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

MPAI Technical Specification

Portable Avatar Format (MPAI-PAF)

V1.3

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 Notices and Disclaimers.

Technical Specification
Portable Avatar Format (MPAI-PAF)
V1.3

Contents

1   Foreword

The international, unaffiliated, non-profit ***Moving Picture, Audio, and Data Coding by Artificial Intelligence (MPAI)*** organisation was established in September 2020 in the context of:
1. **Increasing** use of Artificial Intelligence (AI) technologies applied to a broad range of domains affecting millions of people
2. **Marginal** reliance on standards in the development of those AI applications
3. **Unprecedented** impact exerted by standards on the digital media industry affecting billions of people

believing that AI-based data coding standards will have a similar positive impact on the Information and Communication Technology industry.

The design principles of the MPAI organisation as established by the MPAI Statutes are the development of AI-based Data Coding standards in pursuit of the following policies:
1. Publish upfront clear Intellectual Property Rights licensing frameworks.
2. Adhere to a rigorous standard development process.
3. Be friendly to the AI context but, to the extent possible, remain agnostic to the technology thus allowing developers freedom in the selection of the more appropriate – AI or Data Processing – technologies for their needs.
4. Be attractive to different industries, end users, and regulators.
5. Address five standardisation areas:

1. *Data Type*, a particular type of Data, e.g., Audio, Visual, Object, Scenes, and Descriptors with as clear semantics as possible.
2. *Qualifier*, specialised Metadata conveying information on Sub-Types, Formats, and Attributes of a Data Type.
3. *AI Module* (AIM), processing elements with identified functions and input/output Data Types.
4. *AI Workflow* (AIW), MPAI-specified configurations of AIMs with identified functions and input/output Data Types.
5. *AI Framework* (AIF), an environment enabling dynamic configuration, initialisation, execution, and control of AIWs.

6. Provide appropriate Governance of the ecosystem created by MPAI Technical Specifications enabling users to:
   1. *Operate* Reference Software Implementations of MPAI Technical Specifications provided together with Reference Software Specifications
   2. *Test* the conformance of an implementation with a Technical Specification using the Conformance Testing Specification.
   3. *Assess* the performance of an implementation of a Technical Specification using the Performance Assessment Specification.
   4. *Obtain* conforming implementations possibly with a performance assessment report from a trusted source through the MPAI Store.

Today, the MPAI organisation operated on four solid pillars:
1. The MPAI Patent Policy specifies the MPAI standard development process and the Framework Licence development guidelines.
2. *Technical Specification: Artificial Intelligence Framework (MPAI-AIF)* V2.1 specifies an environment enabling initialisation, dynamic configuration, and control of AIWs in the standard AI Framework environment depicted in Figure 1. An AI Framework can execute AI applications called AI Workflows (AIW) typically including interconnected AI Modules (AIM). MPAI-AIF supports small- and large-scale high-performance components and promotes solutions with improved explainability.

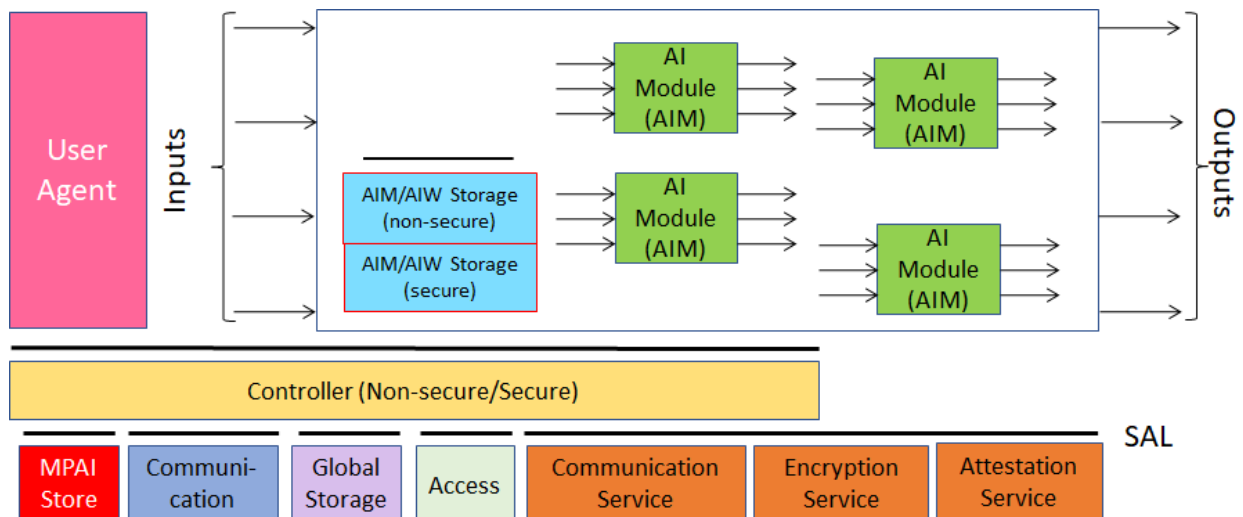*Figure 1 – The AI Framework (MPAI-AIF) V2 Reference Model*

