



Public document

N1366

2023/08/23

Source Requirements (XRV)

Title Use Cases and Functional Requirements: XR Venues (MPAI-XRV) – Live Theatrical Stage Performance

Target MPAI Community

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. This document, part of the MPAI-XRV – Live Theatrical Stage Performance Call for Technologies set of documents addresses the Functional Requirements of the Live Theatrical Stage Performance Use Case.

Responses to the Call should be received by the [MPAI Secretariat](#) by 2023/11/20T23:59 UTC.

Contents

Abstract	1
1 Introduction	2
2 Terms and definitions	4
3 References	5
4 A Real-Virtual Interaction Model	6
5 Live Theatrical Stage Performance Use Case	7
5.1.1 Purpose	7
5.1.2 Description and flow of actions	7
6 Functional requirements	8
6.1.1 Typical Configuration	8
6.1.2 Reference Architecture	9
6.1.3 AI Modules	10
6.1.4 AIM I/O Data Formats	11
7 Technologies requested	12
7.1 Scene Descriptors	12
7.2 Participant Descriptors	12
7.3 Participant Status	12
7.4 Script	13
7.5 Cue point	13
7.6 Interpreted Operator Control	13
7.7 Action Descriptors	13
7.8 Real Experience Generation	13
7.8.1 Lighting	13
7.8.2 FX (Effects)	14
7.8.3 Audio-Visual (A/V)	14

7.8.4	Real Experience Venue specification.....	14
7.9	Virtual Experience Generation.....	14
7.9.1	Virtual Experience Descriptors	14
7.9.2	Audio-Visual (A/V).....	14
7.9.3	Virtual Experience Venue specification.....	15
Annex 1 -	Basics about MPAI.....	16
1	General	16
2	Governance of the MPAI Ecosystem	16
3	AI Framework	17
4	Personal Status	18
4.1	General	18
4.2	Personal Status Extraction.....	18
Annex 2 -	Other MPAI-XRV Use Cases	20
5	eSports Tournament (XRV-EST).....	20
5.1	Purpose	20
5.2	Description	20
6	Experiential retail/shopping.	21
6.1	Purpose	21
6.2	Description and flow of actions	21
7	Collaborative immersive laboratory	21
7.1	Purpose	21
7.2	Description	22
7.3	Specific application areas.....	22
7.3.1	Microscopic dataset visualisation.....	22
7.3.2	Macroscopic dataset visualisation and simulation	22
7.3.3	Educational lab.....	23
7.3.4	Collaborative CAD.....	23
8	Immersive art experience.	23
8.1	Purpose	23
8.2	Description	23
9	DJ/VJ performance at a dance party.	24
9.1	Purpose	24
9.2	Description	24
10	Live concert performance.....	24
10.1	Purpose	24
10.2	Description	24
11	Experiential marketing/branding.....	25
11.1	Purpose	25
11.2	Description	25
12	Meetings/presentations.....	25
12.1	Purpose	25
12.2	Description	25
Annex 3 -	MPAI-wide terms and definitions	26
Annex 4 -	Notices and Disclaimers Concerning MPAI Standards	29

1 Introduction

This Use Case and Functional Requirements: XR Venues (MPAI-XRV) – Live Theatrical Stage Performance document describes the Use Case and identifies the Functional Requirements of Live Theatrical Stage Performance Use Case. This document is part of the MPAI XR Venues (MPAI-

XR) 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. This document should be considered jointly with the Call for Technologies: XR Venues (MPAI-XRV) – Live Theatrical Stage Performance [1] and Framework Licence: XR Venues (MPAI-XRV) – Live Theatrical Stage Performance [2].

The purpose of the planned Technical Specification is to define interfaces of AI Modules able to perform functions that facilitate live multisensory immersive stage performances which ordinarily require extensive on-site show control staff to operate. Use of the AI Modules organised in AI Workflows enabled by the Technical Specification¹ will allow more direct, precise yet spontaneous show implementation and control to achieve the show director’s vision. It will also free staff from repetitive and technical tasks allowing them to amplify their artistic and creative skills.

The Call for Technologies [1] seeks to obtain technologies that support some of and preferably all the Functional Requirements identified in this document that MPAI intends to use in the development of the planned Technical Specification. Those proposing technologies in response to the Call are requested to state their availability to license their technologies, if adopted by MPAI, in conformity with the Framework Licence [2].

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. eSports Tournament.
2. Experiential retail/shopping.
3. Collaborative immersive laboratory.
4. Immersive art experience.
5. DJ/VJ performance at a dance party.
6. Live concert performance.
7. Experiential marketing/branding.
8. 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^{2,3}.

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](#) in order not to miss future MPAI Calls for Technologies related to those Use Cases.

¹ See Annex 1 - Basics about MPAI and with the MPAI process to develop standards [3] for additional information addition

² https://openvoicenetwork.org/documents/ovn_ethical_guidelines_voice_experiences.pdf

³ <https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html>

2 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 7*.

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 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.
- <i>Action</i>	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.
- <i>Scene</i>	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.

