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

**Public document**

|  |  |
| --- | --- |
| **N1848** | 2024/06/12 |
| **Source** | Requirements (XRV) |
| **Title** | Use Cases and Functional Requirements: XR Venues (MPAI-XRV) – Collaborative Immersive Laboratory (XRV-CIL) WD0.4 |
| **Target** | MPAI-45 |

# Abstract

XR Venues (MPAI-XRV) is an MPAI project addressing a multiplicity of use cases enabled by Extended Reality (XR), the combination of Augmented Reality (AR), Virtual Reality (VR) and Mixed Reality (MR) technologies and enhanced by Artificial Intelligence (AI) technologies. The word Venue is used as a synonym for Real and Virtual Environments.

Nine Use Cases have been identified among which is the Collaborative Immersive Laboratory (XRV-CIL). This document addresses CRV-CIL Use Cases and Functional Requirements.

Contents

[Abstract 1](#_Toc157703664)

[1 Introduction 2](#_Toc157703665)

[2 Terms and definitions 3](#_Toc157703666)

[3 References 5](#_Toc157703667)

[4 Collaborative immersive laboratory 6](#_Toc157703668)

[4.1 Scope 6](#_Toc157703669)

[4.2 Use cases 6](#_Toc157703670)

[4.2.1 3D modelling of microscopic medical images 6](#_Toc157703671)

[4.2.2 Animation of medical images 7](#_Toc157703672)

[4.2.3 Modelling of astronomical space-time 7](#_Toc157703673)

[4.3 Description 7](#_Toc157703674)

[4.4 Specific application areas 7](#_Toc157703675)

[4.4.1 Microscopic dataset visualisation 7](#_Toc157703676)

[4.4.2 Macroscopic dataset visualisation and simulation 7](#_Toc157703677)

[4.4.3 Educational lab 8](#_Toc157703678)

[4.4.4 Collaborative CAD 8](#_Toc157703679)

[Annex 1 - Basics about MPAI 9](#_Toc157703680)

[1 General 9](#_Toc157703681)

[2 Governance of the MPAI Ecosystem 9](#_Toc157703682)

[3 AI Framework 10](#_Toc157703683)

[4 Personal Status 11](#_Toc157703684)

[4.1 General 11](#_Toc157703685)

[4.2 Personal Status Extraction 11](#_Toc157703686)

[Annex 2 - Other MPAI-XRV Use Cases 13](#_Toc157703687)

[5 eSports Tournament (XRV-EST). 13](#_Toc157703688)

[5.1 Purpose 13](#_Toc157703689)

[5.2 Description 13](#_Toc157703690)

[6 Experiential retail/shopping. 14](#_Toc157703691)

[6.1 Purpose 14](#_Toc157703692)

[6.2 Description and flow of actions 14](#_Toc157703693)

[7 Immersive art experience. 14](#_Toc157703694)

[7.1 Purpose 14](#_Toc157703695)

[7.2 Description 14](#_Toc157703696)

[8 DJ/VJ performance at a dance party. 15](#_Toc157703697)

[8.1 Purpose 15](#_Toc157703698)

[8.2 Description 15](#_Toc157703699)

[9 Live concert performance. 15](#_Toc157703700)

[9.1 Purpose 15](#_Toc157703701)

[9.2 Description 15](#_Toc157703702)

[10 Experiential marketing/branding. 16](#_Toc157703703)

[10.1 Purpose 16](#_Toc157703704)

[10.2 Description 16](#_Toc157703705)

[11 Meetings/presentations. 16](#_Toc157703706)

[11.1 Purpose 16](#_Toc157703707)

[11.2 Description 16](#_Toc157703708)

[Annex 3 - MPAI-wide terms and definitions 17](#_Toc157703709)

[Annex 4 - Notices and Disclaimers Concerning MPAI Standards 20](#_Toc157703710)

# Introduction

This Use Case and Functional Requirements: XR Venues (MPAI-XRV) – Collaborative Immersive Laboratory document describes the Use Case and identifies the Functional Requirements of a Collaborative Immersive Laboratory. This document is part of the MPAI XR Venues (MPAI-XRV) project addressing contexts enabled by Extended Reality (XR) – any combination of Augmented Reality (AR), Virtual Reality (VR) and Mixed Reality (MR) technologies – and enhanced by Artificial Intelligence (AI) technologies. The word “Venue” is used as a synonym for Real and Virtual Environments.

MPAI is also considering other Use Cases in the XR Venue space, some of which may employ identical or similar technologies to the Live Theatrical Stage Performance Use Case. This is facilitated by the MPAI approach of defining AI standards using AI Workflows (AIW) composed on AI Modules (AIM). Annex 2 - Other MPAI-XRV Use Cases provides additional details:

1. Live Theatrical Stage Performance
2. Collaborative immersive laboratory.
3. eSports Tournament.
4. Experiential retail/shopping.
5. Immersive art experience.
6. DJ/VJ performance at a dance party.
7. Live concert performance.
8. Experiential marketing/branding.
9. Meetings/presentations.

Important note:

All MPAI-XRV use cases and many of those not considered here involve the collection of large amounts of potentially sensitive Participant Data. This document does not address the processes that oversee the collection and processing of Participant Data. Rather, this document assumes that whatever processing is carried out, it conforms with the necessary and ethical/legal constraints, e.g., with the consent of the right holders of the data[[1]](#footnote-2),[[2]](#footnote-3).

