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

N1305	2023/07/12
Source	Requirements (CAV)
Title	Use Cases and Functional Requirements – Connected Autonomous Vehicle (MPAI-CAV) – Architecture
Target	MPAI Community

1 Introduction

MPAI, Moving Picture, Audio, and Data Coding by Artificial Intelligence – the international, unaffiliated, non-profit organisation developing standards for AI-based data coding – is publishing a Call for Technologies related to the architecture and the data exchanged by the components of the architecture.

MPAI intends to develop a Technical Specification for the architecture of a Connected Autonomous Vehicle (CAV), to be called Technical Specification – Connected Autonomous Vehicle – Architecture. MPAI defines a CAV as a system that:

1. Moves in an environment like the one depicted in *Figure 1*

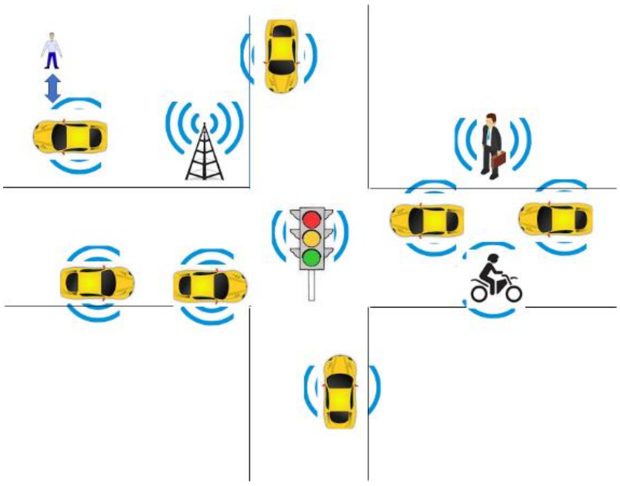

Figure 1 - An environment of CAV operation

2. Has the capability to autonomously reach a target destination by:

2.1. Understanding human utterances, e.g., the human's request to be taken to a certain location.
2.2. Planning a Route.
2.3. Sensing the external Environment and building Representations of it.
2.4. Exchanging such Representations and other Data with other CAVs and CAV-aware entities, such as, Roadside Units and Traffic Lights.
2.5. Making decisions about how to execute the Route.
2.6. Acting on the CAV motion actuation to implement the decisions.

The CAV architecture is composed of four Subsystems depicted in *Figure 2*.

1. Human-CAV Interaction (HCI).
2. Environment Sensing Subsystem (ESS)
3. Autonomous Motion Subsystem (AMS).
4. Motion Actuation Subsystem (MAS).

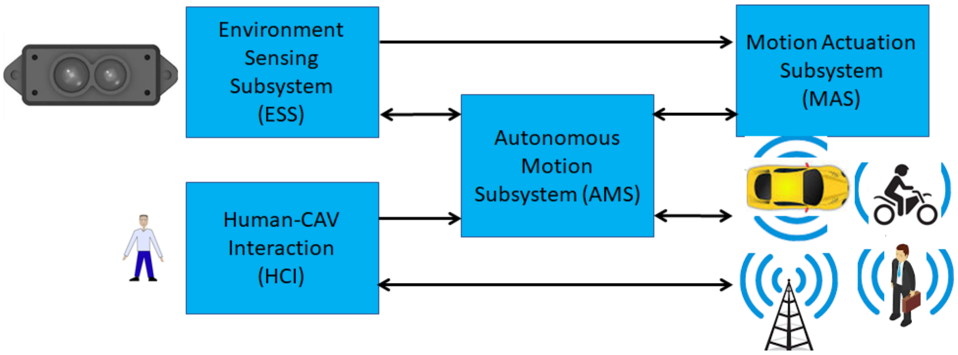
Environment Sensing Subsystem (ESS)
Motion Actuation Subsystem (MAS)
Autonomous Motion Subsystem (AMS)
Human-CAV Interaction (HCI)

Figure 2 – The CAV subsystems

MPAI does not intend to include the mechanical parts of a CAV in the planned Technical Specification: Connected Autonomous Vehicle – Architecture. MPAI only intends to refer to the interfaces of the Motion Actuation Subsystem with such mechanical parts.

The functions of the Subsystems are summarily described in *Table 1* and specified in each of the Chapters 4-5-6-7.