3. *Technical Specification: Data Types, Formats, and Attributes (MPAI-TFA) V1.2* specifies Qualifiers, a type of metadata supporting the operation of AIMs receiving data from other AIMs. Qualifiers convey information on Sub-Types (e.g., the type of colour), Formats (e.g., the type of compression and transport), and Attributes (e.g.,

semantic information in the Content). Although Qualifiers are human-readable, they are only intended to be used by AIMs. Therefore, Text, Speech, Audio, Visual, and other Data exchanged by AIWs and AIMs should be interpreted as being composed of Content (Text, Speech, Audio, and Visual as appropriate) and associated Qualifiers. Therefore a Text Object is composed of Text Data and Text Qualifier. The specification of most MPAI Data Types reflects this point.

4. *Technical Specification: Governance of the MPAI Ecosystem (MPAI-GME) V1.1* defines the following elements:
    1. Standards, i.e., the ensemble of Technical Specifications, Reference Software, Conformance Testing, and Performance Assessment.
    2. Developers of MPAI-specified AIMs and Integrators of MPAI-specified AIWS (Implementers).
    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.

The interaction between and among actors of the MPAI Ecosystem are depicted in Figure 2.
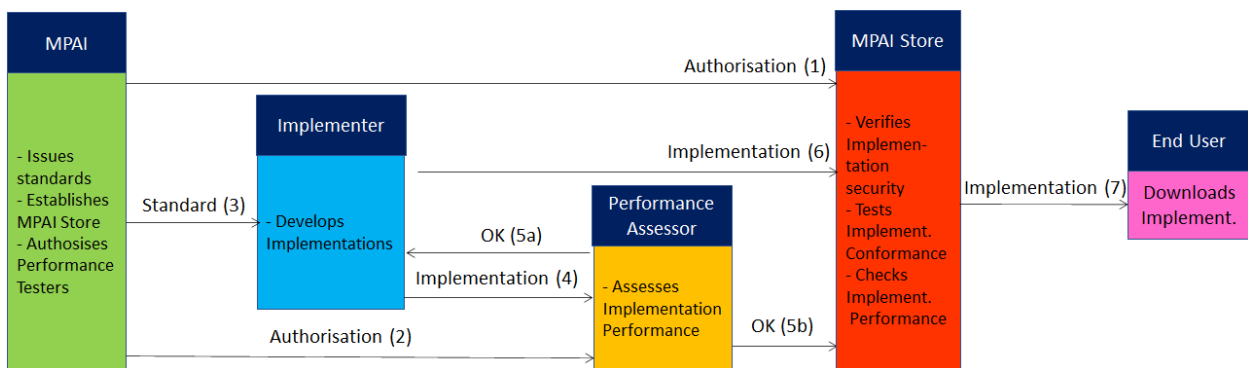

Figure 2 – The MPAI Ecosystem

2   Introduction (Informative)

There is a long history of computer-created objects called "digital humans", i.e., digital objects having a human appearance when rendered. In most cases the underlying assumption of these objects has been that creation, animation, and rendering is done in a closed environment. Such digital humans had little or no need for standards.

In a communication and more so in a metaverse context, there are many cases where a digital human is not constrained within a closed environment thus requiring forms of standardisation. *Technical Specification: Portable Avatar Format* (**MPAI-PAF**) **V1.3** – in the following also called MPAI-PAF V1.3 or MPAI-PAF – is a response to the requirements of new usage contexts. MPAI-PAF specifies a standard for Portable Avatar Format (PAF) enabling a receiving party to render a digital human as intended by the sending party.