3 References

1. MPAI; MPAI XR Venues (MPAI-XRV) - Live Theatrical Stage Performance Call for Technologies; N1365; <https://mpai.community/standards/mpai-xrv/call-for-technologies/>

2. MPAI; MPAI XR Venues (MPAI-XRV) - Live Theatrical Stage Performance Framework Licence; N1367; <https://mpai.community/standards/mpai-xrv/framework-licence/>
3. MPAI; Patent Policy; <https://mpai.community/about/the-mpai-patent-policy/>
4. MPAI; Technical Specification: Governance of the MPAI Ecosystem (MPAI-GME) V1.1; <https://mpai.community/standards/mpai-gme/>.
5. MPAI; Technical Specification: Artificial Intelligence Framework (MPAI-AIF) V1.1; <https://mpai.community/standards/mpai-aif/>.
6. MPAI; Technical Specification: Multimodal Conversation (MPAI-MMC) V1.2; <https://mpai.community/standards/mpai-mmc/>.
7. MPAI; Technical Specification: Context-based Audio Experience (MPAI-MMC) V1.4; <https://mpai.community/standards/mpai-mmc/>.
8. MPAI; Technical Specification: The Governance of the MPAI Ecosystem V1, 2021; <https://mpai.community/standards/mpai-gme/>.
9. MPAI; Technical Report: MPAI Metaverse Model (MPAI-MMM) – Functionalities; V1; <https://mpai.community/standards/mpai-mmm/>.
10. MPAI; Technical Report: MPAI Metaverse Model (MPAI-MMM) – Functionality Profiles; V1; <https://mpai.community/standards/mpai-mmm/>.
11. Universal Scene Description; <https://openusd.org/>

4 A Real-Virtual Interaction Model

An important feature of MPAI-XRV is the strong interaction with – and sometimes even interchangeability of – a Real/Virtual Environment and a Virtual/Real World Environment. The MPAI-XRV model, depicted in *Figure 1*, is helpful to guide the analysis of the MPAI-XRV use cases.

The model assumes that there is a complete symmetry between the actions performed and the data formats exchanged between a Real Environment and a Virtual Environment (e.g., a metaverse).

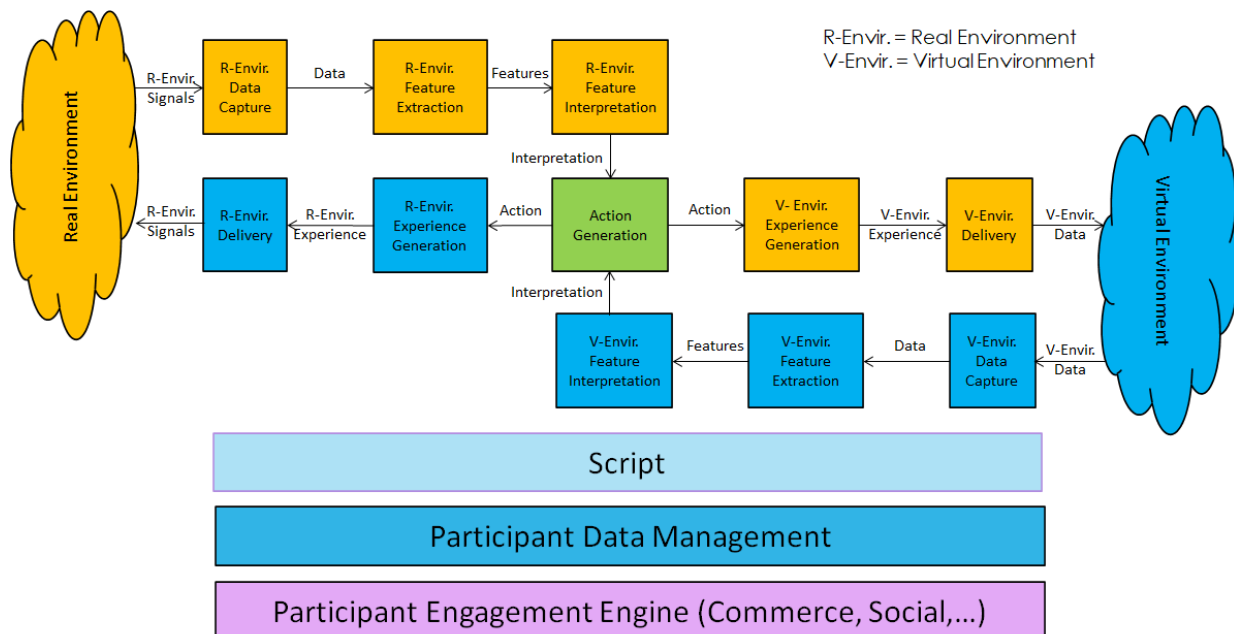


Figure 1 – Real World (yellow) and Virtual (blue) Interactions.

Table 2 defines the functions of the processing elements that bidirectionally capture and process data from a Real Environment which are used to generate actions and deliver experiences in a Virtual Environment and vice versa. Table 2 describes the functions of the identified components.

Table 2 - The functions of the components in the MPAI-XRV Model

Data Capture	Captures Environment as collections of signals and/or Data.
Feature Extraction	Analyses Data to extract Descriptors.
Feature Interpretation	Analyses Descriptors to yield Interpretations.
Action Generation	Analyses Interpretations to generate Actions.
Experience Generation	Analyses Actions to generate Environment.
Environment Delivery	Delivers Environment as collections of signals and/or Data.

5 Live Theatrical Stage Performance Use Case

5.1.1 Purpose

Theatrical stage performances such as Broadway theatre, musicals, dramas, operas, and other performing arts increasingly use video scrims, backdrops, and projection mapping to create digital sets rather than constructing physical stage sets, allowing animated backdrops, and reducing the cost of mounting shows.

The use of immersion domes – especially LED volumes – promises to surround audiences with virtual environments that the live performers can inhabit and interact with. In addition, Live Theatrical Stage Performance can extend into the metaverse as a digital twin implement the model depicted in *Figure 1*. In this case, elements of the Virtual Environment experience can be projected in the Real Environment and elements of the Real Environment experience can be rendered in the Virtual Environment (metaverse).