Implementers must take great care in data security, assuring the correct possibility to opt-in/opt-out offered and proper use of the data. Also, care must be taken in training and testing AI Models to assure conformance with local laws and regulations and to prevent offensive or unintended experiences.

Interested parties should contact the [MPAI Secretariat](mailto:secretariat@mpai.community) in order not to miss future MPAI Calls for Technologies related to those Use Cases.

# Terms and definitions

The meaning of terms used in capital letters in this document have the meaning given in *Table 1*. The Terms of MPAI-wide applicability are defined in *Table 4*.

*Table 1* *–Terms and Definitions*

|  |  |
| --- | --- |
| **Term** | **Definition** |
| Actuator | A mechanism for modulating an experience in a real or virtual world. |
| AI Module | (AIM) A processing element receiving AIM-specific Inputs and producing AIM-specific Outputs according to according to its Function. An AIM may be an aggregation of AIMs. |
| 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. |
| Avatar | A rendered animated 3D digital object representing a real or fictitious person. |
| Biometric Data | Biological data collected from participants or performers including heart rate, electromyographic (EMG) data, skin conductance, etc. |
| Cognitive State | An estimation of the internal status of a human or avatar or a group thereof reflecting their understanding of the Environment, such as:  For a person: “Confused”, “Confident” and “Assured”.  For a group: “my team is going to lose”, or “we are winning”. |
| Controller | A manual control interface for participants, performers, or operators. |
| Cue Point | The position in the Script at any given time that an AIM, operator, or performer uses to generate Action according to the Script. |
| Data | Information in digital form. |
| * Format | The syntax and semantics of Data. |
| Descriptor | Digital representation of a Feature. |
| * *Action* | Components of the description that are used to create the complete user experience – in both Real and Virtual Environments – in accordance with the script. This includes all aspects of the experience including the performers and objects’ position, orientation, gesture, costume, audio, video, etc. |
| * Extraction | The process that extracts Descriptors from Data. |
| * Generation | The process that generates attributes to be applied to a scene or object according to the Script. |
| * Interpretation | The process that assigns a meaning to Descriptors. |
| * *Scene* | A Descriptor used to describe a Scene in the Real or Virtual Environment. |
| Dome display | Wrap-around immersive display surrounding an audience, using a projection screen or LED panels, technically known as Spatial Augmented Reality. |
| Emotion | An estimation of the internal status of a human or avatar or a group thereof resulting from their interaction with the Environment, such as:  For a person, “Victorious”, “Fearful” and “Angry”.  For a group: “Victorious”, “Fearful” and “Disappointed” |
| Environment | A portion of a real or a virtual world. |
| Extended Reality (XR) | Any combination of Augmented Reality (AR), Virtual Reality (VR) and Mixed Reality (MR). |
| Feature | An attribute of an object or a scene in a Real or Virtual Environment. |
| Interpreted Operator Controls | The assignment of meaning of data from control surfaces. |
| LiDAR Data | Data provided by LiDAR, an active time-of-flight sensor operating in the µm range – ultraviolet, visible, or near infrared light (900 to 1550 nm). |
| MoCap | Data capturing the movement of people or objects. |
| Participant | A human in a Real or Virtual Environment (Venue). |
| * Data | Data provided by or collected from Participants. |
| * Data Management | The set of legal, ethical, marketing, maintenance, etc. rules guiding the acquisition, retention, and processing of Data related to and provided by Participants. |
| * Engagement Engine | Algorithms to engage participants during or after the event allowing social interaction, commerce, and other engagement modalities. |
| * Status | The ensemble of information, expressed by Emotion, Cognitive State and Social Attitude, derived from observing the collective behaviour of participants in a Real and on-line Environment (via audio, video, interactive controllers, and smartphone apps). |
| Performer | A live actor performing on the Real Environment stage or represented by an avatar in the Virtual Environment. |
| Script | A collection of descriptors that the director/producer selects for execution at runtime controlling the action/experience in both Real and Virtual Worlds. |
| Sensor | A mechanism capturing data from a real or virtual Environment. |
| Show Control | Externally generated commands from an operator or show control system that pertain to desired Actions in the Real or Virtual Environment |
| Social Attitude | An element of the internal status related to the way a human or avatar intends to position vis-à-vis the Environment, e.g.:  For person: “Confrontational”, “Collaborative” and “Aggressive”.  For groups: “Confrontational”, “Collaborative” and “Aggressive” |
| Use Case | A particular instance of the Application domain target of an MPAI Application Standard. |
| Real Environment Venue Specification | The specification of the physical venue and all experiential technologies including lighting (fixture type and placement, DMX profile), FX, video displays, acoustics, sound reinforcement, stage props, preset cues and participant interactive devices. |
| Virtual Environment Venue Specification | The specification of the virtual venue and all available experiential systems for the delivery of audio, video, 3D environments, preset cues, and participant interactive devices. |
| Volumetric Visual Data | A set of samples representing the value of visual 3D data represented as textured meshes, points clouds, or UV + depth. |
| XR Venue | A combination of Real or Virtual Environments addressed by MPAI-XRV Use Cases. |