Table 1 – The Functions of the MPAI-CAV Subsystems

Subsystem name	**Function**
Human-CAV Interaction (HCI)	1. Recognises the humans having rights to the CAV. 2. Receives and passes to the AMS instructions about the target destination. 3. Interacts with humans by assuming the shape of an avatar. 4. Activate other Subsystems as required by humans. 5. Provides the Full Environment Representation received from the AMS for passengers to use.
Environment Sensing Subsystem (ESS)	1. Acquires and processes information from the Environment. 2. Produces the Basic Environment Representation 3. Sends the Basic Environment Representation to the AMS.
Autonomous Motion Subsystem (AMS)	1. Computes the Route to destination based on information received from the HCI. 2. Receives the Basic Environment Representation of the ESS and of other CAVs in range. 3. Creates the Full Environment Representation. 4. Issues commands to the MAS to drive the CAV to the intended destination.
Motion Actuation Subsystem (MAS)	1. Sends its Spatial Attitude and other Environment information to the ESS. 2. Receives/actuates motion commands in the Environment. 3. Sends feedback to the AMS

The following high-level workflow illustrates a CAV operation example and the role of CAV Subsystems:

1. A *human* with appropriate credentials requests the CAV, via *Human-CAV Interaction*, to take the human to a given Pose.
2. *Human-CAV Interaction* authenticates the human, interprets the request, communicates with the HCIs of other CAVs on matters that directly impact the human passengers, and passes

commands to the *Autonomous Motion Subsystem*. The human may subsequently integrate/correct their instructions.

3. *Autonomous Motion Subsystem*:
    a. Requests *Environment Sensing Subsystem* to provide the current Pose.
    b. Computes the Route and may offer options to authenticated humans.
4. *Environment Sensing Subsystem* computes and sends the Basic Environment Representation to the *Autonomous Motion Subsystem.*
5. *Autonomous Motion Subsystem*:
    a. Receives the Basic Environment Representations from the Environment Sensing Subsystem
    b. Exchanges the Basic Environment Representation with other CAVs and computes the Full Environment Representation.
    c. Makes decision on how to best move the CAV to reach the destination, e.g., by avoiding a car suddenly appearing on the horizon.
    d. Issues appropriate commands to the Motion Actuation Subsystem.
6. While the CAV moves, the humans in the cabin may:
    a. Interact and hold conversation with other humans on board and the Human-CAV Interaction Subsystem.
    b. Issue commands.
    c. Request the Full Environment Representation to render the environment.
    d. Interact with (humans in) other CAVs.

MPAI assumes that each of the four Subsystems of a CAV is an implementation of MPAI Technical Specification: AI Framework (MPAI-AIF) V2 [2]. A AI Framework (AIF) V2 executes an AI Workflow composed of AI Modules in a secure environment. Annex 3 - Chapter 1 provides a concise description of the AI Framework.

Each of the four Chapters 4-5-6-7 addresses a *Subsystem* (corresponding to an AI Workflow of Annex 3 - Chapter 1) providing the following:

1. The Function of the Subsystem.
2. The input/output data of the Subsystem.
3. The topology of the Components (AI Modules) of the Subsystem.
4. For each AI Module of the Subsystem:

4.1. The Function.
4.2. The input/output data.

A fifth Chapter includes the elements of the so-called Communication Device enabling a CAV to communicate with other CAVs.

Note that this document:

1. Does not make any assumption regarding the Location carrying out the processing required by Subsystem or AI Modules.
2. Assumes that information processing, collection, and storage is performed according to the laws of the Location.

This Technical Report has been developed by the Connected Autonomous Vehicles group of the Requirements Standing Committee. MPAI may publish more versions of this Technical Report and intends to publish a Technical Specification where the AIM and AIW I/O Data Formats will all be specified.

2 Terms and definitions

Table 2 defines the terms used in this document. The general MPAI Terms are defined in *Table 16*.