Use of AI in Live Theatrical Stage Performance will allow:

1. Rapid mounting of a show into a variety of real and virtual venues.
2. Orchestration of the complex lighting, video, audio, and stage set cues that must adapt to the pace of live performers without extensive staff.
3. Large shows to tour to smaller venues that otherwise could not support complex productions.
4. Live performances spanning both Virtual- and Real-Environments, including in-person or remote participants and performers with enhanced participant interactivity.
5. A more direct connection between the artist and participants by consolidating many complex experiential modalities into a simple user interface.
6. Artists to access a large amount of data from opted-in individuals and which can be incorporated into the visual and musical performance. Each show can thus be unique for each audience.

5.1.2 Description and flow of actions

The typical set up can be described as follows:

1. A physical stage.
2. Lighting, projections (e.g., dome, holograms, AR goggles), and Special Effects (FX).
3. Audience (standing or seated) in the real and virtual venue and external audiences via interactive streaming.
4. Interactive interfaces to allow audience participation (e.g., voting, branching, real-virtual action generation).
5. Performers on stage, platforms around domes or moving through the audience (immersive theatres).

6. Multisensory experience delivery system (immersive video and spatialised audio, touch, smell).
7. Capture of biometric data from audience and/or performers from wearables, sensors embedded in the seat, remote sensing (e.g., audio, video, lidar).
8. Show operator(s) to allow manual augmentation and oversight of an AI that has been trained by show operator activity.
9. Virtual Environment (metaverse) that mirrors selected elements of the Real Environment. For example, performers on the stage are mirrored by digital twins in the metaverse, using:
 - a. Capture body motion (MoCap) to animate an avatar.
 - b. Keyed 2D image mapped on a plane.
 - c. Volumetrically captured 3D images producing photorealistic digital embodiments.
10. Real Environment can also mirror selected elements of the Metaverse, similar to in-camera visual effects/virtual production techniques. For instance, elements of the Metaverse such as, avatars, landscape, sky, objects can be represented in the Real Environment through:
 - a. Immersive displays
 - b. The floor of the stage itself and set pieces on the stage may be projection-mapped or wrapped with LED to integrate them into the immersive environment. This allows, for instance set pieces such as a tree, to come alive with moving leaves, blooming flowers, or ripening fruits, and for the tree to cast a virtual shadow across the stage from a virtual light source moving across an immersive dome. Many of these elements may be extracted from the metaverse and projected into the real-world immersive environment.
 - c. Augmented reality overlays using AR glasses, “hologram” displays or scrims.
 - d. Lighting and FX.
11. The physical stage and set pieces blend seamlessly into the virtual 3D backdrop projected onto the dome such that the spectators perceive as a single immersive environment.
12. Real performers enter the stage. As they move about the stage, whether dancing, acting, etc., their performance may be mirrored in the Virtual Environment (metaverse) by tracking performer’s motion, gesture, vocalisation, and biometrics. The performance is accompanied by music, lighting, and FX.
13. In addition, virtual performers in the Virtual Environment (metaverse) may be projected onto the real-world immersive environment via immersive display, AR, etc.
14. The Script or cue list describes the show events, guiding and synchronising the actions of all AI Modules (AIM) as the show evolves from cue to cue and scene to scene. In addition to performing the show, the AIMs might spontaneously innovate show variations amplify the actions of performers or respond to commands from operators by modifying the Real or Virtual Environment within scripted guidelines.

6 Functional requirements

6.1.1 Typical Configuration

A view of the devices and data required for a typical live theatrical stage performance configuration with both a Real Environment and Virtual Environment venues are provided by Figure 2. The objective of the Live Theatrical Stage Performance Use Case is to automate as many of these functions as possible using AI Modules (AIMs).

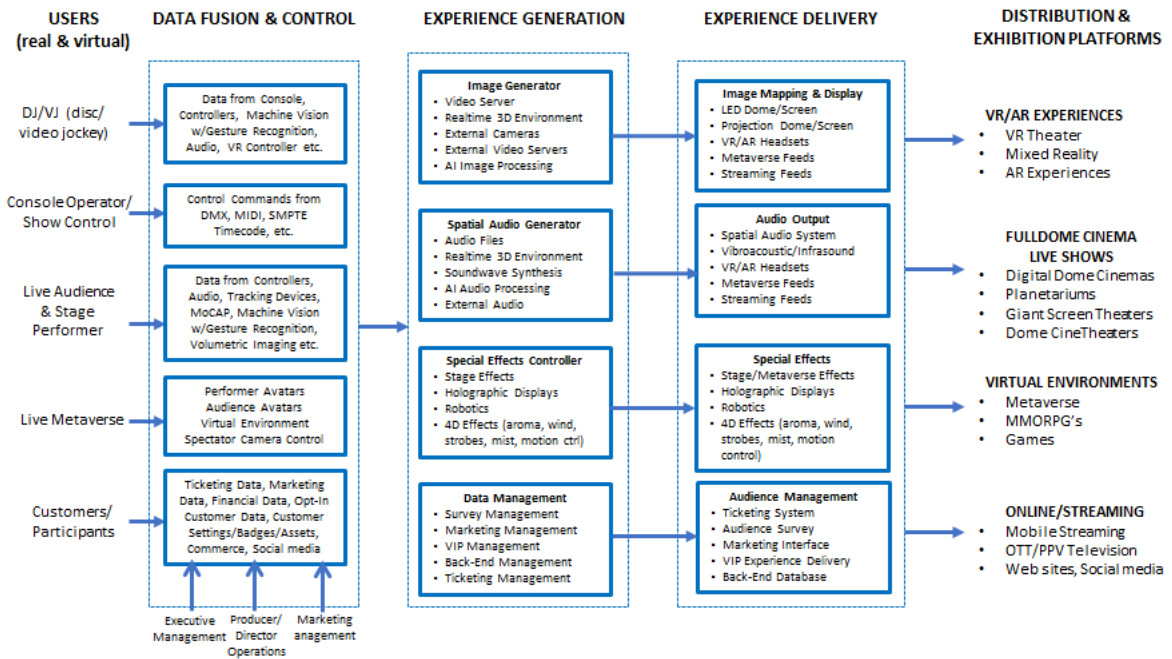


Figure 2 – Elements and data types for typical live theatrical stage performance system

6.1.2 Reference Architecture

Figure 3 provides the Reference Model of the Live Theatrical Stage Performance Use Case incorporating AI Modules (AIM's). In this diagram, data extracted from the Real and Virtual Environment (on the left) are processed and injected into the same Real and Virtual Environments (on the right).