MPAI-PAF V1.2 specifies the Avatar-Based Videoconference (PAF-ABV) AI Workflow where:
    1. **Client Transmitters** send PAFs containing:
        o At the beginning: Avatar Models, Language Selector, and Speech Object and Face Object for participant authentication.
        o Continuously: Avatar Descriptors, and Speech Objects to a Server.
    2. **Avatar Videoconference Server**:

- o At the beginning:
    - Selects an Environment, i.e., a meeting room and equips it with objects, i.e., meeting table and chairs.
    - Places Avatar Models around the table.
    - Distributes for each participant a PAF containing Environment, Avatar Models, and their positions to all receiving clients.
  - o Continuously sends to receiving clients:
    - Translated Speech from participants according to Language Selectors.
    - Sends PAFs containing Avatar Descriptors and translated Speech.
3. **Client Receivers**:
  - o At the beginning: receive Environment and PAFs containing Avatar Models and Language Selectors from the server.
  - o Continuously from the server:
    - Receive PAFs containing Avatar Descriptors and translated Speech.
    - Create Audio and Visual Scene Descriptors.
    - Render the Audio-Visual Scene as seen from the human participant-selected Point of View.

In all Chapters and Sections, Terms beginning with a capital letter are defined in *Table 1* if they are specific to this Technical Specification and in *Table 2* if they are common to all MPAI Technical Specifications. All Chapters, and Sections are Normative unless they are labelled as Informative.

3 Scope

Technical Specification: Portable Avatar Format (MPAI-PAF) V1.3 specifies:

1. The *Portable Avatar Format* and related *Data Types* enabling a receiver to decode and render an Avatar and its Environment as intended by the sender.
2. The *Personal Status Display* Composite AI Module allowing the conversion of a Text and a Personal Status to a Portable Avatar.
3. The *Audio-Visual Scene Rendering* Composite AI Module rendering Audio-Visual Scene Descriptors and/or Portable Avatars as Audio and Visual information.
4. The AI Framework (MPAI-AIF)-conforming *AI Workflows and AI Modules* of the Avatar-Based Videoconference Use Case also using Data Types and AI Modules from other MPAI Technical Specifications.

The Use Case normatively defines:

1. The <u>Functions</u> of the AIWs and of the AIMs.
2. The <u>Connections</u> between and among the AIMs
3. The <u>Semantics</u> and the <u>Formats</u> 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 function specified in the relevant Section of AI Workflows.
    2. Use AIMs connected with the topology and connections conforming with the AI Workflows specifications.
    3. Have input and output data with the formats specified by the relevant AI Workflows specifications.
2. An *AIM*, shall:
    1. Perform the AIM function specified in the relevant Section of AI Modules and AI Modules from other Technical Specifications.
    2. Be composed of AIMs specified in the relevant Section of AI Modules and AI Modules from other Technical Specifications in case the AIM is Composite.

      3. Receive and produce the data specified in the relevant Section of Data Types and Data Types from other Technical Specifications

Users of this Technical Specification should note that:

1. Implementers may use the Reference Software of this Technical Specification to develop their Implementations.
2. The Conformance Testing specification can be used to test the conformity of an Implementation to this Standard.
3. Performance Assessors can assess the level of Performance of an Implementation based on the Performance Assessment associated with the MPAI-PAF Technical Specification.
4. Implementers and Users should consider the notices and disclaimers.

This Technical Specification includes the following elements:

1. Scope (This Chapter)
2. Definitions
3. References
4. AI Workflows for Avatar-Base Videoconference
5. AI Modules
6. Data Types.

The current version of the Technical Specification has been developed by the Portable Avatar Format Development Committee (PAF-DC). MPAI may issue new versions of MPAI-PAF that extend or replace the scope of the current Technical Specification.

4 Definitions

Capitalised Terms have the meaning defined in **_Table 1_**. Terms applicable to all MPAI Technical Specifications are defined in _Table 2_. Non-capitalised terms letter have the meaning commonly defined for the context in which they are used or represent an entity in the real world. For instance,

1. Table 1 defines _Object, Scene,_ and _User_ but does not define _object, scene,_ and _human_.
2. Object indicates an Item but object indicates an entity in the Universe commonly classified as object.

A dash "-" preceding a Term in Table 1 means the following:

1. If the font is normal, the Term in Table 1 without a dash and preceding the one with a dash should be placed <u>before</u> that Term. The notation is used to concentrate in one place all the Terms that are composed of, e.g., the word Data <u>followed</u> by one of the words Format and Type.
2. If the font is _italic,_ the Term in the table without a dash and preceding the one with a dash should be placed <u>after</u> that Term. The notation is used to concentrate in one place all the Terms that are composed of, e.g., the word Descriptor <u>preceded</u> by one of the words Face and Body.