Table 2 – Terms and Definitions

Term	Definition
Accelerometer Data	Data related to the acceleration forces acting on a CAV produced by the electronic sensor accelerometer.
Alert	Elements in an Environment Representation that should be treated with priority by the Obstacle Avoider AIM.
AMS-MAS Command	The AMS Command instructing the Motion Actuation Subsystem to change the Ego CAV's Spatial Attitude SA_A at time t_A to Spatial Attitude SA_B at time t_B.
AMS-HCI Response	Response generated by the Motion Actuation Subsystem during and after the execution of an HCI-AMS Command.
AMS-MAS Command	A Command issued by the AMS to the MAS designed to drive the CAV to reach a Goal.
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.
Audio Data	The serialised output of a microphone array capturing the target Environment to create the Audio Scene Description used to incorporate Environment Audio information in the Basic and Full Environment Representation.
Audio Object	Digital Representation of Audio information with its metadata.
Audio Scene	The Audio Objects of an Environment with Spatial Object metadata.
Audio Scene Descriptors	Descriptors enabling the description of the outdoor and indoor sound field in terms of individually Identified Audio Objects with a Spatial Attitude.
Audio-Visual Object	Coded representation of Audio-Visual information with its metadata. An Audio-Visual Object can be a combination of Audio-Visual Objects.
Audio-Visual Scene (AV Scene)	The Audio-Visual Objects of an Environment with Object Spatial Attitude.
Audio-Visual Scene Descriptors	Descriptors enabling the description of the outdoor and indoor Audio-Visual Scene in terms of Audio-Visual Objects having a common time-base, associating co-located audio and visual objects if both are available, and supporting the physical displacement and interrelation (e.g., occlusion) of Audio and Visual Objects over time.
Avatar	An animated 3D object representing a real or fictitious person in a virtual space rendered to a physical space.
Avatar Model	The Model of a human that a user selects to impersonate the CAV's HCI as rendered by the Personal Status Display AIM.
Basic Environment Representation (BER)	A Digital Representation of the Environment that integrates the Ego CAV's Spatial Attitude, the Scene Descriptions produced by the available Environment Sensing Technology-specific Road Topology, and Other Environment Data.
Body Descriptors	Descriptors representing the motion and conveying information on the Personal Status of the body of a human or an avatar.
Brakes Command	The result of the interpretation of AMS-MAS Command to the Brakes.

Brakes Response	The Response of Brakes to the AMS Command Interpreter.
Camera Data	Serialised data provided by a variety of sensor configurations operating in the visible frequency range.
Camera Scene Descriptors	Descriptors produced by the Camera Scene Description AIM using Camera Data and previous Basic Environment Representations.
CAV Centre	The point in the CAV selected as represented by coordinates (0,0,0).
CAV Identifier	A code uniquely identifying a CAV carrying information, such as Country where the CAV has been registered, Registration number in that country, CAV manufacturer identifier, CAV model identifier.
Cognitive State	An element of the internal status of a human or avatar reflecting their understanding of the Environment, such as "Confused" or "Dubious" or "Convinced".
Connected Autonomous Vehicle	A vehicle able to autonomously reach a Pose by: 1. Understanding human utterances in the Subsystem (HCI). 2. Planning a Route (AMS). 3. Sensing and building a Representations of the external Environment (ESS). 4. Exchanging such Representations and other Data with other CAVs and CAV-aware entities (AMS). 5. Making decisions about how to execute the Route (AMS). 6. Acting on the MAS.
Data	The digital representation of information.
Data Format	The standard digital representation of Data.
Decision Horizon	The time within which a decision is assumed will be implemented.
Descriptor	Coded representation of a feature of text, audio, speech, or visual.
Digital Representation	A data structure corresponding to a physical entity.
Ego CAV	The object in the representation of an environment that the CAV recognises itself.
Emotion	An element of the internal status of a human or avatar resulting from their interaction with the Environment or subsets of it, such as "Angry", and "Sad".
Environment	The portion of the real world of current interest to the CAV.
Environment Representation	A digital representation of the Environment produced by an Environment Sensing Technology in CAV.
Environment Sensing Technology (EST)	One of the technologies used to sense the Environment by the Environment Sensing Subsystem, e.g., RADAR, Lidar, Video, Ultrasound, and Audio, The Offline Map is considered as an EST.
Face	The portion of a 2D or 3D digital representation corresponding to the face of a human.
Face Descriptors	Descriptors representing the motion and conveying information on the Personal Status of the face of a human or an avatar.
Face ID	The Identifier of a human belonging to a group of humans inferred from analysing the face of the human.
Factor	One of Cognitive State, Emotion, and Social Attitude
Full Environment Representation (FER)	A digital representation of the Environment using the Basic Environment Representations of the ego CAV and of other CAVs in range or Roadside Units.