Data is collected from both the Real and Virtual Environments including audio, video, volumetric or motion capture (mocap) data from stage performers, signals from control surfaces and more. One or more AIMs extract features from participants and performers which are output as Participant and Scene Descriptors. These Descriptors are further interpreted by Performance and Participant Status AIMs to determine the Cue Point in the show (according to the Script) and Participants Status (in general, an assessment of the audience's reactions).

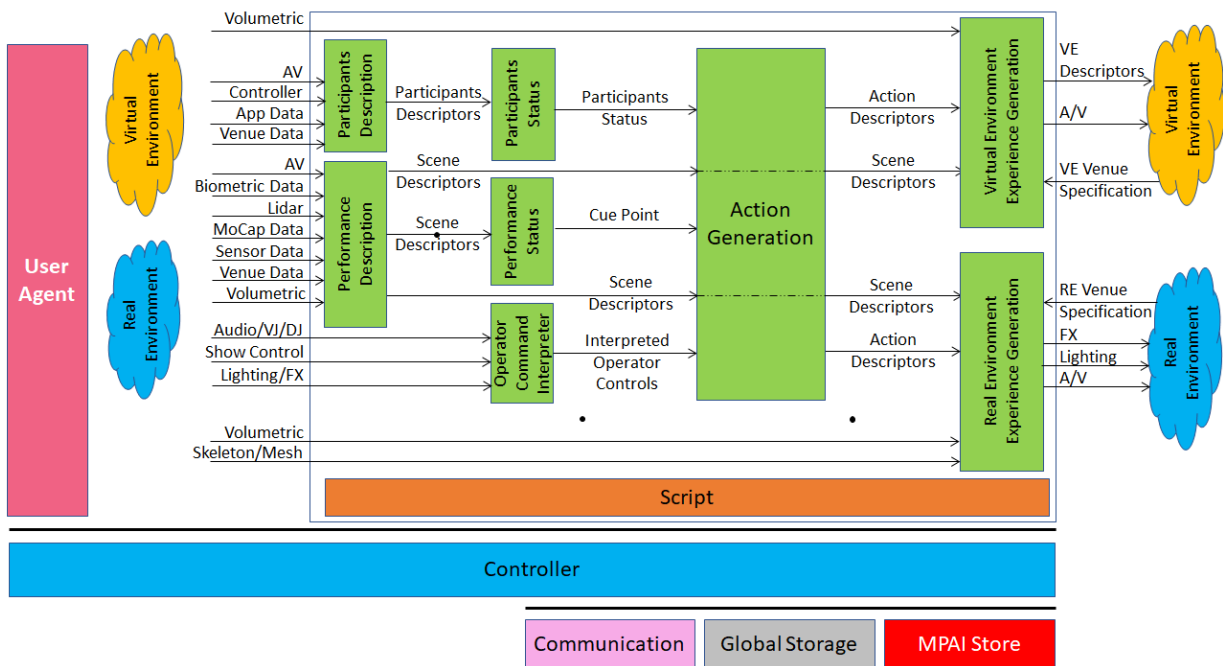


Figure 3 - Live theatrical stage performance architecture (AI Modules shown in green)

Likewise, data from the Show Control computer or control surface, consoles for audio, DJ, VJ, lighting and FX (typically commanded by operators) – if needed – are interpreted by the Operator Command Interpreter AIM. The Action Generation AIM accepts Participant Status, Cue Point and Interpreted Operator Commands and uses them to direct action in both the Real and Virtual Environments via Scene and Action Descriptors. These general descriptors are converted into actionable commands required by the Real and Virtual Environments – according to their Venue Specifications – to enable multisensory Experience Generation in both the Real and Virtual Environments. In this manner, the desired experience can automatically be adapted to a variety of real and virtual venue configurations.

6.1.3 AI Modules

Table 3 gives the list of AIMs, their functions, and their input/output data.

Table 3 - AIMs, their functions, and I/O data

AIM name	Function	Input	Output
Participants Description	Descriptors extraction	- Audio - Video - Controllers - Apps - Venue Data	- Participants Descriptors
Performance Description	Descriptors extraction Separate and track the individual (R or V) performers	- Lidar/Video of the stage - MoCap data - Volumetric data - Sensor data - Biometric data - Audio - Venue Data	- Scene Descriptors
Participants Status	Extract Participants Status from Descriptors	- Participants Descriptors	- Participants Status
Performance Status	Interpret of Scene Description including Performers, e.g., to create a digital twin and control environment	- Scene Descriptors	- Cue Points
Operator Command Interpreter	Interprets operator commands from control surfaces / computers	- Operator consoles (DJ consoles VJ, lighting and show control console / computer)	- Interpreted Operator Commands
Action Generator	Generates Action Descriptors for RE and VE based on interpretation of participants, performers, operators, and Cue Point according to the Script, and accessing stored assets	- Participants Status - Cue Point - Interpreted Operator Control - Scene Descriptors	- Action Descriptors for RE & VE. Descriptors have a generic data format.
RW Experience Generation	Creates RW multisensory experience for participants (RW/live streaming)	- Camera orientation - Multisensory Scene and Actions Descriptors.	- FX - Lighting - AV
VW Experience Generation	Creates VW multisensory experience for participants (VW)	- Camera orientation - Multisensory Scene and Actions Descriptors.	- VE Descriptors - AV

6.1.4 AIM I/O Data Formats

Table 4 records comments on the data formats.