Table 1 – Terms and Definitions

Term	Definition
Attitude	
– _Social_	A Factor of the Personal 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 Digital 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 device or a `human is what it states it is.
Avatar	A rendered Digital Human.
– Model	An inanimate Avatar exposing animation interfaces.
– *Portable*	A Data Type including Avatar ID, Time, Audio-Visual Scene Descriptors, Spatial Attitude, Avatar Model, Body Descriptors, Face Descriptors, Language Preference, Speech Coding, Speech Data, Text, and Personal Status.
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	An element of the internal status reflecting the way a human or avatar understands the Environment, such as "Confused", "Dubious", "Convinced".
Context	Additional information about a communication emitted by an Entity, such as language, culture etc.
Data	Information in digital form.
– Format	The syntax and semantics of a Data Type.
– Type	A particular type of Data.
Descriptor	The Digital Representation of a feature of an Object.
– *Body*	A Data Type including the digital representation of the features of the body of a real or digital human.
– *Face*	A Data Type including the digital representation of a feature of the face of a real or digital human.
Device	A piece of equipment used to interact and have Experience in a Digital Environment.
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 real or Digital Human
Environment	A Virtual Space that may be null or may include an Audio-Visual Scene.
Experience	The state of a human whose senses are continuously affected for a meaningful period.
Face	A digital representation of a human face.
Factor	One of Emotion, Cognitive State, and Spatial 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.
Grade	The intensity of a Factor.
Human	

Digital	A Digitised or a Virtual Human in a Virtual Space.
Digitised	An Object in a Virtual Space that has the appearance of a specific human when rendered.
Virtual	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.
Language Preference	The Language(s) acceptable by a participant.
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 other Audio Objects.
– *Audio-Visual*	Coded representation of Audio-Visual information with its metadata.
– Descriptor	The digital representation of the feature of an Object.
– *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 Visual Object can include other Visual Objects.
– *Virtual*	An Object not representing an object in a Real Environment.
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.
Point of View	The Spatial Attitude of a Digital Human watching an Environment.
Portable Avatar	A Data Type representing an Avatar and its Context.
Position	The coordinates of a representative point for an object in a Virtual Space with respect to a set of Coordinate Axes.
Rendering	The process of instantiating a Virtual Space as a human-perceptible entity.
Scene	A Digital Environment populated by Objects.
– *Audio*	The Audio Objects of an Environment with Object metadata such as Spatial Attitude.
– *Audio-Visual*	(AV Scene) The Audio-Visual Objects of an Environment Object metadata such as Spatial Attitude.
– *Visual*	The Visual Objects of an Environment with Object metadata such as Spatial Attitude.
Scene Descriptors	The digital representation of a feature of a scene.
– *Audio*	A Data Type including the digital representation of the audio features of a real or digital scene.
– *Audio-Visual*	A Data Type combining the Audio or Visual Scene Descriptors.
– *Visual*	A Data Type including the digital representation of the visual features of a real or digital scene.
Representation	Data that digitally represents an entity of a real environment.

Scene Geometry	The digital representation of the object arrangement of a scene.
– *Audio*	A Data Type describing the spatial arrangement of the Visual Objects of a Scene.
– *Audio-Visual*	A Data Type describing the spatial arrangement of the Audio, Visual, and Audio-Visual Objects of a Scene.
– *Visual*	A Data Type describing the spatial arrangement of the Visual Objects of a Scene.
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].
Virtual Space	A space generated and maintained by a computing platform that can be rendered.

The Terms used in this standard whose first letter is capital and are not already included in Table 1 are defined in Table 2. To concentrate in one place all the Terms that are composed of a common name followed by other words (e.g., the word Data <u>followed</u> by one of the words Format, Type, or Semantics), the definition given to a Terms preceded by a dash "-" applies to a Term composed by that Term without the dash preceded by the Term that precedes it in the column without a dash.

Table 2 – 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 Model (AIM)	A data processing element receiving AIM-specific Inputs and producing AIM-specific Outputs according to according to its Function. An AIM may be an aggregation of AIMs.
AI Workflow (AIW)	A structured aggregation of AIMs implementing a Use Case receiving 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.
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 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.
– 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) 2. Pass the Conformance Testing (Level 2) of an Application Standard 3. Pass the Performance Testing (Level 3) of an Application Standard.	
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 theirs Connections in an AIW.	
Reference Software	A technically correct software implementation of a Technical Specification containing source code, or source and compiled code.	
Reliability	The attribute of an Implementation that performs as specified by the Application Standard, profile, and version the Implementation refers to, e.g., within the application scope, stated limitations, and for the period of time specified by the Implementer.	
Replicability	The attribute of an Implementation whose Performance, as Assessed by a Performance Assessor, can be replicated, within an agreed level, by another Performance Assessor.	
Robustness	The attribute of an Implementation that copes with data outside of the stated application scope with an estimated degree of confidence.	
Scope	The domain of applicability of an MPAI Application Standard	
Service Provider	An entrepreneur who offers an Implementation as a service (e.g., a recommendation service) to Users.	
Standard	A set of Technical Specification, Reference Software, Conformance Testing, Performance Assessment, and Technical Report of an MPAI application Standard.	
Technical Specification	(Framework) the normative specification of the AIF. (Application) the normative specification of the set of AIWs belonging to an application domain along with the AIMs required to Implement the AIWs that includes: 1. The formats of the Input/Output data of the AIWs implementing the AIWs.	