# References

1. https://docs.oracle.com/javadb/10.6.2.1/ref/rrefblob.html#:~:text=A%20BLOB%20(binary%20large%20object,do%20not%20hold%20character%20data
2. https://www.amnh.org/research/hayden-planetarium/digital-universe
3. https://www.astronomer.io/integrations/azure-blob-storage/
4. https://docs.google.com/document/d/1ew9rz\_DJLrWB\_9g3LFe0AfIObMflxPh36fq98XgUAIQ/edit?usp=sharing
5. Configuration of hyperspectral and multispectral imaging systems. / Amigo, José Manuel; Grassi, Silvia.
6. Hyperspectral Imaging. ed. / José Manuel Amigo. Elsevier, 2020. p. 17-34 (Data Handling in Science and Technology, Vol. 32). Research output: Chapter in Book/Report/Conference proceeding › Book chapter › Research › peer-review

# Scope of Collaborative Immersive Laboratory

Scope of XRV-CIL is to define a Common Intermediate Format able to:

1. Losslessly represent CT-MRI data with their formats.
2. Record the navigation parameters used by a user to navigate the slices up to the point on interest.

|  |  |
| --- | --- |
| **Term** | **Definition** |
| Slice | The basic information stored in the File corresponding to a section of a real object separated by a distance. |
| Object | A user-defined 3D structure extracted from the slices. |
| Contrast | The state of a portion of an image having a significantly different luminance or colour from an adjacent portion. |
| Contrasting agent | Substance that allows a sensor to capture an image with greater contrast than it could capture without contrasting agent. |
| Opacity |  |
| Threshold |  |
| Cursor |  |
| User Point of View | The Point of View selected by a user to view a 3D Object created from the slices from inside and outside of the Object. |

The parameters are:

1. The Spatial Attitude of the cursor in the Object.
2. The Euler angles of the Object.

A visualiser of multidimensional data will allow citizen scientists and researchers to join physically or virtually via avatar or volumetric representation of themselves for navigation, inspection, analysis, and simulations of scientific or industrial 3D/spatial models/datasets ranging from microscopic to macroscopic.

Examples are:

* View data in its actual 3D or 4D (over time) form through Immersive Reality.
* Present very large data sets that are generated by microscopes, patients, and industrial scanners.
* Format/reformat, qualify, and quantify sliced dataset with enhanced visualisation and analysis tools or import results for rapid correction of metadata for volumetric import.
* Provide tools for investigators to understand complex data sets completely and communicate their findings efficiently.

# Functional Requirements

## Content

The format shall be able to:

### In case of no time (static)

1. Create a virtual 3D object by ray-cast (<https://tinyurl.com/mvxfsvtr>) through the slices to the point of view of the user.
2. Add points at, draw lines (e.g., contours) around, highlight, extract, navigate inside the virtual 3D Object created by ray-casting.
3. Represent in 3D:
   1. All the information included in the slices.
   2. Specific objects with their spatial information extracted from the slices at the time the file was created.
4. Provide the volumes of 2.b in their spatial positions:
   1. Including the visual information between the volumes.
   2. Without the visual information between the 3D volumes (it may hide important features of the objects).
5. Facilitate the user’s “walking around” an object.
6. Provide number of objects in a user-specified volume.
7. Provide histograms of different parameters.
8. Compare different user-specified volumes.
9. Provide a slice of an object of a user-specified plane.

### In case of time (dynamic)

1. Rate of change of object features
2. User can select on from a panel of enhancements.

### Collaboration features

1. A master user drives the rendering of what other users see.

## Metadata

1. Information provided by the Metadata shall include:
   1. The distance between slices.
   2. A description of the objects that have been extracted in Content 2.b.
   3. Average number of objects per unit volume.
   4. Features of information inside an opaque object
      1. The diameter of a cell.
      2. The two diameters of an elongate object.

# Application Domains

1. Hyperspectral data
2. Anatomy
3. Medical
4. Biology (dense)
   1. Confocal (type of microscopy using a pin hole, very common)
   2. Micro Computational Topography (CT)
   3. Light sheet
5. Multispectral imaging of documents, painting, etc.
6. Spectroscopy
7. Material science
8. Non-destructive testing
9. Astronomy (representation of portions of the universe).
   1. Sparse but maybe not so sparse

## Hyperspectral data

Hyperspectral datablob [5,6] is another N-dimensional data type where dimensions are:

Examples of dimensions are:

* Time
* Space
* Other metadata used to annotate the time-space dimensions.

Hyperspectral imaging is a technique that collects and processes information across the electromagnetic spectrum to obtain the spectrum for each pixel in an image. This allows for the identification of objects and materials by analysing their unique spectral signatures. Applications of hyperspectral imaging include food quality & safety, waste sorting and recycling, and control and monitoring in pharmaceutical production.

A hyperspectral camera captures a scene’s light, separated into its individual wavelengths or spectral bands. It provides a two-dimensional image of a scene while simultaneously recording the spectral information of each pixel in the image.

The result is a hyperspectral image, where each pixel represents a unique spectrum. This unique spectrum can be compared to fingerprints. Since every material and compound reacts with light differently, their spectral signatures are also different. Just like fingerprints can be used to identify a person, the spectra can identify and quantify the materials in the scene.

