scene Rendering	Renders the content of the Portable Avatar.

Table 5 gives the I/O Data of the AIMs of HMC-CEC. Note that an ID can either be specified as an [Instance Identifier](#) or refer to a generic identifier.

Table 5 – Input/Output Data of AI Modules

AIM	Receives	Produces
Audio-Visual Scene Integration and Description	Input Portable Avatar	Audio-Visual Scene Descriptors
Audio-Visual Scene Description	Input Audio Input Visual	Audio-Visual Scene Descriptors
Entity Context Understanding	Audio-Visual Scene Descriptors Input Text Input Selector	Audio-Visual Scene Geometry Personal Status Entity ID Text Meaning Instance Identifier

Entity Dialogue Processing	Audio-Visual Scene Geometry Personal Status Entity ID Text Meaning Instance Identifier	Machine Personal Status Machine Avatar ID Machine Text
Text-to-Text Translation	Machine Text Machine Personal Status	Machine Translated Text
Personal Status Display	Machine Personal Status Machine Avatar ID Machine Text	Output Portable Avatar
Audio-Visual Scene Rendering	Output Portable Avatar	Output Text Output Audio Output Visual

7.1.5 AIW, AIMS, and JSON Metadata

Table 6 – AIW, AIMS, and JSON Metadata

AIW	AIMs/1	AIMs/2	AIMs/3	Name	JSON
HMC-CEC	HMC-SID			Communicating Entities in Context	X
	HMC-SID			AV Scene Integration and Description	X
	OSD-AVS			Audio-Visual Scene Description	X
		CAE-ASD		Audio Scene Description	X
			CAE-AAT	Audio Analysis Transform	X
			CAE-ASL	Audio Source Localisation	X
			CAE-ASE	Audio Separation and Enhancement	X
			CAE-AST	Audio Synthesis Transform	X
			CAE-AMX	Audio Descriptors Multiplexing	X
		OSD-VSD		Visual Scene Description	X
		OSD-AVA		Audio-Visual Alignment	X
	HMC-ECU			Entity And Context Understanding	X
		OSD-SDX		Audio-Visual Scene Demultiplexing	X
		MMC-ASR		Automatic Speech Recognition	X
		OSD-VOI		Visual Object Identification	X
			OSD-VDI	Visual Direction Identification	X
			OSD-VOE	Visual Object Extraction	X
			OSD-VII	Visual Instance Identification	X
		CAE-AOI		Audio Object Identification	X
		MMC-NLU		Natural Language Understanding	X
		MMC-PSE		Personal Status Extraction	X
			MMC-ITD	Entity Text Description	X
			MMC-ISD	Entity Speech Description	X
			PAF-IFD	Entity Face Description	X
			PAF-IBD	Entity Body Description	X
			MMC-PTI	PS-Text Interpretation	X

			MMC-PSI	PS-Speech Interpretation	X
			PAF-PFI	PS-Face Interpretation	X
			PAF-PGI	PS-Gesture Interpretation	X
			MMC-PMX	Personal Status Multiplexing	X
		MMC-TTT		Text-to-Text Translation	X
	MMC-EDP			Entity Dialogue Processing	X
	MMC-TTT			Text-to-Text Translation	X
	PAF-PSD			Personal Status Display	X
		MMC-TTS		Text-to-Speech	X
		PAF-IFD		Entity Face Description	X
		PAF-IBD		Entity Body Description	X
		PAF-PMX		Portable Avatar Multiplexing	X
	PAF-AVR			Audio-Visual Scene Rendering	X

8 AI Modules

HMC-CEC specifies one Composite and one Basic AI Module.

Table 7 – Basic and Composite AI Modules

Acronym	Specification	JSON
HMC-ECU	Entity and Context Understanding	X
HMC-SID	AV Scene Integration and Description	X

Table 8 provides the full list of with web links to the AI Modules utilised by HMC-CEC organised according to the Technical Specifications specifying them.