	2. The Connections of the AIMs of the AIW.
	3. The formats of the Input/Output data of the AIMs belonging to the AIW.
Testing Laboratory	A laboratory accredited to Assess the Grade of Performance of Implementations.
Time Base	The protocol specifying how Components can access timing information
Topology	The set of AIM Connections of an AIW.
Use Case	A particular instance of the Application domain target of an Application Standard.
User	A user of an Implementation.
User Agent	The Component interfacing the user with an AIF through the Controller
Version	A revision or extension of a Standard or of one of its elements.
Zero Trust	A cybersecurity model primarily focused on data and service protection that assumes no implicit trust.

5   References

5.1   Normative References

Technical Specification: Portable Avatar Format (MPAI-PAF) normatively references the following documents, both from MPAI and other standards organisations. Referenced MPAI standards are publicly available at the URL indicated in the reference.

1. MPAI; Technical Specification: Governance of the MPAI Ecosystem (MPAI-GME) V1.1.
2. MPAI; Technical Specification: Artificial Intelligence Framework (MPAI-AIF) V2.1.
3. MPAI; Technical Specification: Context-based Audio Enhancement (MPAI-CAE) V2.2.
4. MPAI; Technical Specification: Object and Scene Description (MPAI-OSD) V1.2.
5. MPAI; Technical Specification: AI Module Profiles (MPAI-PRF) V1.0.
6. MPAI; Technical Specification: Data Types, Formats, and Attributes; V1.2.
7. Khronos; Graphics Language Transmission Format (glTF); October 2021.
8. ISO/IEC 19774-1:2019 Information technology – Computer graphics, image processing and environmental data representation – Part 1: Humanoid animation (HAnim) architecture.
9. ISO/IEC 19774-2:2019 Information technology – Computer graphics, image processing and environmental data representation – Part 2: Humanoid animation (HAnim) motion data animation.

5.2   Informative References

These references are provided for information purposes.
10. MPAI; The MPAI Statutes.
11. MPAI; The MPAI Patent Policy.
12. MPAI; Technical Specification: Connected Autonomous Vehicles (MPAI-CAV) – Architecture (CAV-ARC) V1.1.
13. MPAI; Technical Specification: Connected Autonomous Vehicles (MPAI-CAV) – Technologies (CAV-TEC) V1.0.
14. MPAI; Technical Specification: Human and Machine Communication (MPAI-MMM) V1.1

15. MPAI; Technical Specification: MPAI Metaverse Model (MPAI-MMM) – Architecture (MMM-ARC) V1.1.
16. MPAI; Technical Specification: MPAI Metaverse Model (MPAI-MMM) – Technologies V1.1.

6 AI Workflows

6.1 Technical Specifications

Technical Specification: Portable Avatar Format (MPAI-PAF) V1.3, jointly with other MPAI Technical Specifications, provides technologies for the digital representation of 3D Model Data that enable the **Avatar-Based Videoconference**, a form of videoconference held in a Virtual Environment populated by speaking Avatars and implemented as an AI Workflow specified by **Technical Specification: AI Framework (MPAI-AIF) V2.1**.

Table 1 displays the full list of AIWs specified by MPAI-PAF V1.3. Click a listed AIW to access its dedicated page, which includes its functions, reference model, I/O Data, Functions of AIMs, I/O Data of AIMs, and a table providing links to the AIW-related AIW, AIMs, and JSON metadata. All previously specified MPAI-PAF AI-Workflows are superseded by those specified by V1.3 but may be used if their version is explicitly mentioned.