Full Environment Representation Audio	The A output of the Full Environment Representation Viewer.
Full Environment Representation Commands	Commands issued by a CAV passenger to the HCI to enable navigation in the Full Environment Representation, e.g., select a Point of View, zoom in/out, control sound level.
Full Environment Representation Visual	The Visual output of the Full Environment Representation Viewer.
Gesture	A movement of the body or part of it, such as the head, arm, hand, and finger, often a complement to a vocal utterance.
Global Navigation Satellite System (GNSS)	One of the systems providing global navigation information such as GPS, Galileo, Glonass, BeiDou, Quasi Zenith Satellite System (QZSS) and Indian Regional Navigation Satellite System (IRNSS).
Goal	The Spatial Attitude planned to be reached at the end of a Decision Horizon.
HCI-AMS Command	High-level instructions issued by HCI to AMS to instruct it to reach a final destination or messages that HCI has received from the HCI of other CAVs.
Identifier	A Label that is uniquely associated with a human, an avatar, or an object.
Inertial Measurement Unit	An inertial positioning device, e.g., odometer, accelerometer, speedometer, gyroscope etc.
Instance ID	Instance of a class of Objects and the Group of Objects the Instance belongs to.
LiDAR Data	Serialised data provided by a LiDAR sensor, an active time-of-flight sensor operating in the µm range – ultraviolet, visible, or near infrared light (900 to 1550 nm).
LiDAR Scene Descriptors	Descriptors produced by the LiDAR Scene Description AIM using LiDAR Data and previous Basic Environment Representations.
Machine Avatar	The rendered face and body of an Avatar produced by the Personal Status Display
Machine Speech	The rendered synthetic speech generated by the Personal Status Display.
Map Scene Descriptors	Descriptors produced by the Map Scene Description AIM using Offline Map Data and previous Basic Environment Representations.
MAS-AMS Response	The Response of AMS Command Interpreter integrating the Response from Brakes, Wheel Directions, and Wheel Motors. The MAS-AMS Responses contain the value of a Spatial Attitude’s at an intermediate Pose with the corresponding Time.
Meaning	Information extracted from an input text such as syntactic and semantic information.
Microphone Array Geometry	Audio Data captured by an array of microphones arranged as specified by the Microphone Array Geometry and providing sensing characteristics of the microphone(s) used (e.g., cardioid), sampling frequency, number of bits/sample etc.
Modality	One of Text, Speech, Face, or Gesture.
Model	A Data Format representing an object with their features ready to be animated.

Object	An Object is a data structure representing an object sensed by an EST and produced by an EST-specific Scene Description. Elements characterising and object are: 1. Timestamp. 2. Identifier of the Scene Description AIM that has generated the Object. 3. Alerts 4. Spatial Attitude of the Object and its estimated accuracy measured from the CAV Centre. 5. Bounding box. 6. Object type (2D, 2.5D, and 3D):
Object ID	The Identifier uniquely associated with a particular class of Objects, e.g., hammer, screwdriver, etc.
Odometer Data	The distance from the start up to the current Pose measured by the number of wheel rotations times the tire circumference (π x tire diameter).
Offline Map	A previously created digital map of an Environment and associated metadata.
Offline Map Data	Data provided by an Offline Map in response to a given set of coordinate values.
Orientation	The set of the 3 roll, pitch, yaw angles indicating the rotation around the principal axis (x) of an Object, its y axis having an angle of 90˚ counter-clockwise (right-to-left) with the x axis and its z axis (perpendicular to and out of the ground). See Figure 5.
Other Environment Data	Additional Data acquired by the Motion Actuation Subsystem and complementing the spatial data such as weather, temperature, air pressure, humidity, ice and water on the road, wind, fog etc.
Path	A sequence of Poses $pi = (x_i, y_i, z_i, \alpha_i, \beta_i, \gamma_i)$.
Personal Status	The ensemble of information internal to a person expressed by 3 Factors (Cognitive State, Emotion, Social Attitude) conveyed by one or more Modalities (Text, Speech, Face, and Gesture Modalities).
Pose	Position and Orientation of the CAV.
Position	The current coordinates of a CAV as obtained from the CAV's sensors.
RADAR Data	Serialised data provided by a RADAR sensor, an active time-of-flight sensor operating in the 24-81 GHz range.
RADAR Scene Descriptors	Descriptors produced by the RADAR Scene Description AIM using RADAR Data and previous Basic Environment Representations.
Road State	Data about the state of the road the CAV is traversing inferred by the AMS from internally available information or received from an external source via a communication channel such as detours and road conditions.
Road Topology	A data structure containing the Position of the Road Signs (Traffic Poles, Road Signs, Traffic Lights) and a Taxonomy-based semantics of the Road Signs.
Roadside Unit	A wireless communicating device located on the roadside providing information to CAVs in range.
Route	A sequence of Way Points.
Scene Description	The organised collection of Descriptors that enable an object-based description of a scene.
Scene Description Format	The combination of EST-specific 2D, 2.5D, or 3D Scene Descriptors used by an EST Scene Description in an EST-specific time window.