## Anatomy

### Imaging types

**Several data formats already. We should define a Common Intermediate Format that can understand the different existing formats.**

## Medical

### Imaging types

1. CT
2. MRI
3. PET
4. Ultrasound
5. 2D pathology

### Viewing requirements

1. Visualiser should be usable by medical people and lay people (patient, students, etc.)
2. Visualise people numerical measurement (non-visual)
   1. Blood statistics (e.g., Panel 7)
   2. Oxygenation level
3. Visualise to give medical practitioners hints.
4. Visualise to help prognosis.
5. Visualise an anatomical reference (e.g., that of a particular human) in the context of viewing a particular sample (Visible human project – National Library of Medicine).
6. Allow a third party to follow one’s visualisation.

## Biology

### Imaging types

1. Confocal (type of microscopy using a pin hole, very common)
2. Micro Computational Topography (CT)
3. Light sheet
4. Electron microscopy

### Viewing requirements

1. Able to visualise 3/4 D datasets
2. Able to handle large datasets (tens of TBs)
3. Enables different annotations in different portion of the dataset (incl. time)
4. Extract numerical data from annotations
5. Enables communication of findings to parties.

# Use cases

## Virtual cadaver anatomy in the context of curriculum

## 3D modelling of microscopic medical images

Create 3D models of the fascia from 2D slices sampling microscopic medical images, classify cells based on their spatial phenotype morphology, enable the user to explore, interact with, zoom in the 3D model, count cells, and jump from a portion of the endoderm to another.

The system is composed of

1. An AI Module
   1. Reads the file containing the digital capture of 2D slices, e.g., of the endocrine system.
   2. Creates the 3D model
      1. Of the fascia.
      2. Converting a confocal image stack.
   3. Finds the cells in the model.
   4. Classifies the cells.
2. Humans project themselves into a common virtual collaborative environment to
   1. Navigate the 3D model.
   2. Interact with the 3D model.
   3. Zoom in the 3D model (e.g., x2000).
   4. Share information about their Points of View so that all can vernally comment on and annotate a particular feature.

Another Use Case relates to a human (e.g., an athlete) exercising a sport.

The human has previously had their body MRI scanned.

A motion capture system captures the relevant data.

An AIM creates the model of the human body including:

1. The surface of the body.
2. The organs of the body based on the MRI scan.

A human can zoom into a library of images and correlate progression of part of the body over time.

Relevant data formats are:

1. Image Data: TIFF, PNG, JPEG, DICOM, VSI, OIR, IMS, CZI, ND2, and LIF files
2. Mesh Data: OBJ, FBX, and STEP files
3. Volumetric Data: OBJ, PLY, XYZ, PCG, RCS, RCP and E57[[3]](#footnote-4)
4. Supplemental Slides from Powerpoint/Keynote/Zoom
5. 3D Scatterplots from CSV files (LiDAR)
6. Confocal.

## Animation of medical images

Animate data over time such as Functional MRI data or cardiac rhythm.

Display, navigate, zoom in 3d representations of body organs (brain, heart) captured with different technologies (fMRI, echocardiogram, …) rendered in a virtual space populated by digital humans.

Other examples are: a participant shares with other participants their Point of View so that everybody can see the 3D object from the same Point of View.

Participants can collaborate if

1. They use the same “viewer”.
2. The 3D data
   1. Are represented in a standard format.
   2. Navigation commands are standardised.

## Modelling of astronomical space-time

Assemble hyperspectral data from imaging of portions of the universe. Data blob is the format used. Omniverse from NVIDIA.

## Description

## Specific application areas

### Microscopic dataset visualisation

1. Deals with different object types, e.g.:
   1. 3D Visual Output of a microscope.
   2. 3D model of the brain of a mouse.
   3. Molecules captured as 3D objects by an electronic microscope.
2. Create and add metadata to a 3D audio-visual object:
   1. Define a portion of the object – manual or automatic.
   2. Assign physical properties to (different parts) of the 3D AV object.
   3. Annotate a portion of the 3D AV object.
   4. Create links between different parts of the 3D AV object.
3. Enter, navigate and act on 3D audio-visual objects:
   1. Define a portion of the object – manual or automatic.
   2. Count objects per assigned volume size.
   3. Detect structures in a (portion of) the 3D AV object.
   4. Deform/sculpt the 3D AV object.
   5. Combine 3D AV objects.
   6. Call an anomaly detector on a portion with an anomaly criterion.
   7. Follow a link to another portion of the object.
   8. 3D print (portions of) the 3D AV object.

### Macroscopic dataset visualisation and simulation

1. Deals with different dataset types, e.g.:
   1. Stars, 3D star maps (HIPPARCOS, Tycho Catalogues, etc.).
   2. Deep-sky objects (galaxies, star clusters, nebulae, etc.).
   3. Deep-sky surveys (galaxy clusters, large-scale structures, distant galaxies, etc.).
   4. Satellites and man-made objects in the atmosphere and above, space junks, planetary and Moon positions.
   5. Real-time air traffic.
   6. Geospatial information including CO2 emission maps, ocean temperature, weather, etc.
2. Simulation data
   1. Future/past positions of celestial objects.
   2. Stellar and galactic evolution.
   3. Weather simulations.
   4. Galaxy collisions.
   5. Black hole simulation.
3. Create and add metadata to datasets and simulations:
   1. Assign properties to (different parts) of the datasets and simulations.
   2. Define a portion of the dataset – manual or automatic.
   3. Annotate a portion of the datasets and simulations.
   4. Create links between different parts of the datasets and simulations.
4. Enter, navigate, and act on 3D audio-visual objects:
   1. Search data for extra-solar planets.
   2. Count objects per assigned volume size.
   3. Detect structures and trends in a (portion of) the datasets and simulations.
   4. Call an anomaly detector on a portion with an anomaly criterion.