Acronym	Names and Specifications of AI Workflows	JSON
PAF-CTX	Videoconference Client Transmitter	X
MMC-VMS	Virtual Meeting Secretary	X
PAF-AVS	Avatar Videoconference Server	X
PAF-CRX	Videoconference Client Receiver	X

Figure 1 depicts the system composed of four types of subsystems specified as AI Workflows.
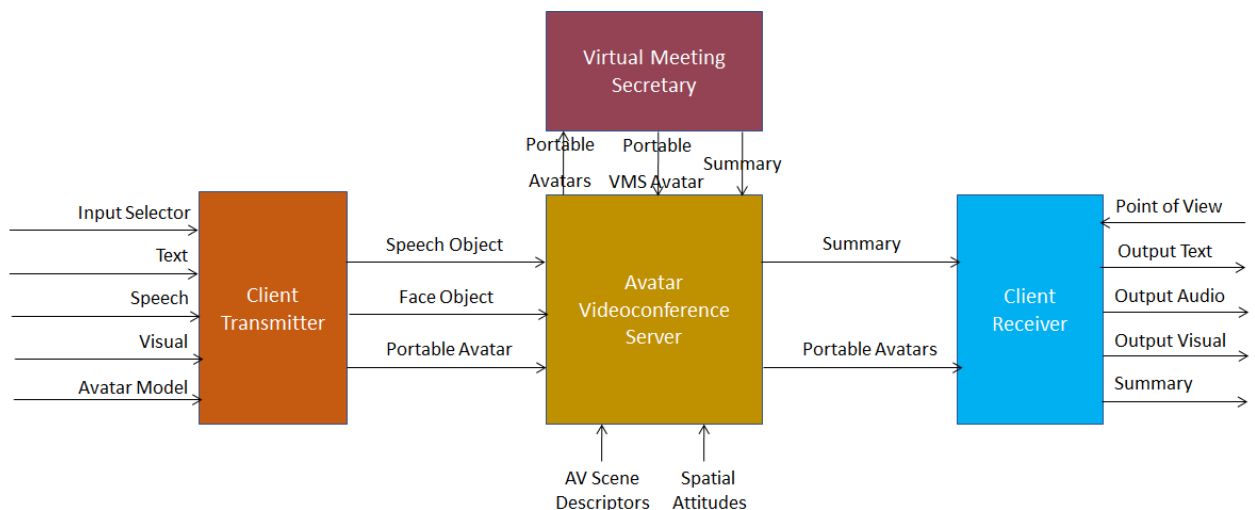

Figure 1 – Avatar-Based Videoconference end-to-end diagram

The components of the PAF-ABV system:
1. **Participant:** a human joining an ABV either individually or as a member of a group of humans in the same physical space.

2. **Audio-Visual Scene:** a Virtual Audio-Visual Environment equipped with Visual Objects such as a Table and an appropriate number of chairs and Audio Objects described by Audio-Visual Scene Descriptors.
3. **Portable Avatar:** a data set specified by MPAI-PAF including data representing a human participant.
4. **Videoconference Client Transmitter:**
   o <u>At the beginning</u> of the conference:
     ▪ Receives from Participants and sends to the Server Portable Avatars containing the Avatar Models and Language Selectors.
     ▪ Sends to the Server Speech Object and Face Object for Authentication.
   o <u>Continuously</u> sends to the Server Portable Avatars containing Avatar Descriptors and Speech.
5. The **Avatar Videoconference Server**
   o <u>At the beginning</u> of the conference:
     ▪ Selects the Audio-Visual Descriptors, e.g., a Meeting Room.
     ▪ Equips the Room with Objects, i.e., Table and Chairs.
     ▪ Places Avatar Models around the Table with a given Spatial Attitude.
     ▪ Distributes Portable Avatars containing Avatars Models, their Speech Objects and Spatial Attitudes, and Audio-Visual Scene Descriptors to all Receiving Clients.
     ▪ Authenticates Speech and Face Objects and assigns IDs to Avatars.
     ▪ Sets the common conference language.
   o <u>Continuously</u>:
     ▪ Translates Speech to Participants according to their Language Selectors.
     ▪ Sends Portable Avatars containing Avatar Descriptors, Speech, and Spatial Attitude of Participants and Virtual Meeting Secretary to all Receiving Clients and Virtual Meeting Secretary.
6. **Virtual Meeting Secretary** is an Avatar not corresponding to any Participant that <u>continuously</u>:
   o Uses the common meeting language.
   o Understands Text Objects and Speech Objects of all Avatars and extracts their Personal Statuses.
   o Drafts a Summary of its understanding of Avatars' Text Objects, Speech Objects, and Personal Status.
   o Displays the Summary either to:
     ▪ Outside of the Virtual Environment for participants to read and edit directly, or
     ▪ The Visual Space for Avatars to comment, e.g., via Text Objects.
   o Refines the Summary.
   o Sends its Portable Avatar containing its Avatar Descriptors to the Server.
7. **Videoconference Client Receiver**:
   o <u>At the beginning</u> of the conference:
     ▪ Receives Audio-Visual Scene Descriptors and Portable Avatars containing Avatar Models with their Spatial Attitudes.
   o <u>Continuously</u>:
     ▪ Receives Portable Avatars with Avatar Descriptors and Speech.
     ▪ Produces Visual Scene Descriptors and Audio Scene Descriptors.
     ▪ Renders the Audio-Visual Scene by spatially adding the Avatars' Speech Objects to the Spatial Attitude of the respective Avatars' Mouths. Rendering may be done from a Point of View, possibly different from the

Position assigned to their Avatars in the Visual Scene, selected by participant who use a device of their choice (Head Mounted Display or 2D display/earpad) to experience the Audio-Visual Scene.