Table 4 – Commented data formats

Data type	Comments
MoCap	e.g., Text ⁴ , Rokoko ⁵
Volumetric data	e.g., USD, FBX ⁶
Lidar	e.g., LAS ⁷ , E57 ⁸
Sensor data	
- Accelerometer	D ₂ x,D ₂ y,D ₂ z
- Positional tracker	x,y,z;a,b,c;t
- Object tracker	x,y,z;a,b,c;t
Performer behaviour	Gestures (waving, pointing, etc.), Mudra ⁹ , Dance notation (Choreography) ¹⁰ , BMN ¹¹ , DanceForms software ¹² , Gesture recognition ¹³
Operator consoles	Sliders, knobs, etc. (DMX, MIDI, NDI, Dante, ArtNet, UltraLeap, etc.), Gesture and speech recognition and interpretation
Participant Controllers	Data formats describing action of wands, control boxes on the seats, Interactive participant app or, vision system capturing audience motion.
Participants Status	Data formats expressing Participants Status composed of Emotion, Cognitive State and Attitude.
Script	Organised list of directions for all RE-VE Actions and Experiences based on time or events.
Biometric Data	Numeric data representing heart rate, electromyographic (EMG) signal or skin conductance plus metadata specifying individual performer and/or participant.
App Data	Compiled data from participants using mobile app, web interface or metaverse controls. Metadata defines data fields and app function. Fields may include personally identifying information, participant messaging, touchscreen, button or slider hits or swipes.
Venue Data	Miscellaneous data from real or virtual world.
FX	Fog, strobes, wind, mist, pyrotechnics, stage robotics, motion-base, rigging, etc.
AV (Video)	Data formats delivering pixels for dome, stage, set pieces, scrims, “hologram” effects, AR glasses, etc. (HDMI, SDI, NDI, SMPTE 2110). May also include video mixing and control data and metadata including camera or display number.

⁴ https://en.wikipedia.org/wiki/List_of_motion_and_gesture_file_formats

⁵ <https://www.rokoko.com/>

⁶ <https://www.hhi.fraunhofer.de/en/departments/vca/research-groups/multimedia-communications/research-topics/volumetric-video-formats.html>

⁷ https://en.wikipedia.org/wiki/LAS_file_format

⁸ <https://library.carleton.ca/guides/help/lidar-formats>

⁹ [https://en.wikipedia.org/wiki/List_of_mudras_\(yoga\)](https://en.wikipedia.org/wiki/List_of_mudras_(yoga))

¹⁰ https://en.wikipedia.org/wiki/Dance_notation

¹¹ https://en.wikipedia.org/wiki/Benesh_Movement_Notation

¹² <https://amt-lab.org/reviews/2020/3/lets-get-digital-visualizing-movement-in-danc>

¹³

AV (Audio)	Spatial Audio including 5.1, 7.1, 11.2, Atmos, Ambisonic, Stereo Binaural, Wavefront synthesis, etc. May also include audio mixing and control data and metadata such as speaker or microphone number.
AV (Camera orientation commands)	The commands to capture audio and video from RE or VE participants, performers, virtual characters etc. Commands include 6D camera position/orientation, zoom, focus, and aperture of any number of real or virtual cameras.

7 Technologies requested

Respondents to the Call are requested to propose Formats for some or preferably all of the Data whose functional requirements are described below.

Please note that in the following, performer, participants, and objects may be present in the Real or Virtual Environment or both. The term immersive will be used to refer to both real and virtual.

7.1 Scene Descriptors

The Scene Descriptors have the following features:

1. The descriptors describe, within both the real and virtual (immersive) environment:
 - a. Visually:
 - i. Position and skeletal orientation of immersive performers
 - ii. Position and orientation of immersive objects
 - b. Audio:
 - i. Stage/virtual environment audio (instruments, vocal, playback, etc.).
 - ii. Spatial position of source.
 - iii. Text of speech.
2. The visual scene descriptors have a format that is either existing and known (e.g., raw video, volumetric, MoCap, etc) or new.
3. The audio scene descriptors have a format that is either existing and known (e.g., Multichannel Audio, Spatial Audio, Ambisonics) or new.
4. The described performers and objects in the Real or Virtual Environment:
 - a. Have their spatial and AV component accurately described.
 - b. Be individually accessible and processable.
 - c. Have a representation that is independent of the capturing technology.
 - d. Be associated with the identity of each performer or object.
 - e. Provide a clear association of Audio objects with Visual objects.

Action Generation may use some Scene Descriptors to produce Action Descriptors.

In addition, Scene Descriptors may feed into Action Generation for modification and reformatting into data and commands that comply with the Specifications of the local venue. This allows, for instance, the creation of a digital twin in the Virtual Environment that mirrors the scene in the Real Environment.

7.2 Participant Descriptors

Are expressed by:

1. Visual behaviour of the audience (hand waving, standing, etc.)
2. Participants audio reaction (clapping, laughing, booing, etc.)
3. Audience choice (voting, motion controller, text, etc.)

7.3 Participant Status

Is expressed either with:

1. A Format supporting the semantics of a set of statuses over time.
 - a. Sentiment (e.g., measurement of spatial position-based audience reaction)
 - b. Expression of choice (e.g., voting, physical movement of audience)
 - c. Emergent behaviour (e.g., pattern emerging from coordinated movement)
2. A Language describing the status both at a time and as a trend.

7.4 Script

Script includes written Show Script including character dialog, song lyrics, stage action, etc. plus a Master Cue Sheet with a corresponding technical description of all experiential elements including sound, lighting, follow spots, set movements, cue number, and show time of cue.