Scene Descriptors	The individual attributes of the coded representation of the objects in a scene, including their location.
Shape	The digital representation of the volume occupied by a CAV.
Social Attitude	An element of the internal status of a human or avatar related to the way they intend to position themselves vis-à-vis the Environment or subsets of it, e.g., “Confrontational”, “Respectful”.
Spatial Attitude	CAV’s Position, Orientations and their velocities and accelerations at a given time.
Speaker ID	The Identifier of a human belonging to a group of humans inferred from analysing the speech of the human.
Speech	Digital representation of analogue speech sampled at a frequency between 8 kHz and 96 kHz with 8, 16 and 24 bits/sample, and non-linear and linear quantisation.
Speech Model	The collection of Speech Descriptors characteristic of a speaker used to generate the synthetic speech of the Personal Status Display.
Speedometer Data	The speed of the CAV as measured by the electronic sensor that measures the instantaneous speed of the CAV.
Subsystem	One of the 4 components making up the CAV.
Text	A series of characters drawn from a finite alphabet represented using a Character Set.
Text (Language Understanding)	Text resulting from the refinement of Text produced by a Speech Recognition AIM by the Language Understanding AIM.
Traffic Rules	The digital representation of the traffic rules applying to an Environment as extracted from the local Traffic Signals based on the local traffic rules.
Traffic Signals	The digital representations of the traffic signals on a road and around it, their Spatial Attributes, and the semantics of the traffic signals.
Trajectory	A sequence of Spatial Attitudes s_i ($s_1,s_2,\ldots s_i$) and the expected time each Spatial Attitude will be reached.
Ultrasound Data	Serialised data provided by an ultrasonic sensor, an active time-of-flight sensor typically operating in the 40 kHz to 250 kHz range, measuring the distance between objects within close range.
Ultrasound Scene Descriptors	Descriptors produced by the Ultrasound Scene Description AIM using Ultrasound Data and previous Basic Environment Representations.
Video	Data generated by a camera.
Viewpoint	The Spatial Attitude of a user looking at the Environment.
Visual Object	Coded representation of Visual information with its metadata.
Visual Scene	The Visual Objects of an Environment with Spatial Object metadata.
Visual Scene Descriptors	Descriptors enabling the description of the outdoor and indoor visual scene in terms of individually Identified Visual Objects with a Spatial Attitude.
Waypoint	A point w_i on an Offline Map.
Wheel Direction Command	The result of the interpretation of AMS-MAS Command to the Wheel Direction.
Wheel Direction Feedback	The Response of Wheel Direction to the AMS Command Interpreter.
Wheel Motor Command	The result of the interpretation of AMS-MAS Command to the Wheel Motor.
Wheel Motor Response	The Response of the Wheel Motor to AMS Command Interpreter.

