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

|  |  |
| --- | --- |
| **N945** | 2022/11/21 |
| **Source** | Requirements (ARA) |
| **Title** | Technical Specification - Avatar Representation and Animation (MPAI-ARA) WD0.2 |
| **Target** | MPAI Community |

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

**MPAI Technical Specifiction**

**Avatar Representation and Animation**

**MPAI-ARA**

|  |
| --- |
| **WD0.0.0.1** |

**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 use of this Technical Specification.

Readers are invited to review Annex 2 - Notices and Disclaimers.

© Copyright MPAI 2021-22. All rights reserved

**Technical Specification**

**Avatar Representation and Animation**

**(Under development)**

[1 Introduction 4](#_Toc119158138)

[2 Scope 6](#_Toc119158139)

[2.1 General 6](#_Toc119158140)

[2.2 Normative content of the Use Case 6](#_Toc119158141)

[3 Terms and Definitions 6](#_Toc119158142)

[4 References 9](#_Toc119158143)

[4.1 Normative References 9](#_Toc119158144)

[4.2 Informative References 10](#_Toc119158145)

[5 Avatar-Based Videoconference 10](#_Toc119158146)

[5.1 Scope of Use Case 10](#_Toc119158147)

[5.2 Transmitting Client 11](#_Toc119158148)

[5.2.1 Function 11](#_Toc119158149)

[5.2.2 Architecture 11](#_Toc119158150)

[5.2.3 Input and output data 12](#_Toc119158151)

[5.2.4 AI Modules 13](#_Toc119158152)

[5.3 Server 13](#_Toc119158153)

[5.3.1 Function 13](#_Toc119158154)

[5.3.2 Architecture 14](#_Toc119158155)

[5.3.3 I/O data 14](#_Toc119158156)

[5.3.4 AI Modules 15](#_Toc119158157)

[5.4 Virtual Secretary 15](#_Toc119158158)

[5.4.1 Architecture 15](#_Toc119158159)

[5.4.2 I/O Data 17](#_Toc119158160)

[5.4.3 AI Modules 17](#_Toc119158161)

[5.5 Receiving Client 17](#_Toc119158162)

[5.5.1 Function 17](#_Toc119158163)

[5.5.2 Architecture 18](#_Toc119158164)

[5.5.3 I/O Data 18](#_Toc119158165)

[5.5.4 AI Modules 18](#_Toc119158166)

[6 Composite AI Modules 19](#_Toc119158167)

[6.1 Personal Status Extraction 19](#_Toc119158168)

[6.1.1 Scope of Composite AIM 19](#_Toc119158169)

[6.1.2 Reference architecture 19](#_Toc119158170)

[6.1.3 Input/output data 20](#_Toc119158171)

[6.1.4 AI Modules 20](#_Toc119158172)

[6.2 Personal Status Display 20](#_Toc119158173)

[6.2.1 Scope of Composite AIM 20](#_Toc119158174)

[6.2.2 Reference Architecture 21](#_Toc119158175)

[6.2.3 Input/output data 21](#_Toc119158176)

[6.2.4 AI Modules 21](#_Toc119158177)

[7 Data Formats 22](#_Toc119158178)

[7.1 Audio Scene Descriptors 22](#_Toc119158179)

[7.2 Cognitive State 22](#_Toc119158180)

[7.3 Digital Human Descriptos 22](#_Toc119158181)

[7.4 Emotion 22](#_Toc119158182)

[7.5 Face Descriptors for Motion 22](#_Toc119158183)

[7.6 Face Descriptors for Recognition 22](#_Toc119158184)

[7.7 Face Descriptors for Personal Status 22](#_Toc119158185)

[7.8 Gesture Descriptors for Personal Status 22](#_Toc119158186)

[7.9 Language identifier 22](#_Toc119158187)

[7.10 Meaning 22](#_Toc119158188)

[7.11 Personal Status 22](#_Toc119158189)

[7.12 Social Attitude 22](#_Toc119158190)

[7.13 Spatial Attitude 22](#_Toc119158191)

[7.14 Speech Descriptors for Recognition 22](#_Toc119158192)

[7.15 Text 22](#_Toc119158193)

[7.16 User ID 22](#_Toc119158194)

[7.17 Visual Scene Descriptors 22](#_Toc119158195)

[Annex 1 - General MPAI Terminology 23](#_Toc119158196)

[Annex 2 - Notices and Disclaimers Concerning MPAI Standards (Informative) 26](#_Toc119158197)

[Annex 3 - The Governance of the MPAI Ecosystem (Informative) 28](#_Toc119158198)

[Annex 4 - Industry terms 30](#_Toc119158199)

# Introduction

In recent years, Artificial Intelligence (AI) and related technologies have been introduced in 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 rule, MPAI standards include four documents: Technical Specification, Reference Software Specifications, Conformance Testing Specifications, and Performance Assessment Specifications.