The cue sheet advances to the next cue point based on quantifiable or clearly defined actions such as spoken word, gesture, etc. Various formats for show scripts and cue sheet exist, however, here we need a format for script that is machine readable and actionable.

MPI requests:

1. An extensible set of clearly defined events/criteria for triggering the cue.
2. Language for expressing Action Descriptors, which define the experiential elements associated with each cue.
3. A format for script that is machine readable and actionable.

7.5 Cue point

1. Expression of the current cue point based on interpretation of Scene Descriptors.
2. Cue points may be defined by: phrase, gesture, dance motion, prop status etc. that can be real or virtual.

7.6 Interpreted Operator Control

Interpreted Controls from manual operator consoles including:

1. Show control consoles (may include rigging, elevator stages, prop motions, and pyro).
2. Audio control consoles (controls audio mixing and effects).
3. DJ/VJ control consoles (real-time AV playback and effects).

All consoles may include sliders, buttons, knobs, gesture/haptic interfaces, joystick, touch pads.

Interpreted Controls may be Script-dependent.

7.7 Action Descriptors

Action Descriptors describe the Actions necessary to create the complete experience – in both the Real and Virtual Environments – in accordance with the script. Action Descriptors are able to express all aspects of the experience including the performers' and objects' position, orientation, gesture, costume, etc.

They can be expressed in a format that may be either existing and known (e.g., text prompts) or new.

Descriptors are generic in the sense that they are independent of the specifications of a particular venue. The Descriptors are processed by the Experience Generation AIM and translated into commands that are actionable in the Real or Virtual Environment.

7.8 Real Experience Generation

7.8.1 Lighting

Commands and data for all lighting systems, devices, and elements, typically using the DMX protocol or similar. Lighting systems may include video provided by Real Experience Generation.

7.8.2 FX (Effects)

Commands and data for all Effects generators (e.g., fog, rain, pyro, mist, etc. machines, 4D seating, stage props, rigging etc.), typically using various standard protocols. FX systems may include video and audio provided by Real Experience Generation.

7.8.3 Audio-Visual (A/V)

Data and commands for all A/V experiential elements, including audio, video, and capture cameras/microphones.

1. Camera control
 - a. Camera #n
 - b. Camera on/off
 - c. Keyframe based:
 - i. Spatial Attitude
 - ii. Optical parameters (aperture, focus, zoom)
 - iii. Frame rate
 - iv. Camera
2. Audio
 - a. M Audio channels (generated by or passed through the AIM)
 - b. Audio source location designation (channel number or spatial orientation of STEM).
 - c. MIDI device commands
 - d. Audio server commands
 - e. Mixing console commands
3. Video
 - a. N Video channels (generated by or passed through the AIM)
 - b. Video display location designation (display number or spatial orientation and mapping details).
 - c. Video server commands
 - d. VJ console commands/data
 - e. Video mixing console commands

7.8.4 Real Experience Venue specification

An input to the Real Experience Generation AIM defining protocols, data, and command structures for the specific Real Environment Venue. This could include number, type, and placement of lighting fixtures, special effects, sound, and video display resources.

7.9 Virtual Experience Generation

7.9.1 Virtual Experience Descriptors

A variety of controls for 3D geometry, shading, lighting, materials, cameras, physics, etc. may be used to affect the Virtual Environment. A protocol such as OpenUSD [11] may be used. The actual format used may depend on the current Virtual Environment Venue Specification.

7.9.2 Audio-Visual (A/V)

Data and commands for all A/V experiential elements with the virtual environment, including audio and video.

1. Audio
 - a. M Audio channels (generated by or passed through the AIM).
 - b. Placement of audio channels, including attachments to objects or characters.
 - c. Virtual Audio device commands, e.g., MIDI.
2. Video

- a. N Video channels (generated by or passed through the AIM).
- b. Placement of Video channels, including mapping to objects and characters.
- c. Virtual VJ console commands/data.
- d. Virtual Video mixing console commands.

7.9.3 Virtual Experience Venue specification

An input to the Virtual Experience Generation AIM defining protocols, and data and command structures for the specific Virtual Environment Venue.

Annex 1 - Basics about MPAI

1 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.

2 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 4 depicts the MPAI ecosystem operation for conforming MPAI implementations.

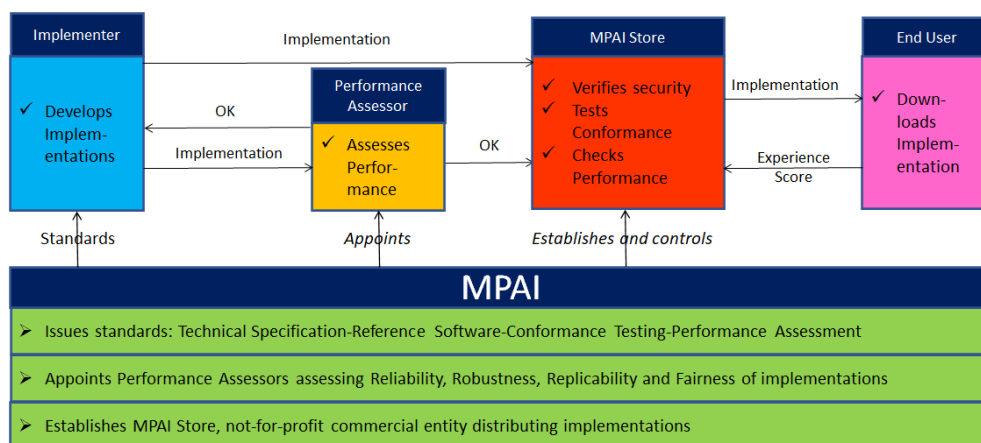


Figure 4 – The MPAI ecosystem operation

Technical Specification: Governance of the MPAI Ecosystem Table 5 identifies the following roles in the MPAI Ecosystem:

Table 5 - 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) ¹⁴ . Verifies security and Tests Implementation Conformance.
Users	Download Implementations and report their experience to MPAI.