### Educational lab

1. Experiential learning models simulations for humans.
2. Group navigation across datasets and simulations.
3. Group interactive curricula.
4. Evaluation maps.

### Collaborative CAD

1. Building information management.
2. Collaborative design and art.
3. Collaborative design reviews.
4. Event simulation (emergency planning etc.).
5. Material behaviour simulation (thermal, stress, collision, etc.).
6. Basics about MPAI

# General

In recent years, Artificial Intelligence (AI) and related technologies have been introduced in a broad range of applications affecting the life of millions of people and are expected to do so much 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. In addition, some AI technologies may carry inherent risks, e.g., in terms of bias toward some classes of users making the need for standardisation more important and urgent than ever.

The above considerations have prompted the establishment of the international, unaffiliated, not-for-profit Moving Picture, Audio and Data Coding by Artificial Intelligence (MPAI) organisation with the mission to develop *AI-enabled data coding standards* to 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 – and new in standardisation – type of Specification includes standard operating procedures that 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.

# Governance of the MPAI Ecosystem

The technical foundations of the MPAI Ecosystem [4] are currently provided by the following documents developed and maintained by MPAI:

1. Technical Specification.
2. Reference Software Specification.
3. Conformance Testing.
4. Performance Assessment.
5. Technical Report

An MPAI Standard is a collection of a variable number of the 5 document types.

*Figure 1* depicts the MPAI ecosystem operation for conforming MPAI implementations.

Diagram

Description automatically generated

*Figure 1 – The MPAI ecosystem operation*

Technical Specification: Governance of the MPAI Ecosystem Table 2 identifies the following roles in the MPAI Ecosystem:

Table 2 - Roles in the MPAI Ecosystem

|  |  |
| --- | --- |
| MPAI | Publishes Standards.  Establishes the not-for-profit MPAI Store.  Appoints Performance Assessors. |
| Implementers | Submit Implementations to Performance Assessors. |
| Performance Assessors | Inform Implementation submitters and the MPAI Store if Implementation Performance is acceptable. |
| Implementers | Submit Implementations to the MPAI Store. |
| MPAI Store | Assign unique ImplementerIDs (IID) to Implementers in its capacity as ImplementerID Registration Authority (IIDRA)[[4]](#footnote-5).  Verifies security and Tests Implementation Confor­mance. |
| Users | Download Implementations and report their experience to MPAI. |

# AI Framework

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.

*Figure 2* depicts the MPAI-AIF Reference Model under which Implementations of MPAI Application Standards and user-defined MPAI-AIF Conforming applications operate [5].

A computer screen shot of a computer

Description automatically generated

*Figure 2 – The AI Framework (AIF) Reference Model*

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.

An AIW is defined by its Function and input/output Data and by its AIM topology. Likewise, an AIM is defined by its Function and input/output Data. MPAI standard are silent on the technology used to implement the AIM which may be based on AI or data processing, and implemented in software, hardware or hybrid software and hardware technologies.

MPAI also defines 3 Interoperability Levels of an AIF that executes an AIW. Table 3 gives the characteristics of an AIW and its AIMs of a given Level:

Table 3 - MPAI Interoperability Levels

|  |  |  |
| --- | --- | --- |
| *Level* | AIW | AIMs |
| *1* | An implementation of a use case | Implementations able to call the MPAI-AIF APIs. |
| *2* | An Implementation of an MPAI Use Case | Implementations of the MPAI Use Case |
| *3* | An Implementation of an MPAI Use Case certified by a Performance Assessor | Implementations of the MPAI Use Case certified by Performance Assessors |

# Personal Status

## General

*Personal Status* is the set of internal characteristics of a human and a machine making a conversation. Reference [6] identifies three Factors of the internal state:

1. *Cognitive State* is a typically rational result from the interaction of a human/avatar with the Environment (e.g., “Confused”, “Dubious”, “Convinced”).
2. *Emotion* is typically a less rational result from the interaction of a human/avatar with the Environment (e.g., “Angry”, “Sad”, “Determined”).
3. *Social Attitude* is the stance taken by a human/avatar who has an Emotional and a Cognitive State (e.g., “Respectful”, “Confrontational”, “Soothing”).

The Personal Status of a human can be displayed in one of the following Modalities: *Text, Speech,* *Face,* or *Gesture*. More Modalities are possible, e.g., the body itself as in body language, dance, song, etc. The Personal Status may be shown only by one of the four Modalities or by two, three or all four simultaneously.

## Personal Status Extraction

Personal Status Extraction (PSE) is a composite AIM that analyses the Personal Status conveyed by Text, Speech, Face, and Gesture – of a human or an avatar – and provides an estimate of the Personal Status in three 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., at least one of Emotion, Cognitive State, and Attitude).