3 References

1. MPAI; Technical Specification: Governance of the MPAI Ecosystem (MPAI-GME) V1.1; https://mpai.community/standards/mpai-gme/.
2. MPAI; Technical Specification: AI Framework (MPAI-AIF) V2; https://mpai.community/standards/mpai-aif/.
3. MPAI; Technical Specification: Avatar Representation and Animation (MPAI-ARA) V1; https://mpai.community/standards/mpai-ara/.
4. MPAI; Technical Specification: Multimodal Conversation (MPAI-MMC) V2; https://mpai.community/standards/mpai-mmc/.
5. Technical Specification: Context-based Audio Enhancement (MPAI-CAE) V2, https://mpai.community/standards/mpai-cae/.
6. Universal Coded Character Set (UCS): ISO/IEC 10646; December 2020
7. ISO/IEC 14496-10; Information technology – Coding of audio-visual objects – Part 10: Advanced Video Coding.
8. ISO/IEC 23008-2; Information technology – High efficiency coding and media delivery in heterogeneous environments – Part 2: High Efficiency Video Coding.
9. ISO/IEC 23094-1; Information technology – General video coding – Part 1: Essential Video Coding.
10. ISO 8855:2011 – Road vehicles – Vehicle dynamics and road-holding ability – Vocabulary
11. SAE; Vehicle Dynamics Terminology J670_202206; https://www.sae.org/standards/content/j670_202206/
12. SAE; Levels of Driving Automation J3016; https://www.sae.org/binaries/content/assets/cm/content/blog/sae-j3016-visual-chart_5.3.21.pdf

4 Human-CAV Interaction (HCI)

4.1 Functions of Subsystem

The Human-CAV Interaction (HCI) Subsystem performs the following high-level functions:

1. Authenticates humans e.g., for the purpose of letting them into the CAV.
2. Interprets and executes commands provided by humans, possibly after a dialogue, e.g., to go to a Waypoint, issue commands such as turn off air conditioning, open window, call a person, search for information, etc.
3. Displays Full Environment Representation to passengers via a viewer and allows passengers to control the display.
4. Interprets conversation utterances with the support of the extracted Personal Statuses of the humans, e.g., on the fastest way to reach a Waypoint because of an emergency, or during a casual conversation.
5. Displays itself as a Body and Face with a mouth uttering Speech showing a Personal Status comparable to the Personal Status that a human counterpart (e.g., driver, tour guide, interpreter) would display in similar circumstances.

The HCI operation is highly influenced by the notion of *Personal Status*, the set of internal characteristics of conversation humans and machines. See Annex 1 Section 1.

Reference Architecture of Subsystem

Reference Architecture of Subsystem

4.2 Reference Architecture of Subsystem

Figure 3 gives the Human-CAV Interaction (HCI) Reference Model supporting the case of a group of humans approaching the CAV from outside the CAV and sitting inside the CAV.