Each component of the Avatar-Based Videoconference Use Case is implemented as an AI Workflow (AIW) composed of AI Modules (AIMs). Each AIW includes the following elements:

1	Functions of the AIW	The functions performed by the AIW implementing the Use Case.
2	Reference Model of the AIW	The Topology of AIMs in the AIW.
3	Input and Output Data of the AIW	Input and Output Data of the AIW.
4	Functions of the AIMs	Functions performed by the AIMs.
5	Input and Output Data of the AIW	Input and Output Data of the AIMs.
6	AIW, AIMs, and JSON Metadata	Links to summary specification on the web of the AIMs and corresponding JSON Metadata [2].

6.2 Reference Software

As a rule, MPAI provides Reference Software implementing the Technical Specification released with the BSD-3-Clause licence and the following disclaimers

1. The purpose of the Reference Software is to demonstrate a working Implementation of an AIW, not to provide a ready-to-use product.
2. MPAI disclaims the suitability of the Software for any other purposes than those of the MPAI-PAF Standard, and does not guarantee that it offers the best performance and that it is secure.
3. Users shall verify that they have the right to use any third-party software required by this Reference Software, e.g., by accepting the licences from third-party repositories.

Note that at this stage only part of the AIMs required to operate the MPAI-PAF AIWs have a Reference Software Implementation.

6.3 Conformance Testing

An implementation of an AI Workflow conforms with MPAI-PAF if it accepts as input and produces as output Data and/or Data Objects (Data of a Data Type and its Qualifier) conforming with those specified by MPAI-PAF.

The Conformance of an instance of a Data is to be expressed by a sentence like "Data validates against the Data Type Schema". This means that:

- Any Data Sub-Type is as indicated in the Qualifier.
- The Data Format is indicated by the Qualifier.
- Any File and/or Stream have the Formats indicated by the Qualifier.
- Any Attribute of the Data is of the type or validates against the Schema specified in the Qualifier.

The method to Test the Conformance of a Data or Data Object instance is specified in the *Data Types* chapter.

6.4 Performance Assessment

Performance is a multidimensional entity because it can have various connotations. Therefore, the Performance Assessment Specification should provide methods to measure how well an AIW performs its function, using a metric that depends on the nature of the function, such as:

1. Quality: the Performance of a Videoconference Client Transmitter AIW can measure how well the AIW represents the human Participant.

2. Bias: Performance of a Videoconference Client Receiver AIW can reproduce the avatar videoconference.
3. Legal compliance: the Performance of an AIW can measure the compliance of the AIW to a regulation, e.g., the European AI Act.
4. Ethical compliance: the Performance Assessment of an AIW can measure the compliance of an AIW to a target ethical standard.

Note that at this stage MPAI-PAF AIWs do not have a Performance Assessment Specification.

7 AI Modules

7.1 Technical Specifications

Table 1 provides the links to the specifications and the JSON syntax of all AIMs specified by **Technical Specification: Portable Avatar Format (MPAI-PAF) V1.3**. MPAI-PAF V1.3 AI-Modules supersede those previously specified which may still be used if their version is by explicitly signaled. AIMs in bold are Composite.

Table 1 - Specifications and JSON syntax of AIMs used by MPAI-PAF V1.3

Acronym	AIM Name	JSON	Acronym	AIM Name	JSON
PAF-AVC	Audio-Visual Scene Creation	X	PAF-PDX	Portable Avatar Demultiplexing	X
PAF-AVR	Audio-Visual Scene Rendering	X	PAF-PMX	Portable Avatar Multiplexing	X
PAF-EBD	Entity Body Description	X	PAF-PFI	PS-Face Interpretation	X
PAF-EFD	Entity Face Description	X	PAF-PGI	PS-Gesture Interpretation	X
PAF-FIR	Face Identity Recognition	X	**PAF-PSD**	**Personal Status Display**	**X**
PAF-FPS	**Face Personal Status Extraction**	**X**	PAF-SPA	Service Participant Authentication	X
PAF-GPS	**Gesture Personal Status Extraction**	**X**	PAF-VSC	Visual Scene Creation	X

7.2 Reference Software

As a rule, MPAI provides Reference Software implementing the AI Modules released with the BSD-3-Clause licence and the following disclaimers:
1. The purpose of the Reference Software is to provide a working Implementation of an AIM, not a ready-to-use product.
2. MPAI disclaims the suitability of the Software for any other purposes than those of the MPAI-PAF Standard, and does not guarantee that it offers the best performance and that it is secure.
3. Users shall verify that they have the right to use any third-party software required by this Reference Software, e.g., by accepting the licences from third-party repositories.