The last type of Specification includes standard operating procedures to enable users of MPAI Implementations to make informed decision about their applicability based on the notion of Performance, defined as a set of attributes characterising a reliable and trustworthy implementation.

In the following, Terms beginning with a small letter have the commonly used meaning and a capital letter are defined in *Table 1* if they are specific to this Technical Report and in *Table 14* if they are common to all MPAI Standards.

In general, MPAI Application Standards are defined as aggregations – called AI Workflows (AIW) – of processing elements – called AI Modules (AIM) – executed in an AI Framework (AIF). 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 AIF that executes an AIW. The AIW and its AIMs 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 reference [7].

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

MPAI Application Standards normatively specify the 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.

Diagram

Description automatically generated

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

It should be noted that 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 Standards are designed to enable a User to obtain, via standard protocols, an Implementation of an AIW and of the set of corresponding AIMs and execute it in an AIF Implementation. The MPAI Store in *Figure 1* is the entity from which Implementations are downloaded. MPAI Standards assume that the AIF, AIW, and AIM Implementations may have been developed by independent implementers. A necessary condition for this to be possible, is that any AIF, AIW, and AIM implementations be uniquely identified. MPAI has appointed an ImplementerID Registration Authority (IIDRA) to assign unique ImplementerIDs (IID) to Implementers.[[1]](#footnote-1)

A necessary condition to make possible the operations described in the paragraph above is the existence of an ecosystem composed of Conformance Testers, Performance Assessors, the IIDRA and an instance of the MPAI Store. Reference [7] provides an informative example of such ecosystem.

# Scope

## General

Avatar Representation and Animation (MPAI-ARA) specifies the technologies enabling the implementation of the Avatar-Based Videoconference Use Case specified in Chapter 5 - Avatar-Based Videoconference, specifically:

1. The Representation of a Digital Environment populated with objects.
2. The Representation of a Digital Human.
3. The Representation of the features of a human.
4. The Representation of a Digital Human Model of the human.
5. The Representation of the motion of a human.
6. The Animation of a Digital Human

The current version of the Avatar Representation and Animation Technical Specification has been developed by the Requirements Standing Committee. MPAI may issue new versions of MPAI-ARA extending or replacing the current Technical Specification.

## Normative content of the Use Case

The Use Case normatively defines:

1. The Functions of the AIWs and of the AIMs.
2. The Connections between and among the AIMs
3. The Semantics and the Formats of the input and output data of the AIW and the AIMs.

The word *normatively* implies that an Implementation claiming Conformance to:

1. An *AIW*, shall:
   1. Perform the AIW function specified in the appropriate Section of Chapter.
   2. All AIMs, their topology and connections should conform with the AIW Architecture specified in the appropriate Section of Chapter.
   3. The AIW and AIM input and output data should have the formats specified in the appropriate Subsection of Section.
2. An *AIM*, shall:
   1. Perform the AIM function specified by the appropriate section of Chapter.
   2. Receive and produce the data specified in the appropriate Subsection of Section.
   3. Receive as input and produce as output data having the format specified in Section.

3. A data *Format*, the data shall have the format specified in Section.

Users of this Technical Specification should note that:

1. This Technical Specification defines Interoperability Levels but does not mandate any.
2. Implementers decide the Interoperability Level their Implementation satisfies.
3. Implementers can use the Reference Software of this Technical Specification to develop their Implementations.
4. The Conformance Testing specification can be used to test the conformity of an Implemen­tation to this Standard.
5. Performance Assessors can assess the level of Performance of an Implementation based on the Performance Assessment specification of this Standard.
6. Implementers and Users should consider the notices and disclaimers of Annex 2.

# Terms and Definitions

The terms used in this standard whose first letter is capital have the meaning defined in *Table 1* where Terms that are preceded by a dash “-”

1. Refer to a Term preceding it that has no dash.
2. Mean the following:
   1. If the Term has a normal font, the Term preceding the one with a dash should come after that Term.
   2. If the Term has an *italic font,* the Term without dash preceding the one with a dash should be placed before that Term.

Notation 2.a is used to concentrate in one place all Terms that refer to Metaverse, e.g., Metaverse Asset, Component, Instance, Manager, Object, Operator, Partner, Profile, Service and End User, or *Audio, Human, and Object* Metadata.