Table 8 – AI Modules utilised by HMC-CEC

CAE		MMC		OSD		PAF	
Audio Analysis Transform	X	Automatic Speech Recognition	X	Audio-Visual Alignment	X	Audio-Visual Scene Rendering	X
Audio Descriptors Multiplexing	X	Entity Dialogue Processing	X	Audio-Visual Event Description	X	Face Identity Recognition	X
Audio Object Identification	X	Entity Speech Description	X	Audio-Visual Scene Demultiplexing	X	Entity Body Description	X
Audio Scene Description	X	Entity Text Description	X	Audio-Visual Scene Description	X	Entity Face Description	X
Audio Separation and Enhancement	X	Natural Language Understanding	X	Visual Direction Identification	X	Portable Avatar Demultiplexing	X
Audio Source Localisation	X	Personal Status Extraction	X	Visual Instance Identification	X	PS-Face Interpretation	X
Audio Synthesis Transform	X	Personal Status Multiplexing	X	Visual Object Extraction	X	PS-Gesture Interpretation	X
HMC		PS-Speech Interpretation	X	Visual Object Identification	X	Personal Status Display	X
AV Scene Integration and Description	X	PS-Text Interpretation	X	Visual Scene Description	X	Portable Avatar Multiplexing	X

Entity and Context Understanding	X	Speaker Identity Recognition	X			
		Text and Speech Translation	X			
		Text-To-Speech	X			
		Text-to-Text Translation	X			

8.1 Entity and Context Understanding

8.1.1 Functions

The functions of Entity and Context Understanding (HMC-ECU) allow a Machine to achieve understanding the information conveyed by an Entity and its Context in order to enable the Entity Dialogue Processing AIM to produce a pertinent communication.

Therefore, Entity and Context Understanding (HMC-ECC):

1. Receives the Audio-Visual Scene Descriptors.
2. Separates the components of the Audio-Visual Scene Descriptors.
3. Performs
 1. Recognition of Speaker ID.
 2. Recognition of Face ID.
 3. Recognition of Entity's Speech.
 4. Recognition of Audio Object and Visual Object.
 5. Understanding of Entity's Natural Language expressed as Text in the Context of the Audio and Visual Instance.
 6. Extraction of the Entity's Personal Status.
 7. Translation of the Entity's Text.
4. Produces:
 1. Audio-Visual Scene Geometry
 2. Entity ID
 3. Audio Instance ID
 4. Visual Instance ID
 5. Personal Status
 6. Translated and Refined Text
 7. Meaning.

8.1.2 Reference Model

Figure 10 depicts the Reference Architecture of the Entity and Context Understanding AIM.

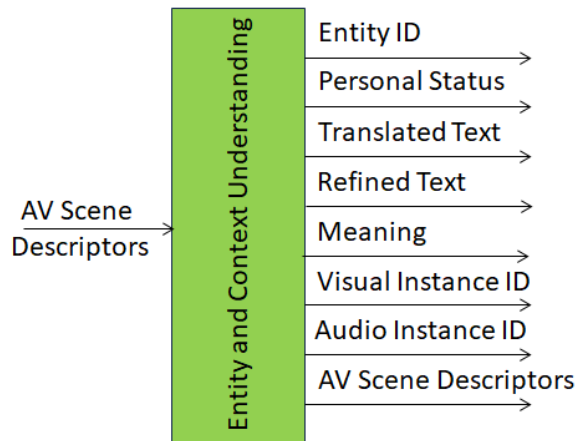


Figure 10 – The Entity and Context Understanding Composite AIM

8.1.3 I/O Data

Table 9 specifies the Input and Output Data of the of the Entity Context Understanding AIM.

Table 9 – I/O Data of the Entity Context Understanding Composite AIM

Input	Description
Audio-Visual Scene Descriptors	The digital representation of the Audio, Visual, and Audio-Visual Objects of the Scene and their spatial arrangement.
Output	Description
Entity ID	
Personal Status	Personal Status of Entity having the Entity ID.
Translated Text	Translated Text of Text Object or of Text conveyed by Speech Object.
Refined Text	Refined Text of Speech Object.
Meaning	Other name for Refined Text Descriptors.
Visual Instance ID	The Identifier of the specific Visual Object belonging to a level in the taxonomy.
Audio-Visual Scene Descriptors	As in Input
Audio Instance ID	The Identifier of the specific Audio Object belonging to a level in the taxonomy.

8.1.4 SubAIMs

HMC-ECU is a Composite AIM having the Reference Model depicted in Figure 11.