Note that at this stage only part of the MPAI-PAF AIMs have a Reference Software Implementation.

7.3 Conformance Testing

An implementation of an AI Module conforms with MPAI-PAF V1.3 if it accepts as input and produces as output Data and/or Data Objects (combination of Data of a Data Type and its Qualifier) conforming with those specified by MPAI-PAF V1.3.

The Conformance of an instance of a Data is to be expressed by a sentence like "Data validates against the Data Type Schema". This means that:

- Any Data Sub-Type is as indicated in the Qualifier.
- The Data Format is indicated by the Qualifier.
- Any File and/or Stream have the Formats indicated by the Qualifier.
- Any Attribute of the Data is of the type or validates against the Schema specified in the Qualifier.

The method to Test the Conformance of a Data or Data Object instance is specified in the *Data Types* chapter.

7.4 Performance Assessment

Performance is a multidimensional entity because it can have various connotations. Therefore, the Performance Assessment Specification should provide methods to measure how well an AIW performs its function, using a metric that depends on the nature of the function, such as:

1. *Quality*: Performance Assessment measures how well an AIM performs its function, using a metric that depends on the nature of the function, e.g., how well a Face Identity Recognition (FIR) AIM identifies Faces.
2. *Bias*: Performance Assessment measures the preference given by an AIM to certain elements, using a metric that depends on a bias related to certain attributes of the AIM. For instance, a Face Identity Recognition (FIR) AIM tends to have a higher correct identification of Face having a particular skin colour.
3. *Legal* compliance: Performance Assessment measures how well an AIM performs its function, using a metric that assesses its accordance with a certain legal standard.
4. *Ethical* compliance: the Performance Assessment of an AIM can measure the compliance of an AIM to a target ethical standard.

The current MPAI-PAF V1.3 does not provide AIM Performance Assessment methods.

8 Data Types

8.1 Technical Specifications

This page gives the links to the specification of Data Types of *Technical Specification: Portable Avatar Format (MPAI-PAF) V1.3*. All previously specified MPAI-PAF Data Types that are specified by V1.3 are superseded. Use of earlier versions is permitted if their version is explicitly signaled..

Table 1 provides the Data Types specified by MPAI-PAF V1.3. MPAI-PAF AIWs and AIMs also utilise Data Types specified by other MPAI Technical Specifications.

Table 1 - Data Types specified by MPAI-PAF V1.3

3D Model Object	Avatar	3D Model Basic Scene Descriptors	3D Model Basic Scene Geometry
3D Model Scene Descriptors	3D Model Scene Geometry	Body Descriptors	Face Descriptors
Gesture Descriptors	Portable Avatar		

8.2 Conformance testing

A Data instance of a Data Type Conforms with MPAI-PAF V1.3 if the JSON Data validate against the relevant MPAI-PAF V1.3 JSON Schema and if the Data Conforms with the relevant Data Qualifier, if present. MPAI-PAF V1.3 does not provide method for testing the Conformance of the Semantics of the Data instance to the MPAI-PAF V1.3 specification.

Conformance testing can be performed by a human using a JSON Validator to verify the Conformance of the syntax of JSON Data to the relevant JSON Schema; and, if the Data has a Qualifier, to verify that the syntax of the Data conforms with the relevant values in the Data Qualifier. Alternatively, Conformance testing can be performed by software implementing the steps above.

8.3 Performance Assessment

Performance Assessment provides methods of assessing the performance of an Data instance. Performance may have various connotations, such as:

1. *Quality*: Performance Assessment measures the quality of the Data instance using a metric that depends on the nature of the Data, e.g., the accuracy of reproduction of face emotions on a synthetic face.
2. *Bias*: Performance Assessment measures the disparity of treatment applied to the Data instance using a metric that depends on a bias related to certain attributes of the Data instance.
3. Legal compliance: Performance Assessment uses an appropriate metric to measure how well the Data instance complies with with a certain legal standard.