*Table 1 – Terms and Definitions*

|  |  |
| --- | --- |
| **Term** | **Definition** |
| Attitude |  |
| * Social | An element of the internal status 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”. |
| * Spatial | Position and Orientation and their velocities and accelerations of a Human and Physical 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. |
| Authentication | The process of determining whether a User is who they state they are. |
| Avatar | A rendered Digital Human. |
| Data | Information in digital form. |
| Descriptor | Coded representation of text, audio, speech, or visual feature. |
| Device | A piece of equipment used to interact and have Experience in a Digital Environment. |
| 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”. |
| Emotion Recognition | An AIM that recognises the best member, or members, of a finite set of Emotions, as conveyed by a Text or Speech Segment or Video. |
| Environment | A Physical or Digital space. |
| * Model | The static audio and visual components of the Environment, e.g., walls, table. And chairs. |
| Experience | The state of a human whose senses are continuously affected for a meaningful period. |
| Face | The portion of a 2D or 3D digital representation corresponding to the face of a human. |
| Factor | One of Emotion, Cognitive State and Attitude. |
| Grade | The intensity of a Factor. |
| Identifier | The label uniquely associated with a User or a Device or an Object. |
| Identity | A name uniquely connected to a User. |
| Metadata | An attribute of Data. |
| Meaning | Information extracted from Text such as syntactic and semantic information, and Personal Status. |
| Modality | One of Text, Speech, Face, or Gesture. |
| Object | A data structure that can be rendered to cause an Experience. |
| * *Audio* | Coded representation of Audio information with its metadata. An Audio Object can include Audio Objects. |
| * *Audio-Visual* | Coded representation of Audio-Visual information with its metadata. An Audio-Visual Object can includeAudio-Visual Objects. |
| * Descriptor | An individual attribute of the coded representation of an object in a Scene, including its Spatial Attitude. |
| * *Digital* | A Digitised or a Virtual Object. |
| * *Digitised* | The digital representation of a real object. |
| * *Visual* | Coded representation of Visual information with its metadata. A Video Object can include Video Objects. |
| * *Virtual* | An Object not representing an object in a Real Environment. |
| Orientation | The 3 yaw, pitch, and roll (α,β,γ) angles of a representative point of an object in the Real and Virtual Space. |
| Persona | A manifestation of a User as a rendered Digital Human. |
| Personal Status | The ensemble of information internal to a person, including Emotion, Cognitive State, and Attitude. |
| Physical Gesture | A movement of the body or part of it, such as the head, arm, hand, and finger, often a complement to a vocal utterance. |
| Pitch | The fundamental frequency of Speech. Pitch is the attribute that makes it possible to judge sounds as "higher" and "lower." |
| Point of View | The Spatial Attitude of a Digital Human watching the Environment. |
| Position | The 3 coordinates (x,y,z) of a representative point of an object in a Real or Virtual space. |
| Scene | An Environment populated by Objects in a Digital space. |
| * *Audio* | The Audio Objects of an Environment with Object location metadata. |
| * *Audio-Visual* | (AV Scene) The Audio-Visual Objects of an Environment with Object location metadata. |
| * Description |  |
| * *Visual* |  |
| * Presentation | The rendering of Data in a format suitable for human perception. |
| Text | A sequence of characters drawn from a finite alphabet. |
| Representation | Data that represent an entity of a Real Environment. |

# References

## Normative References

This standard normatively references the following documents, both from MPAI and other stan­dards organisations. MPAI standards are publicly available at .

1. Technical Specification; AI Framework (MPAI-AIF) V1.1; https://mpai.community/standards/mpai-aif/
2. Technical Specification: Context-based Audio Enhancement (MPAI-CAE) V1.4; https://mpai.community/standards/mpai-cae/
3. ISO 639; Codes for the Representation of Names of Languages — Part 1: Alpha-2 Code.
4. ISO/IEC 10646; Information technology – Universal Coded Character Set
5. MPAI; The MPAI Statutes; https://mpai.community/statutes/
6. MPAI; The MPAI Patent Policy; https://mpai.community/about/the-mpai-patent-policy/.

## Informative References

The references provided here have an information purpose.

1. Technical Specification; The governance of the MPAI ecosystem (MPAI-GME), V1.1; https://mpai.community/standards/mpai-gme/

# Avatar-Based Videoconference

## Scope of Use Case

In this Use Case participants entrust the image of the upper part of their bodies to avatars who utter the participants’ real voice.

Graphical user interface, chart

Description automatically generated

Figure 2 – Avatar-Based Videoconference end-to-end diagram

With reference to Figure 2, the elements characterising the Use Case are:

1. Geographically dispersed humans, some of which are co-located, participate in a videoconference represented by avatars with high similarity with their human twins. The members of a groups have an individual participation in the Virtual Environment where the avatar-based videoconference takes place.
2. The videoconference room in the Virtual Environment is equipped with a table and an appropriate number of chairs hosting:
   1. Virtual Twins of human resembling their Human Twins represented as the upper part of the avatars (waist up).
   2. Avatars not corresponding to any participant, in particular a Virtual Secretary.
3. Each participant selects their avatar model.
4. Each participant uses the transmitting part of their client (Transmitting Client) to send the Server:
   1. The participant’s avatar model, language preference, and speech.
   2. The Avatar Descriptors of the participant’s Face and Gesture.
5. The Server
   1. Authenticates all participants using Speech and Face Descriptors.
   2. Sends the Descriptors of the Visual Objects in the Environment, e.g., walls, furniture, etc. to all participants.
   3. Forwards the avatar models and descriptors to all participants.
   4. Uses the language preferences to translate active speech signals to the requested languages.
   5. Sends the original and translated speech signals to participants according to their language preferences.
6. The Virtual Secretary
   1. Collects the statements made by participating avatars while monitoring the avatars’ Personal Statuses conveyed by their speech, face, and gesture.
   2. Makes a summary by combining all information.
   3. Displays the summary in the Environment for avatars to read and edit the Summary directly.
   4. Edits the Summary based on conversations with avatars using Text, Speech, and Personal Status conveyed by Text, Speech, Face and Gesture.
7. Each participant uses the receiving part of their client (Receiving Client) to:
   1. Place all avatar models on a chair of their choice.
   2. Animate the avatar models with the received descriptors.
   3. Attach the speech signals to the mouths of the corresponding avatars, so that each speech is properly located in the Environment (Spatial Audio).
   4. Select a Point of View, possibly different from the position assigned to his/her avatar in the Environment.
   5. Watch the Virtual Environment with the device of his/her choice (HMD or 2D display) and listen to the resulting spatial audio with the device of his/her choice (HMD or earpads).

It should be noted that Figure 2 depicts a particular implementation of the AVB Use Case where the clients execute all functionalities that a client can meaningfully provide. This is clearly not the only possible partitioning. For instance, a participant could use a dumb client which only controls scene creation and presents the rendered audio-visual scene generated by the server. Ditto for the transmitting part of the client where the Avatar Descriptors could be created by the server at the cost of sending high bitrate private data.

## Transmitting Client

### Function

The function of the Transmitting Client is to:

1. Receive:
   1. Input Audio from the microphone (array).
   2. Input Video of the participant from the camera (array).
   3. Participant’s Avatar Model.
   4. Participant’s spoken language preferences (e.g., EN-US, IT-CH).
2. Send to the Server:
   1. Speech Descriptors (for Authentication).
   2. Face Descriptors (for Authentication).
   3. Participant’s spoken language preferences.
   4. Speech (for distribution to all participants).
   5. Avatar Model (for distribution to all participants).
   6. Avatar Descriptors (for distribution to all participants).

### Architecture

At the start:

1. Each participant sends to the Server:
   1. Language preferences
   2. Avatar model.
2. Speaker Identification sends to Server: Descriptors for Authentication.
3. Face Identification sends to Server: Face Descriptors for Authentication.

During the meeting the following AIMs of the Transmitting Client produce:

1. Audio Scene Description: participants’ speech and location.
2. Visual Scene Description: Face Object and Human Object.
3. Speech Description: Speech Descriptors for Authentication.
4. Speech Recognition: Recognised Text.
5. Face Description1: Face Descriptors.
6. Gesture Description: Gesture Descriptors.
7. Face Description2: Face Descriptors for Authentication.
8. Personal Status Extraction: Personal Status.
9. Language Understanding: Meaning.
10. Avatar Description: Avatar Descriptors.

During the meeting Transmitting Client of each participant sends to Server for distribution to all participants:

1. Speech.
2. Avatar Descriptors.



Figure 3 – Reference Model of Avatar Videoconference Transmitting Client

### Input and output data

*Table 2* gives the input and output data of Transmitting Client AIW:

*Table 2 – Input and output data of Client Transmitting AIW*

|  |  |
| --- | --- |
| **Input** | **Comments** |
| Text | Chat text used to communicate with Virtual Secretary or other participants |
| Language Preference | The language participant wishes to speak and hear at the videoconference. |
| Input Audio | Audio of participant’s Speech and Environment Audio. |
| Input Video | Video of participants’ upper part of the body. |
| Avatar Model | The avatar model selected by the participant. |
| **Output** | **Comments** |
| Language Preference | As in input. |
| Speaker Descriptors (A) | Participants’ Speech Descriptors for Authentication. |
| Participant’s Speech | Speech as separated from Environment Audio. |
| Avatar Descriptors | Descriptors produced by Transmitting Client. |
| Face Descriptors (A) | Participant’s Face Descriptors for Authentication |

### AI Modules

*Table 3* gives the AI Modules of Transmitting Client AIW.