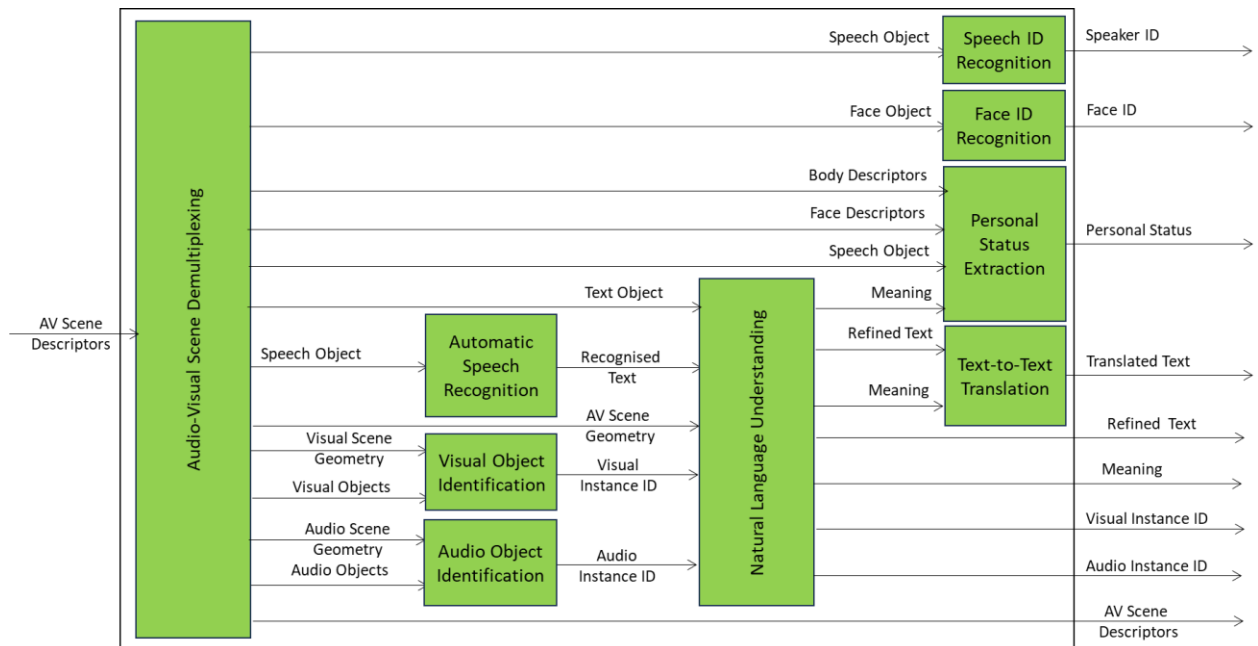


Figure 11 – The Entity and Context Understanding Composite AIM

Table 10 provides the list of AIMs – both Basic and Composite – included in the Entity and Context Understanding Composite AIM.

Table 10 – AIW, AIMs, and JSON Metadata

AIMs	Name
HMC-ECU	Entity And Context Understanding
OSD-SDX	Audio-Visual Scene Demultiplexing
MMC-SIR	Speaker Identity Recognition
PAF-FIR	Face Identity Recognition
MMC-ASR	Automatic Speech Recognition
OSD-VOI	Visual Object Identification
CAE-AOI	Visual Object Identification
MMC-NLU	Natural Language Understanding
MMC-PSE	Personal Status Extraction
MMC-TTT	Text-to-Text Translation

8.1.5 JSON Metadata

<https://schemas.mpai.community/HMC/V1.1/AIMs/EntityContextUnderstanding.json>

8.1.6 Profiles

The Profiles of Entity Context Understanding AIM are [specified](#).

8.2 AV Scene Integration and Description

8.2.1 Functions

AV Scene Integration and Description (HMC-SID) performs the following functions:

Receives	Portable Avatar	
----------	-----------------	--

Adds	The Avatar in the Input Portable Avatar to the Audio-Visual Scene Descriptors conveyed by the Input Portable Avatar with an appropriate Spatial Attitude. If the Input Portable Avatar does not include a Scene, the AV Scene Integration and Description AIM uses a generic Scene.
Produces	The Audio-Visual Scene Descriptors of the resulting Audio-Visual Scene.

8.2.2 Reference Architecture

Figure 12 depicts the HMC-SID Reference Architecture.

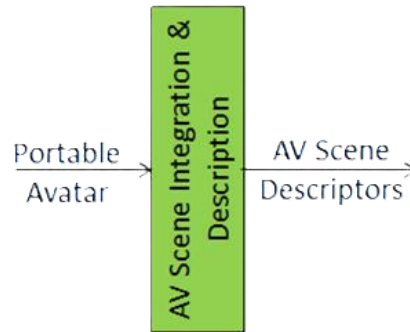


Figure 12 – The AV Scene Integration and Description AIM

8.2.3 I/O Data

The Input and Output Data of are specified in Table 11.

Table 11 – I/O Data of the AV Scene Integration and Description AIM

Input	Description
Portable Avatar	A Communication Item from a Machine Entity.
Output	Description
Audio-Visual Scene Descriptors	The Descriptors of the AV Scene where the Avatar conveyed by the Input Portable Avatar has been added to the Scene with the appropriate Spatial Attitude.