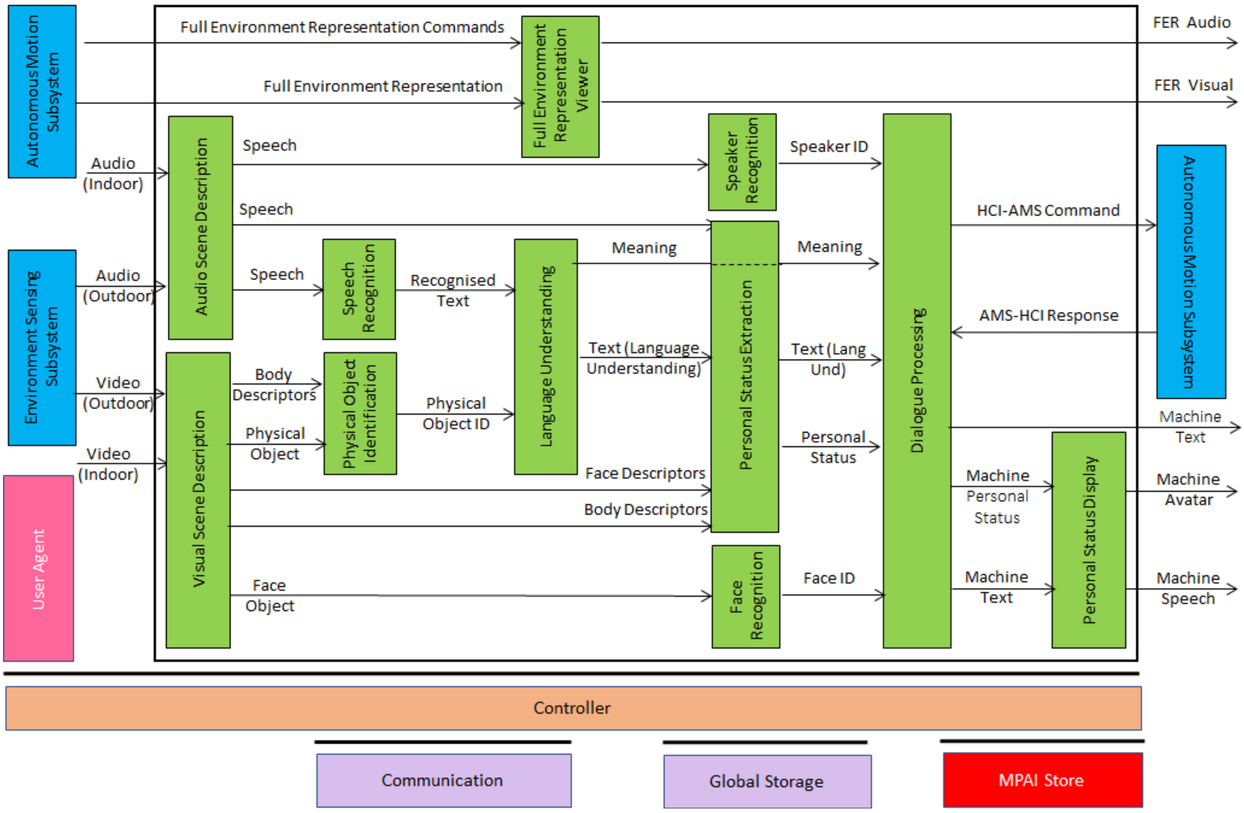

Figure 3 – Human-CAV Interaction Reference Model

The HCI operation is considered in two outdoor and indoor human-CAV interaction scenarios:

1. When a group of humans approaches the CAV from outside the CAV:
   a. The Audio Scene Description AIM creates the Audio Scene Descriptions in the form of Audio (Speech) Objects corresponding to each speaking human in the Environment (close to the CAV).
   b. The Visual Scene Description creates the Visual Scene Description and provides 1) the Face and Physical Objects and 2) the Body and Face Descriptors corresponding to each human in the Environment (close to the CAV).
   c. The Speaker Recognition and Face Recognition AIMs authenticate the humans the HCI is interacting with.
   d. The Speech Recognition AIM recognises the speech of each human.
   e. The Language Understanding AIM produces the refined Text (Language Understanding) and extracts the Meaning.
   f. The Personal Status Extraction AIM extracts the Personal Status of the humans from 1) Speech, 2) Face and Body Descriptors, 3) Text (Language Understanding) and 4) Meaning.
   g. The Dialogue Processing AIM 1) validates the human Identities, 2) responds to human utterances, 3) displays the Face and Body of the HCI Personal Status, and 4) issues commands to the Autonomous Motion Subsystem.
2. When a group of humans sit inside the CAV:
   a. The Audio Scene Description AIM creates the Audio Scene Descriptions in the form of Audio (Speech) Objects corresponding to each speaking human in the cabin.

b. The Visual Scene Description creates the Visual Scene Descriptors in the form of Face Descriptors corresponding to each human in the cabin.
c. The Speaker Recognition and Face Recognition AIMs identify the humans the HCI is interacting with.
d. The Speech Recognition AIM recognises the speech of each human.
e. The Language Understanding AIM extracts the Meaning and produces the refined Text (Language Understanding).
f. The Personal Status Extraction AIM extracts the Personal Status of the humans.
g. The Dialogue Processing AIM validates the human Identities, responds to human utterances, displays the HCI Personal Status, and issues commands to the Autonomous Motion Subsystem.

Notes related to the two scenarios:

1. HCI interacts with the humans sitting in the cabin in two ways:
   h. By responding to commands/queries from one or more humans at the same time, e.g.:
      i. Commands to go to or park at a Waypoint, etc.
      ii. Commands with an effect on the cabin, e.g., turn off air conditioning, turn on the radio, call a person, open window or door, search for information etc.
   i. By conversing with and responding to questions from one or more humans at the same time about travel-related issues (in-depth domain-specific conversation), e.g.:
      i. Humans request information, e.g., time to destination, route conditions, weather at destination, etc.
      ii. Humans ask questions about objects in the cabin or held by humans.
      iii. CAV offers alternatives to humans, e.g., long but safe way, short but likely to have interruptions.
   j. By following the conversation on travel matters held by humans in the cabin if:
      i. The passengers allow the HCI to follow the conversation, and
      ii. The processing is carried out inside the CAV and is held confidential.
2. While in the cabin, passengers can become aware of the external Environment by issuing Full Environment Representation (FER) Commands to navigate the Full Environment Representation.
3. When conversing with the humans in the cabin, the HCI displays itself as a speaking avatar via the Personal Status Display AI Module.