*Table 3 – AI Modules of Transmitting Client AIW*

|  |  |
| --- | --- |
| **AIM** | **Function** |
| Audio Scene Description | Receives Input Audio.  Provides Speech. |
| Visual Scene Description | Receives Input Video  Provides Face Objects and Human Objects. |
| Speech Description | Receives Speech.  Provides Speech Descriptors for Authentication. |
| Speech Recognition | Receives Speech.  Provides Recognised Text. |
| Face Description1 | Receives Face Objects.  Provides the Face channel’s Status. |
| Gesture Description | Receives Human Objects.  Provides the Gesture channel’s Status. |
| Face Description2 | Provides Face Descriptors for Authentication. |
| Personal Status Extraction | Receives Recognised Text, Speech, Text, Face Object, Human Objects.  Provides the Personal Status |
| Language Understanding | Receives Recognised Text and Personal Status.  Provides Meaning. |
| Avatar Description | Receives Meaning, Personal Status, Face Description, and Gesture Description.  Produces the full set of Avatar Descriptions. |

## Server

### Function

The function of the Server is:

1. At the start:
   1. To Authenticate Participants.
   2. To distribute the Environment Model.
   3. To distribute participants’ Avatar Models.
2. During the videoconference
   1. To Translate and send speech to participants according to their preferences.
   2. To Forward Avatars Descriptors to all participants.

### Architecture

The Server:

1. Receives from:
   1. Server manager:
      1. Selected Environment.
      2. Participants’ Identities.
   2. Each Participant:
      1. Speech Descriptors (A).
      2. Face Descriptors (A).
      3. Language Preferences.
      4. Speech.
      5. Avatar Descriptors.
2. Authenticates Participants.
3. Translates Participants’ speech according to preferences.
4. Sends:
   1. Environment Descriptors.
   2. Participants’ IDs.
   3. Participants’ speech and translated participants’ speech.
   4. Participants’ Avatar Descriptors.

Figure 5 gives the architecture of Server AIW.

Graphical user interface, diagram

Description automatically generated

Figure 4 – Reference Model of Avatar-Based Videoconference Server

### I/O data

*Table 4* gives the input and output data of Server AIW.

*Table 4 – Input and output data of Server AIW*

|  |  |
| --- | --- |
| **Input** | **Comments** |
| Environment Selection | Set by Conference Manager |
| Participant Identities (xN) | Assigned by Conference Manager |
| Speech Descriptors (A) (xN) | Participant’s Speech Descriptors for Authentication |
| Face Descriptors (A) (xN) | Participant’s Face Descriptors for Authentication |
| Selected Language (xN) | From all participants |
| Speech (xN) | From all participants |
| Avatar Model (xN) | From all participants |
| Avatar Descriptors (xN) | From all participants |
| **Outputs** | **Comments** |
| Environment Model (xN) | Static Environment Descriptors |
| Participant ID (xN) | As in input |
| Speech (xN) | As in input |
| Avatar Model (xN) | As in input |
| Avatar Descriptors (xN) | As in input |

### AI Modules

*Table 5* gives the AI Modules of Server AIW.

*Table 5 – AI Modules of Server AIW*

|  |  |
| --- | --- |
| **AIM** | **Function** |
| Environment Description | Creates all static Environment Descriptors. |
| Participant Authentication | Authenticates Participants using their Speech and Face Descriptors |
| Translation | For all participants   1. Selects an active speaker. 2. Translates the Speech of that speaker to the set of translated Speech in the Selected Languages. 3. Assigns a translated Speech to the appropriate set of Partici­pants. |

## Virtual Secretary

### Architecture

The function of the Virtual Secretary is to:

1. Listen to the Speech of each avatar.
2. Monitor their Personal Status.
3. Draft a Summary using emojis and text in the meeting’s common language (decided outside).

The Summary can be handled in two different ways:

1. Transferred to an external application so that participants can edit the Summary.
2. Displayed to avatars:
   1. Avatars make Speech comments or Text comments (offline, i.e., via chat).
   2. The Virtual secretary edits the Summary interpreting Speech, Text, and the avatars’ Personal Status.

Figure 5 depicts the architecture of the Virtual Secretary AIW.

Chart

Description automatically generated

Figure 5 – Reference Model of Virtual Secretary

The Virtual Secretary workflow operates as follows (note that the Virtual Secretary processes one avatar at a time):

1. Speech Recognition extracts Text from an avatar speech.
2. Personal Status Extraction extracts Personal Status from Recognised Text, Speech, Face Object, and Human Object.
3. Language Understanding:
   1. Receives Personal Status and Recognised Text.
   2. Creates
      1. The final version of recognised text (Language Understanding-Text).
      2. Meaning of the sentence uttered by an avatar.
4. Summarisation
   1. Receives:
      1. Language Understanding -Text.
      2. Personal Status.
      3. Meaning.
   2. Creates Summary using emojis and text in the meeting’s common language.
   3. Receives Edited Summary from Question and Dialogue Processing.
5. Question and Dialogue Processing
   1. Receives
      1. Language Understanding -Text.
      2. Text from an avatar via chat.
      3. Meaning.
      4. Summary.
   2. Creates edited Summary.
   3. Sends edited Summary back to Summarisation.
   4. Outputs Text and Personal Status.
6. Speech Synthesis (PS)
   1. Receives Output Text and PS (Speech).
   2. Produces Output Speech with conveyed Personal Status.
7. Personal Status Display generates a manifestation of the machine as an avatar.