Figure 3 depicts the Personal Status estimation process:

1. Descriptors are extracted from Text, Speech, Face Object, and Body Object. Depending on the value of Selection, Descriptors can be provided by an AIM upstream.
2. Descriptors are interpreted and the specific indicators of the Personal Status in the Text, Speech, Face, and Gesture Modalities are derived.
3. Personal Status is obtained by combining the estimates of different Modalities of the Personal Status.

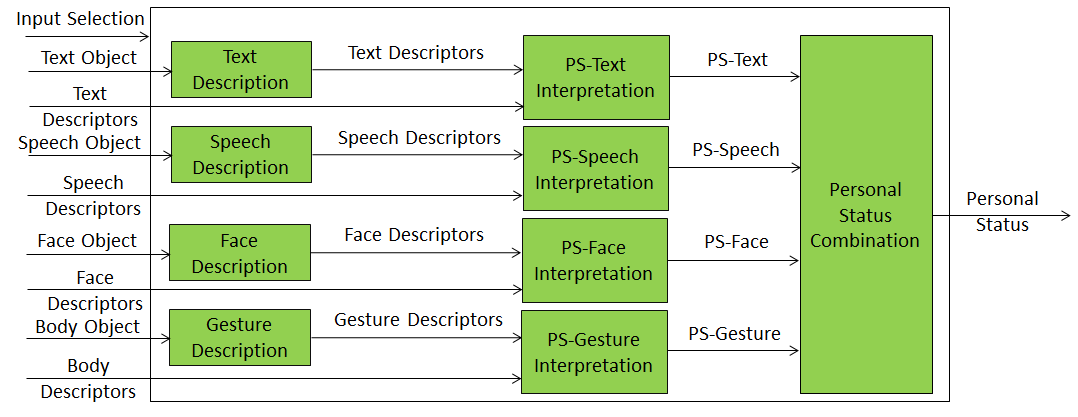


Figure 3 – Reference Model of Personal Status Extraction

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 Body Object without exposing PS-Gesture Descriptors.

1. Other MPAI-XRV Use Cases

# eSports Tournament (XRV-EST).

## Purpose

To define interfaces between components enabling an XR Theatre (RW) to host any pre-existing VW game for the purpose of producing an esports tournament with RW and VW audience interactivity. To the extent that the game possesses the required interfaces, the XR Theatre can drive action within the VW.

## Description

The eSports Tournament Use Case consists of the following:

1. Two teams of 5 RW players are arranged on either side of a RW stage, each using a computer to compete within a common real-time Massively Multiplayer Online (MMO) VW game space.
2. The 10 players in the VW are represented by avatars each driven by
   1. Role (e.g., magicians, warriors, soldier, etc.).
   2. Properties (e.g., costumes, physical form, physical features).
   3. Actions (e.g., casting spells, shooting, flying, jumping) operating in the VW
3. The VW is populated by
   1. Avatars representing the other players.
   2. Autonomous characters (e.g., dragon, monsters, various creatures)
   3. Environmental structures (e.g., terrain, mountains, bodies of water).
4. The action in the VW is captured by multiple VW cameras and
   1. Projected onto an immersive screen surrounding RW spectators
   2. Live streamed to remote spectators as a 2D video.

with all related sounds of the VW game space.

1. A shoutcaster calls the action as the game proceeds.
2. The image of RW players, player stats or other information or imagery may also be displayed on the immersive screen and the live stream.
3. The RW tournament venue is augmented with lighting and special effects, music, and costumed performers.
4. Interactions:
   1. Live stream viewers interact with one another and with commentators through live chats, Q&A sessions, etc.
   2. RW spectators interact through shouting, waving and interactive devices (e.g., LED wands, smartphones) through processing where:
      1. Data are captured by camera/microphone or wireless data interface (see RW data in ***Error! Reference source not found.***).
      2. Features are extracted and interpreted.
5. RW/VW actions can be generated as a result of:
   1. In-person or remote audience behaviour (RW).
   2. Data collected from VW action (e.g., spell casting, characters dying, bombs exploding)
6. At the end of the tournament, an award ceremony featuring the winning players on the RW stage is held with great fanfare.

# Experiential retail/shopping.

## Purpose

To define components and interfaces to facilitate a retail shopping experience enhanced using immersive/interactive technologies driven with AI.

Enhancements includes:

1. Faster locating of products
2. Easy access to product information and reviews.
3. Delivery if special offers
4. Collaborative shopping (members of a group know what other members have purchased)
5. Product annotation according to user preference and theming of the environment according to season and user preferences.
6. Analytics of data collected to inform sales and marketing decisions, inventory control and business model optimisation.
7. Offering remote shoppers the ability to enter a digital twin of real world store as an avatar (as a 3D Graphics or as a volumetric “hologram”) and interact with friends who are physically or virtually present in the real world store.

## Description and flow of actions

The environment displays the following features:

1. It gives the user the impression that it is intelligent because the environment has access to the user’s identity/behaviour/preferences/shopping history/shopping list and is capable to guide the buyer to the area containing products of their supposed interest, propose products, annotate products and to display a particular product and make it flash because the environment thinks it is of interest to the buyer.
2. It broadcasts music etc. to all buyers in the environment driven by the preferences. Friends in the shop at the same time can “meet”, but buyers can opt out from being discoverable (by the store, by friends etc.). Buyers can opt out from the loyalty card and not have the product they buy recorded by the shop.
3. It can be digitally rethemed for different occasions.
4. It offers experience that can takes shape can be anywhere, e.g., in a vehicle or in a public transit space.
5. It enables remote shoppers to virtually enter a digital twin of the store and interact with friends who are physically present in the store for a collaborative shopping experience.