8.2.4 SubAIMs

No SubAIMs.

8.2.5 JSON Metadata

<https://schemas.mpai.community/HMC/V1.1/AIMs/AVSceneIntegrationAndDescription.json>

8.2.6 Profiles

No Profiles

9 Data Types

MPAI-HMC V1.0 only uses Data Types defined by other MPAI Technical Specifications. Table 12 provides the full list with web links of the Data Types utilised by HMC-CEC organised according to the Technical Specifications that specify them.

Table 12 – AI Modules utilised by HMC-CEC

MPAI-AIF		MPAI-MMC		MPAI-OSD		MPAI-PAF	
Neural Network Model	X	Cognitive State	X	Audio-Visual Basic Scene Descriptors	X	Avatar	X
MPAI-CAE		Emotion	X	Audio-Visual Basic Scene Geometry	X	Body Descriptors	X
Audio Basic Scene Descriptors	X	Intention	X	Audio-Visual Event Descriptors	X	Face Descriptors	X
Audio Basic Scene Geometry	X	Meaning	X	Audio-Visual Object	X	Model	X
Audio Object	X	Personal Status	X	Audio-Visual Scene Descriptors	X	Portable Avatar	X
Audio Scene Descriptors	X	Social Attitude	X	Audio-Visual Scene Geometry	X		
Audio Scene Geometry	X	Speech Descriptors	X	Input Visual	X		
		Speech Object	X	Instance Identifier	X		
		Text Descriptors	X	Point of View	X		
		Text Object	X	Selector	X		
				Space-Time	X		
				Spatial Attitude	X		
				Time	X		
				Visual Basic Scene Descriptors	X		
				Visual Basic Scene Geometry	X		
				Visual Object	X		
				Visual Scene Descriptors	X		
				Visual Scene Geometry	X		

10 Profiles

HMC-CEC uses six groups of capability classes to process a Communication Item:

- Receives Communication Items from a Machine or Audio-Visual Scenes from a real space.
- Extracts Personal Status from the Modalities (Text, Speech, Face, or Gesture) in the Communication Item received.
- Understands The Communication Item from the Modalities and the extracted Personal Status, with or without use of the spatial information embedded in the Communication Item.

Translates Using the set of Modalities available to the Machine.
 Generates Response.
 Displays The response using available Modalities.

The Attributes and Sub-Attributes of the HMC-CEC Profiles are defined in Table 12 where the Sub-Attributes are expressed with three characters where the first two are followed by O (representing Object):

1. Audio-Visual Scene represent Text (TXO), Speech (SPO), Audio (AUO), Visual (VIO), and Portable Avatar (PAF) Sub-Attributes, respectively.
2. Personal Status, Understanding, Translation, and Display Response represent Text (TXO), Speech (SPO), Face (FCO), and Gesture (GSO), respectively.

The SPC Sub-Attribute of Understanding represents Spatial Information (SPaCe), i.e., the additional capability of an HMC-CEC implementation to use Spatial Information to understand a Communication Item.

Table 13 – Attribute and Sub-Attribute Codes of HMC-CEC

Attributes	Codes	Sub-Attribute Codes				
Audio-Visual Scene	AVS	TXO	SPO	AUO	VIO	PAF
Personal Status	EPS	TXO	SPO	FCO	GSO	
Understanding	UND	TXO	SPO	FCO	GSO	SPC
Translation	TRN	TXO	SPO	FCO	GSO	
Display Response	RES	TXO	SPO	FCO	GSO	

The formal specification of MPAI-HMC Profiles is provided by the following JSON file:
<https://schemas.mpai.community/HMC/V1.1/data/CECProfiles.json>.

Annex 1 - Notices and Disclaimers Concerning MPAI Standards (Informative)

The notices and legal disclaimers given below shall be borne in mind when [downloading](#) and using approved MPAI Standards.

In the following, “Standard” means the collection of four MPAI-approved and [published](#) 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](#). 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](#). 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](#) to determine the status of any given published MPAI Standard.

Description 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 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](#) 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](#). 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 2 - Patent Declarations

Technical Specification: Human and Machine Communication (MPAI-HMC) has been developed according to the process outlined in the MPAI Statutes [11] and the MPAI Patent Policy [12] using elements already developed in other MPAI Technical Specifications with the addition of a few more new elements.

The following table will include references to the entities declaring to agree to licence their standard essential patents reading on *Technical Specification: Human and Machine Communication (MPAI-HMC)* according to the MPAI-HMC Framework Licence [13]:

Entity	Name	email address