### I/O Data

*Table 6 – I/O data of Virtual Secretary*

|  |  |  |
| --- | --- | --- |
| **Input data** | **From** | **Comment** |
| Text (xN) | Avatars | Remarks on the summary, etc. |
| Speech (xN) | Avatars | Utterances by avatars |
| Face (xN) | Avatars | Faces of avatars |
| Gesture (xN) | Avatars | Gestures of avatars |
| **Output data** | **To** | **Comments** |
| Machine Speech | Avatars | Speech to avatars |
| Machine Face | Avatars | Face to avatars |
| Machine Gesture | Avatars | Gesture to avatars |
| Machine Text | Avatars | Response to chat. |
| Summary | Avatars | Summary of avatars’ interventions |

### AI Modules

*Table 7* gives the AI Modules of the Virtual Secretary depicted in Figure 5.

*Table 7 – AI Modules of Virtual Secretary*

|  |  |
| --- | --- |
| **AIM** | **Function** |
| Speech Recognition | 1. Receives Speech. 2. Produces Recognised Text. |
| Personal Status Extraction | 1. Receives PS-Speech Descriptors, PS-Face Descriptors, PS-Gesture Description. 2. Produces Personal Status. |
| Language understanding | 1. Receives Recognised Text and Personal Status. 2. Produces Language Understanding -Text and Meaning. |
| Summarisation | 1. Receives Meaning and Language Understanding -Text, and edited Summary 2. Produces Summary. |
| Question & Dialogue Processing | 1. Receives Language Understanding -Text, Personal Status, Intention, Meaning. 2. Produces PS (Speech), PS (Face), PS (Gesture), and Output Text, and edited Summary. |
| Personal Status Display | 1. Receives Output Text, PS (Speech), PS (Face), and PS (Gesture). 2. Produces Machine-Text, Machine-Speech, Machine-Face, Machine-Gesture |

## Receiving Client

### Function

The Function of the Receiving Client is to:

1. Create the Environment Using Environment Descriptors.
2. Place and animate the Avatar Models at positions selected by the participant.
3. Add the relevant Speech to each Avatar.
4. Present the Audio-Visual Scene as seen from the Point of View selected by the participant.

### Architecture

The Receiving Client:

1. Creates the AV Scene using:
   1. The Environment Model and Descriptors.
   2. The Avatar Models and Descriptors.
   3. The Speech of each Avatar.
2. Presents the Audio-Visual Scene based on the selected viewpoint in the Environment.

Figure 6 gives the architecture of Client Receiving AIW.

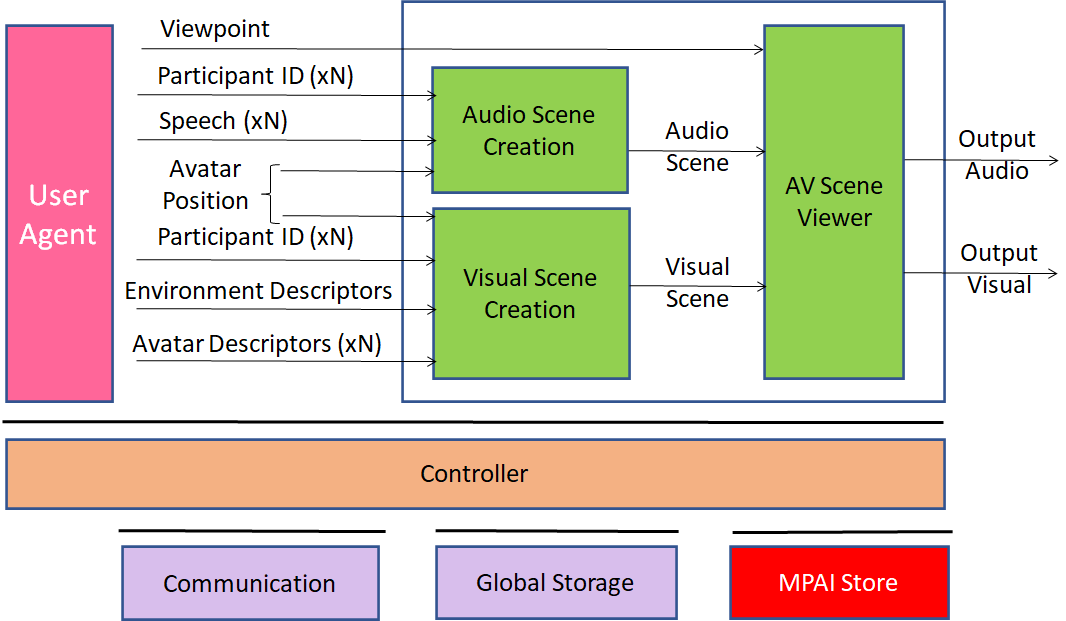


Figure 6 – Reference Model of Avatar-Based Videoconference Receiving Client

### I/O Data

*Table 8* gives the input and output data of Receiving Client AIW.