# Immersive art experience.

## Purpose

Define interfaces and components to enhance magical Environments created by skilled artists to provide each user with a unique interactive experience including the ability to modify the environment per their personal style and preferences.

## Description

Immersive art experiences such as Immersive Van Gogh provide visitors with a visually and aurally immersive experience, often based on the work of a specific artist. These are typically passive walk-through and sit-down experiences. The addition of AI to these Environments allows numerous enhancements including the recognition of individual visitors, allowing them to interact with and modify these environments based on pre-selected preferences and style choices. AI style transfer allows the featured artist’s style to be applied to unique visitor interactions which might include AI voice or text-based image diffusion, gesture-based interactions, proximity effects and more. The addition of AR glasses allows visitors to experience, create and interact with “holograms” within the Environment. Biometric wearables allow the AI to monitor and adjust the multisensory experience to maximize target brain/nervous system states related to well-being, restorative states and more. The XR Venue model also allows visitors in the RW and VW to interact.

# DJ/VJ performance at a dance party.

## Purpose

Define interfaces and components to enhance the overall experience within a nightclub, lounge or dance party Environment. The goal is to empower the DJ/VJ to create and control entertaining immersive and interactive experiences that reduce social inhibitions, encourage play, invoke a greater sense of togetherness, encourage personal connections, evoke altered states of consciousness, amplify user’s self-expression and generally create a highly pro-social experience for participants.

## Description

Dance parties, lounges, clubs, and electronic music festivals use powerful visuals, sound and other effects to captivate participants. The DJ (disc jockey) mixes audio tracks, energizes the crowd and is central to the experience. However, the visual artist or VJ (video jockey) is also an important contributor, often supported by lighting, laser and effects operators, dancers, performers and more. Quite often these venues offer peripheral activities as well to further engage participants off of the dance floor, including interactive screens, spatial art, vendors offering costumes and LED accessories. These venues can be thought of as play spaces. Pro-social intoxicants such as alcohol are sometimes used to lower inhibitions that would otherwise limit social connections. XR Venues can supercharge the dance party experience by providing powerful immersive visuals and by including VW participants Assisted by AI, all music, visuals, lights, and effects can be controlled by a single DJ (or immersive jockey) using gestures, simple control surfaces, vocal commandsm and such. In addition, expanded peripheral activities for deeper engagement might include immersive visuals that respond to emergent crowd behaviours, “photonic go-go booths” that modulate immersive visuals to amplify the creative expression of dancers’ movements, and AI-based matchmaking that fosters connections between like-minded attendees.

# Live concert performance.

## Purpose

Define interfaces and components to enhance live musical concerts with AI-driven visuals and special effects and allow enhanced audience participation while extending concert performances into the metaverse.

## Description

Similar to live theatrical stage performances, musical concerts – whether orchestral or popular music – are increasingly using visuals and other effects to enhance the audience experience. A band or orchestral musicians on stage can be substantially enhanced by video projections from a live VJ, audio responsive visuals, image magnification from cameras and other effects. In addition, skilful live mixing of audio is critical to the audience experience, but it complicated by architectural properties of the physical venue. AI can dynamically optimize the listening experience and allow tight synchronization of visuals with spontaneous musical performances in addition to optimizing the VW experience for remote attendees.

# Experiential marketing/branding.

## Purpose

Define interfaces and components to enhance a wide range of experiences in support of corporate branding.

## Description

Wherever there are a lot of people gathered we often find advertisers or corporate brands seeking visibility. Experiential marketing goes beyond simple advertising or signage by offering memorable experiences to attendees. Experiential marketing often makes use of pop-up venues or storefronts co-located at festivals, sporting events, concerts and more. Digital interactive or immersive experiences are increasingly employed, often incorporating branded story-worlds or iconic brand elements. The XRV allows delivery of a unique experience to each participant and deeper engagement to build brand loyalty. In addition, the experience can be extended into the VW to reach a larger number of attendees.

# Meetings/presentations.

## Purpose

Define interfaces and components to enhance live presentations and dialog, both in RW and VW, using rich multimedia, dialog mapping, AI-based mediation and fact checking.

## Description

Meetings and presentations are increasingly hybrid, including both live and virtual attendees, allowing the sharing of rich multimedia content including documents, videos and website links. Use of an XRV for presentations and especially dialog - including political discourse - presents an opportunity for AI to monitor, track, organize and summarize numerous data in real-time to overlook hyperbole and guide the conversation toward rapid convergence on positive outcomes. Real-time fact-finding/fact-checking, dialog mapping (creating a logical tree showing relationships and dependencies between various points raised), group polling and other advanced methods can be employed in an XRV to guide dialog or facilitate presentations.