3 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 5 depicts the MPAI-AIF Reference Model under which Implementations of MPAI Application Standards and user-defined MPAI-AIF Conforming applications operate [5].

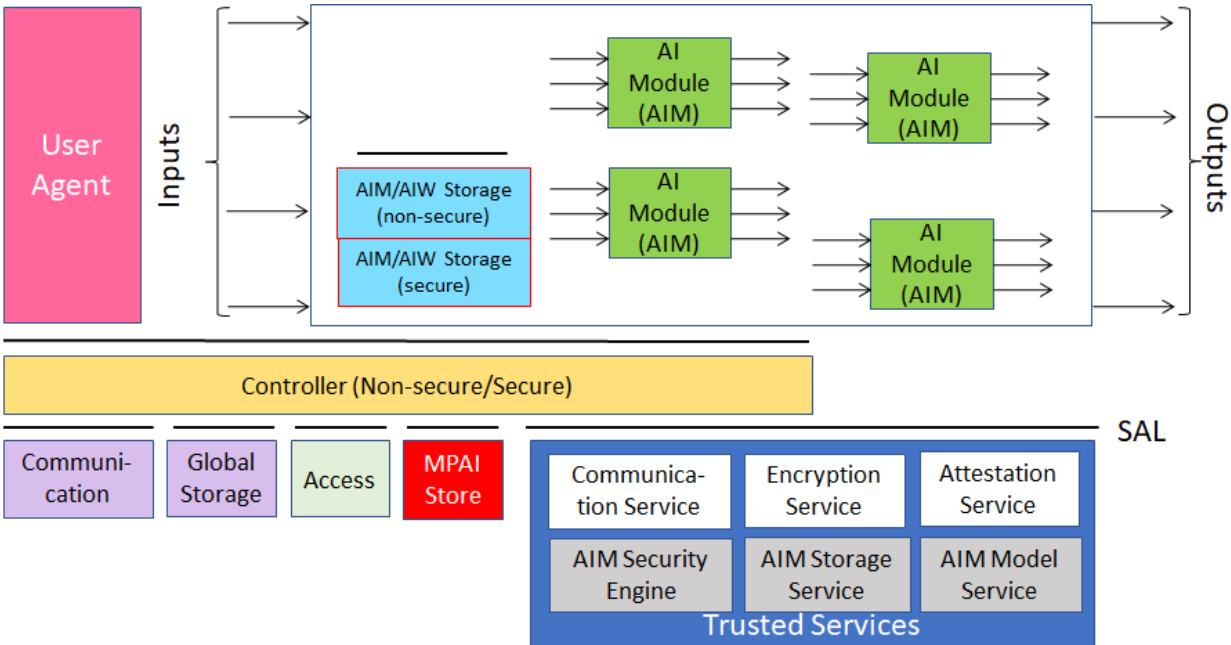


Figure 5 – The AI Framework (AIF) Reference Model

¹⁴ At the time of publication of this Technical Report, the MPAI Store was assigned as the IIDRA.

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 6 gives the characteristics of an AIW and its AIMs of a given Level:

Table 6 - MPAI Interoperability Levels

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

4 Personal Status

4.1 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.

4.2 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 6 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.

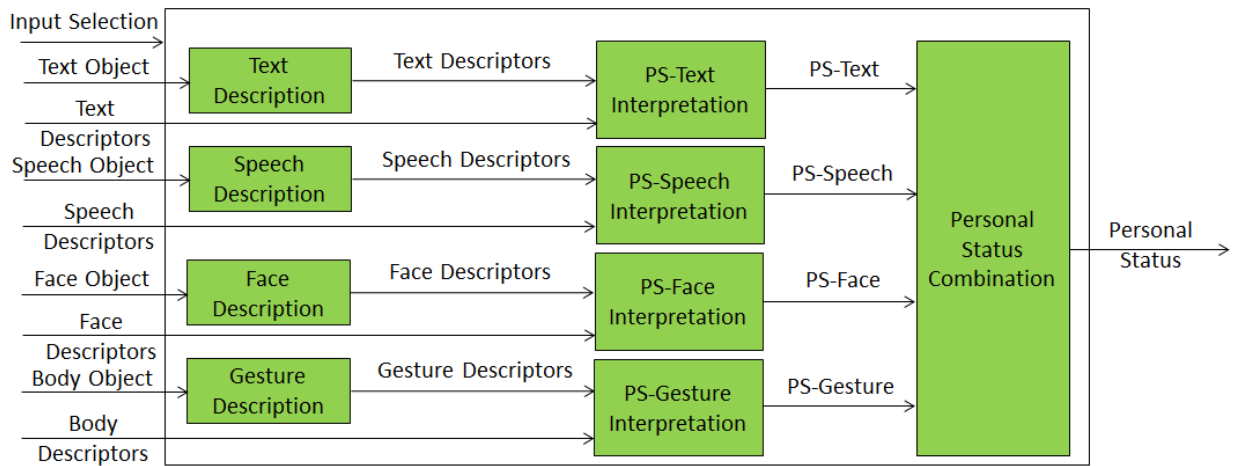


Figure 6 – Reference Model of Personal Status Extraction

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.

Annex 2 - Other MPAI-XRV Use Cases

5 eSports Tournament (XRV-EST).

5.1 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.