*Table 8 – Input and output data of Receiving Client AIW*

|  |  |
| --- | --- |
| **Input** | **Comments** |
| Viewpoint | Participant-selected point to see visual objects and hear audio objects in the Virtual Environment. |
| Avatar Position | Coordinates of Avatars in the Environment |
| Participants’ IDs (xN) | Unique Participants’ IDs |
| Speech (xN) | Possibly translated Participant’s Speech. |
| Environment Descriptors | Static Descriptors of Environment. |
| Avatar Descriptors (xN) | Descriptors of Avatar’s body animation. |
| **Output** | **Comments** |
| Output Audio | Presented using loudspeaker (array). |
| Output Visual | Presented using 2D or 3D display. |

### AI Modules

*Table 9* gives the AI Modules of Receiving Client AIW.

*Table 9 – AI Modules of Client-Based Environment*

|  |  |
| --- | --- |
| **AIM** | **Function** |
| Audio Scene Creation | Creates Audio Scene by combining the speech of speaking Avatars at the respective locations of the Visual Scene. |
| Visual Scene Creation | Creates Visual Scene composed of static Visual Scene Descriptors and Avatars. |
| AV Scene Viewer | Displays Participant’s Audio-Visual Scene from selected Viewpoint. |

# Composite AI Modules

Several Use Cases of different MPAI Specifications find it useful to define and use combinations of AI Modlues called Composite AI Modules in different Use Cases. This chapter specifies them using a format similar to the one adopted for Uses Cases.

## Personal Status Extraction

### Function of Composite AIM

Personal Status Extraction (PSE) is a composite AIM that analyses the Manifestation of a Personal Status conveyed by Text, Speech, Face, and Gesture – of a human or an avatar – and provides the Personal Status estimate. It is used in the Use Case-related figures of this document as a replacement for the combination of the AIMs depicted in Figure 7.

### Reference architecture

Personal Status Extraction produces the estimate of the Personal Status of a human or an avatar by analysing each Modality in 3 steps:

1. *Data Capture* (e.g., characters and words, a digitised speech segment, the digital video containing the hand of a person, etc.).
2. *Descriptor Extraction* (e.g., pitch and intonation of the speech segment, thumb of the hand raised, the right eye winking, etc.).
3. *Personal Status Interpretation* (i.e., one, two, or all three of Emotion, Cognitive State and Attitude).

Figure 7 depicts the Personal Status estimation process:

1. Descriptors are extracted from Text, Speech, Face Object, and Human Object.
2. Descriptors are interpreted and the specific Manifestations of the Personal Status in the Text, Speech, Face, and Gesture channels derived.
3. Personal Status is obtained by fusing the different estimations of the Personal Status.

Diagram

Description automatically generated

Figure 7 – Reference Model of Personal Status Extraction

An implementation can combine, e.g., the PS-Gesture Description and PS-Gesture Interpretation AIMs into one AIM, and directly provide PS-Gesture from a Human Object without exposing PS-Gesture Descriptors.

### Input/output data

*Table 10* gives the input/output data of Conversation with Emotion.

*Table 10 – I/O data of Personal Status Extraction*

|  |  |  |
| --- | --- | --- |
| **Input data** | **From** | **Comment** |
| Text | Keyboard or Speech Recognition | Text or recognised speech. |
| Speech | Microphone | Speech of human. |
| Face Object | Visual Scene Description | The face of the human. |
| Human Object | Visual Scene Description | The upper part of the body. |
| **Output data** | **To** | **Comments** |
| Personal Status | An AIM | For further processing |

### AI Modules

*Table 11* gives the list of the AIMs with their functions.

*Table 11 – AI Modules of Personal Status Extraction*

|  |  |
| --- | --- |
| **AIM** | **Function** |
| PS-Text Description | Receives Text.  Produces PS-Text Descriptors. |
| PS-Speech Description | Receives Speech.  Produces PS-Speech Descriptors. |
| PS-Face Description | Receives Face.  Produces PS-Face Descriptors. |
| PS-Gesture Description | Receives Gesture.  Produces PS- Gesture Descriptors. |
| PS-Text Interpretation | Receives PS-Text Descriptors.  Produces PS-Text. |
| PS-Speech Interpretation | Receives PS-Speech Descriptors.  Produces PS-Speech. |
| PS-Face Interpretation | Receives PS-Face Descriptors.  Produces PS-Face. |
| PS-Gesture Interpretation | Receives PS-Gesture Descriptors.  Produces PS-Gesture. |
| Personal Status Fusion | Receives PS-Text, PS-Speech, PS-Face, PS-Gesture.  Produces Personal Status. |

## Personal Status Display

### Function of Composite AIM

A Personal Status Display (PSD) is a Composite AIM receiving Text and Personal Status and generating an avatar producing Text and uttering Speech with the intended Personal Status while the avatar’s Face and Gesture show the intended Personal Status. The Personal Status driving the avatar can be extracted from a human or can be synthetically generated by a machine as a result of its conversation with a human or another avatar. It is used in the Use Case-related figures of this document as a replacement for the combination of the AIMs depicted in Figure 8.