1. MPAI-wide terms and definitions

The Terms used in this standard whose first letter is capital and are not already included in***Error! Reference source not found.***are defined in Table 4*.*

Table 4 - 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 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 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. |
| Conformance Tester | An entity authorised by MPAI to Test the Conformance of an Implementation. |
| Conformance Testing | The normative document specifying the Means to Test the Conformance of an Implementation. |
| Conformance Testing Means | Procedures, tools, data sets and/or data set characteristics to Test the Conformance of an Implementation. |
| Connection | A channel connecting an output port of an AIM and an input port of an AIM. |
| Controller | A Component that manages and controls the AIMs in the AIF, so that they execute in the correct order and at the time when they are needed. |
| Data | Information in digital form. |
| Data Format | The standard digital representation of Data. |
| Data Semantics | The meaning of Data. |
| Device | A hardware and/or software entity running at least one instance of an AIF. |
| Ecosystem | The ensemble of the following actors: MPAI, MPAI Store, Implementers, Conformance Testers, Performance Testers and Users of MPAI-AIF Im­plementations as needed to enable an Interoperability Level. |
| Event | An occurrence acted on by an Implementation. |
| Explainability | The ability to trace the output of an Implementation back to the inputs that have produced it. |
| Fairness | The attribute of an Implementation whose extent of applicability can be assessed by making the training set and/or network open to testing for bias and unanticipated results. |
| Function | The operations effected by an AIW or an AIM on input data. |
| Global Storage | A Component to store data shared by AIMs. |
| 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). |
| Internal Storage | A Component to store data of the individual AIMs. |
| Interoperability | The ability to functionally replace an AIM/AIW with another AIM/AIW having the same Interoperability Level |
| Interoperability Level | The attribute of an AIW and its AIMs to be executable in an AIF Implementation and to be:   1. Implementer-specific and satisfying the MPAI-AIF Standard *(Level 1)*. 2. Specified by an MPAI Application Standard (*Level 2)*. 3. Specified by an MPAI Application Standard and certified by a Performance Assessor (*Level 3)*. |
| Knowledge Base | Structured and/or unstructured information made accessible to AIMs via MPAI-specified interfaces |
| Message | A sequence of Records. |
| Normativity | The set of attributes of a technology or a set of technologies specified by the applicable parts of an MPAI standard. |
| Performance | The attribute of an Implementation of being Reliable, Robust, Fair and Replicable. |
| Performance Assessment | The normative document specifying the procedures, the tools, the data sets and/or the data set characteristics to Assess the Grade of Performance of an Implementation. |
| Performance Assessment Means | Procedures, tools, data sets and/or data set characteristics to Assess the Performance of an Implementation. |
| Performance Assessor | An entity authorised by MPAI to Assess the Performance of an Implementation in a given Application domain |
| Port | A physical or logical communication interface of an AIM. |
| Profile | A particular subset of the technologies used in MPAI-AIF or an AIW of an Application Standard and, where applicable, the classes, other subsets, options, and parameters relevant to that subset. |
| Record | Data with a specified structure. |
| Reference Model | The AIMs and theirs Connections in an AIW. |
| Reference Software | A technically correct software implementation of a Technical Specific­ation containing source code, or source and compiled code. |
| Reliability | The attribute of an Implementation that performs as specified by the Application Standard, profile and version the Implementation refers to, e.g., within the application scope, stated limitations, and for the period of time specified by the Implementer. |
| Replicability | The attribute of an Implementation whose Performance, as Assessed by a Performance Assessor, can be replicated, within an agreed level, by another Performance Assessor. |
| Robustness | The attribute of an Implementation that copes with data outside of the stated application scope with an estimated degree of confidence. |
| Scope | The domain of applicability of an MPAI Application Standard. |
| Service Provider | An entrepreneur who offers an Implementation as a service (e.g., a recommendation service) to Users. |
| Specification | A collection of normative clauses. |
| Standard | The ensemble of Technical Specification, Reference Software, Conformance Testing and Performance Assessment of an MPAI application Standard. |
| Technical Specification | (Framework) the normative specification of the AIF.  (Application) the normative specification of the set of AIWs belon­ging to an application domain along with the AIMs required to Im­plem­ent the AIWs that includes:   1. The formats of the Input/Output data of the AIWs implementing the AIWs. 2. The Connections of the AIMs of the AIW. 3. The formats of the Input/Output data of the AIMs belonging to the AIW. |
| Testing Laboratory | A laboratory accredited by MPAI to Assess the Grade of Performance of Implementations. |
| Time Base | The protocol specifying how AIF 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. |

1. Notices and Disclaimers Concerning MPAI Standards

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” “Conformance Testing” and, where applicable, “Performance Testing”.

The 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 specially established MPAI Development Committees that operate based on 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 of any kind concerning its Standards, and expressly disclaims all warranties, expressed or implied, concerning any of its Standards, including but not limited to the warranties of merchantability, fitness for a particular purpose, non-infringement etc. MPAI Standards are supplied “AS IS”.

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

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

MPAI alerts users that practising 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, a 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. https://openvoicenetwork.org/documents/ovn\_ethical\_guidlines\_voice\_experiences.pdf [↑](#footnote-ref-2)
2. https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html [↑](#footnote-ref-3)
3. https://info.vercator.com/blog/what-are-the-most-common-3d-point-cloud-file-formats-and-how-to-solve-interoperability-issues [↑](#footnote-ref-4)
4. At the time of publication of this Technical Report, the MPAI Store was assigned as the IIDRA. [↑](#footnote-ref-5)