5.2 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
 - a. Role (e.g., magicians, warriors, soldier, etc.).
 - b. Properties (e.g., costumes, physical form, physical features).
 - c. Actions (e.g., casting spells, shooting, flying, jumping) operating in the VW
3. The VW is populated by
 - a. Avatars representing the other players.
 - b. Autonomous characters (e.g., dragon, monsters, various creatures)
 - c. Environmental structures (e.g., terrain, mountains, bodies of water).
4. The action in the VW is captured by multiple VW cameras and
 - a. Projected onto an immersive screen surrounding RW spectators
 - b. Live streamed to remote spectators as a 2D video.with all related sounds of the VW game space.
5. A shoutcaster calls the action as the game proceeds.
6. The image of RW players, player stats or other information or imagery may also be displayed on the immersive screen and the live stream.
7. The RW tournament venue is augmented with lighting and special effects, music, and costumed performers.
8. Interactions:
 - a. Live stream viewers interact with one another and with commentators through live chats, Q&A sessions, etc.
 - b. RW spectators interact through shouting, waving and interactive devices (e.g., LED wands, smartphones) through processing where:
 - i. Data are captured by camera/microphone or wireless data interface (see RW data in *Figure 1*).
 - ii. Features are extracted and interpreted.
9. RW/VW actions can be generated as a result of:
 - a. In-person or remote audience behaviour (RW).
 - b. Data collected from VW action (e.g., spell casting, characters dying, bombs exploding)
10. At the end of the tournament, an award ceremony featuring the winning players on the RW stage is held with great fanfare.

6 Experiential retail/shopping.

6.1 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.

6.2 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.

7 Collaborative immersive laboratory

7.1 Purpose

Create a collaborative immersive environment allowing 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, patient, 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.

Objective of an exemplary case: to define interfaces of AI Modules that 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.

7.2 Description

There is a file containing the digital capture of 2D slices, e.g., of the endocrine system.

An AIM reads the file and creates the 3D model of the fascia.

Another AIM finds the cells in the model and classifies them.

A human

1. navigates the 3D model.
2. interacts with the 3D model.
3. zooms in the 3D model (e.g., x2000).
4. converts a confocal image stack into a volumetric model.
5. Analyses the movement of an athlete for setting peak performance goals.

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 E57¹⁵
4. Supplemental Slides from Powerpoint/Keynote/Zoom
5. 3D Scatterplots from CSV files

7.3 Specific application areas

7.3.1 Microscopic dataset visualisation

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

7.3.2 Macroscopic dataset visualisation and simulation

1. Deals with different dataset types, e.g.:
 - a. Stars, 3D star maps (HIPPARCOS, Tycho Catalogues, etc.).

¹⁵ <https://info.vercator.com/blog/what-are-the-most-common-3d-point-cloud-file-formats-and-how-to-solve-interoperability-issues>

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

7.3.3 Educational lab

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

7.3.4 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.).

8 Immersive art experience.

8.1 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.

8.2 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.

9 DJ/VJ performance at a dance party.

9.1 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.

9.2 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 commands 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.

10 Live concert performance.

10.1 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.

10.2 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.

11 Experiential marketing/branding.

11.1 Purpose

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

11.2 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.

12 Meetings/presentations.

12.1 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.

12.2 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.

Annex 3 - MPAI-wide terms and definitions

The Terms used in this standard whose first letter is capital and are not already included in Table 2 are defined in Table 7.

Table 7 - MPAI-wide Terms

Term	Definition
Access	Static or slowly changing data that are required by an application such as domain knowledge data, data models, etc.
AI Framework (AIF)	The environment where AIWs are executed.
AI Module (AIM)	A processing element receiving AIM-specific Inputs and producing AIM-specific Outputs according to its Function. 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 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: <ol style="list-style-type: none"> 1. Implementer-specific and satisfying the MPAI-AIF Standard (<i>Level 1</i>). 2. Specified by an MPAI Application Standard (<i>Level 2</i>). 3. Specified by an MPAI Application Standard and certified by a Performance Assessor (<i>Level 3</i>).
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 their Connections in an AIW.
Reference Software	A technically correct software implementation of a Technical Specification containing source code, or source and compiled code.

Reliability	The attribute of an Implementation that performs as specified by the Application Standard, profile and version the Implementation refers to, e.g., within the application scope, stated limitations, and for the period of time specified by the Implementer.
Replicability	The attribute of an Implementation whose Performance, as Assessed by a Performance Assessor, can be replicated, within an agreed level, by another Performance Assessor.
Robustness	The attribute of an Implementation that copes with data outside of the stated application scope with an estimated degree of confidence.
Scope	The domain of applicability of an MPAI Application Standard.
Service Provider	An entrepreneur who offers an Implementation as a service (e.g., a recommendation service) to Users.
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: <ol style="list-style-type: none"> 1. The formats of the Input/Output data of the AIWs implementing the AIWs. 2. The Connections of the AIMs of the AIW. 3. The formats of the Input/Output data of the AIMs belonging to the AIW.
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.

Annex 4 - Notices and Disclaimers Concerning MPAI Standards

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

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

The life cycle of MPAI Standards

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

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

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

Coverage and Applicability of MPAI Standards

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

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

IN NO EVENT SHALL MPAI BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO: THE NEED TO PROCURE SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR

TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE PUBLICATION, USE OF, OR RELIANCE UPON ANY STANDARD, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE AND REGARDLESS OF WHETHER SUCH DAMAGE WAS FORESEEABLE.

MPAI alerts users that 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](#) and does not imply compliance with any regulatory requirements in the context of any jurisdiction. Therefore, it is the responsibility of the MPAI Standard implementer to observe or refer to the applicable regulatory requirements. By publishing an MPAI Standard, MPAI does not intend to promote actions that are not in compliance with applicable laws, and the Standard shall not be construed as doing so. In particular, users should evaluate MPAI Standards from the viewpoint of data privacy and data ownership in the context of their jurisdictions.

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

Copyright

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

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