### Reference Architecture

Figure 8 represents the AIMs required to implement Personal Status Display.

Diagram

Description automatically generated

Figure 8 – Reference Model of Personal Status Display

The Personal Status Display operates as follows:

1. Text is synthesised as Machine Speech using the Personal Status provided by PS (Speech). Text is also passed as output.
2. Machine Speech and PS (Face) are used to produce the animated Face.
3. PS (Gesture) is used to synthesise the animated Gesture.

### Input/output data

*Table 12* gives the input/output data of Personal Status Display.

*Table 12 – I/O data of Personal Status Display*

|  |  |  |
| --- | --- | --- |
| **Input data** | **From** | **Comment** |
| Text | Keyboard, Speech Recognition, Machine |  |
| PS (Speech) | Personal Status Extractor or Machine |  |
| PS (Face) | Personal Status Extractor or Machine |  |
| PS (Gesture) | Personal Status Extractor or Machine |  |
| **Output data** | **To** | **Comments** |
| Machine Text | Human or Avatar |  |
| Machine Speech | Human or Avatar |  |
| Machine Face | Human or Avatar |  |
| Machine Gesture | Human or Avatar |  |

### AI Modules

*Table 13* gives the list of AIMs with their functions.

*Table 13 – AI Modules of Personal Status Extraction*

|  |  |
| --- | --- |
| **AIM** | **Function** |
| Speech Synthesis (PS) | Receives Text and PS-Speech.  Produces Machine Speech. |
| Face Synthesis (PS) | Receives Machine Speech and PS-Face.  Produces Machine Face. |
| Gesture Synthesis (PS) | Receives Text and Machine PS-Gesture.  Produces Machine-Gesture. |

# Data Formats

## Audio Scene Descriptors

Specified in MPAI-CAE V2

## Cognitive State

Specified in MPAI-MMC V2

## Digital Human Descriptos

## Emotion

Specified in MPAI-MMC V2

## Face Descriptors for Motion

## Face Descriptors for Recognition

## Face Descriptors for Personal Status

## Gesture Descriptors for Personal Status

## Language identifier

Specified in MPAI-MMC V2

## Meaning

Specified in MPAI-MMC V2

## Personal Status

Specified in MPAI-MMC V2

## Social Attitude

Specified in MPAI-MMC V2

## Spatial Attitude

## Speech Descriptors for Recognition

## Text

Specified in MPAI-MMC V2

## User ID

## Visual Scene Descriptors

1. General MPAI Terminology

The Terms used in this standard whose first letter is capital and are not already included in *Table 1* are defined in *Table 14.*

*Table 14 – 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. |
| 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 its Function. |
| * Composite AIM | An AIM aggregating more than one AIM. |
| 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. |
| 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. |
| 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. |
| * Tester | An entity authorised by MPAI to Test the Conformance of an Implem­entation. |
| * 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. |
| * Format | The standard digital representation of 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. |
| 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 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 Implementation | The technically correct software implementation of a Technical Specific­ation attached to a Reference Software Specification. |
| 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. |
| * *Technical* | (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. |
| * *Reference Software* | The normative document specifying the use of the Reference Software Implementation. |
| * *Conformance Testing* | The normative document specifying the Means to Test the Conformance of an Implem­entation. |
| * *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. |
| Standard | The ensemble of Technical Specification, Reference Software, Confor­man­ce Testing and Performance Assessment of an MPAI application Standard. |
| Storage |  |
| * Storage | A Component to store data shared by the AIMs. |
| * Storage | A Component to store data of the individual AIMs. |
| 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. |
| * 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. |

1. 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](mailto:secretariat@mpai.community). 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](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](mailto:secretariat@mpai.community).

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.

1. 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.

A Level 1 Implementation shall be an Implementation of the MPAI-AIF Technical Specification executing AIWs composed of AIMs able to call the MPAI-AIF APIs.

|  |  |
| --- | --- |
| 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 | * 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 9* is a high-level description of the MPAI ecosystem operation applicable to fully conforming MPAI implementations as specified in the Governance of the MPAI Ecosystem Specification [**Error! Reference source not found.**]:

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 and MPAI Store.
6. Implementers submit Implementations to the MPAI Store
7. MPAI Store verifies security and Tests Confor­mance of Implementation.
8. Users download Implementations and report their experience to MPAI.

Diagram

Description automatically generated

*Figure 9 – The MPAI ecosystem operation*

1. At the time of publication of this Technical Report, the MPAI Store was assigned as the IIDRA. [↑](#footnote-ref-1)