4.3 Input/Output Data of Subsystem

Table 3 gives the input/output data of the Human-CAV Interaction Subsystem.

Table 3 – I/O data of Human-CAV Interaction

Input data	From	Comment
Audio (ESS)	Environment Sensing Subsystem	User authentication User command User conversation
Audio	Cabin Passengers	User’s social life Commands/interaction with HCI
Video (ESS)	Environment Sensing Subsystem	Commands/interaction with HCI
Video	Cabin Passengers	User’s social life Commands/interaction with HCI
Full Environment	Autonomous Motion	Rendered by Full Environment

Representation	Subsystem	Representation Viewers
Full Environment Representation Commands	Cabin Passengers	To control rendering of Full Environment Representation
Output data	**To**	**Comments**
Output Speech	Humans in Environment Cabin Passengers	HCI’s response to passengers
Output Face	Cabin Passengers	HCI’s face when conversing
Output Body	Cabin Passengers	HCI’s body when conversing
Output Text	Cabin Passengers	HCI’s response to passengers
Full Environment Representation Audio	Passenger Cabin	For passengers to hear external Environment
Full Environment Representation Video	Passenger Cabin	For passengers to view external Environment

4.4 Functions of the AI Modules

Table 4 gives the functions of all Environment Sensing Subsystem AIMs.

Table 4 – AI Modules of the Environment Sensing Subsystem

AIM	Function
Audio Scene Description	Produces the Audio Scene Descriptors using the Audio captured by the appropriate (indoor or outdoor) Microphone Array.
Visual Scene Description	Produces the Visual Scene Descriptors using the visual information captured by the appropriate (indoor or outdoor) visual sensors.
Speech Recognition	Converts speech into Text.
Physical Object Identification	Provides the ID of the class of objects of which the Physical Object is an Instance
Full Environment Representation Viewer	Converts the Full Environment Representation produced by the Autonomous Motion Subsystem into Audio-Visual Scene Descriptors that can be perceptibly rendered.
Language Understanding	Improves the Text from Speech Recognition by using context information (e.g., Instance ID of object).
Speaker Recognition	Provides Speaker ID from Speech.
Personal Status Extraction	Provides the Personal Status of human.
Face Recognition	Provides Face ID from Face.
Dialogue Processing	Provides: 1. Text containing the response of the HCI to the human. 2. Personal Status of HCI congruous with the Text produced by the HCI.
Personal Status Display	Produces Speech, and Machine Face and Body.

4.5 Input/Output Data of AI Modules

Table 5 gives the input/output data of the Human-CAV Interaction AIMs.

Table 5 – AI Modules of Human-CAV Interaction

AIM	Input	Output
Audio Scene Description	Environment Audio (outdoor) Environment Audio (indoor)	Speech Objects
Visual Scene Description	Environment Video (outdoor) Environment Video (indoor)	Face Objects Physical Objects Body Descriptors Face Descriptors
Speech Recognition	Speech Object	Recognised Text
Physical Object Identification	Physical Object Human Object	Object ID
Full Environment Representation Viewer	FER Commands	FER Audio FER Visual
Language Understanding	Recognised Text Personal Status Object ID	Meaning Personal Status Text (Language Understanding)
Speaker Recognition	Speech Descriptors	Speaker ID
Personal Status Extraction	Recognised Text Speech Object Face Object Human Object	Personal Status
Face Recognition	Face Object	Face ID
Dialogue Processing	Speaker ID Meaning Text (Language Understanding) Personal Status Face ID AMS-HCI Response	AMS-HCI Commands Output Text Output Personal Status
Personal Status Display	Machine Text Output Personal Status	Machine Avatar Machine Text Machine Speech

5 Environment Sensing Subsystem (ESS)

5.1 Functions of Subsystem

The Environment Sensing Subsystem (ESS):

1. Uses all Subsystem devices to acquire as much as possible information from the Environment – electromagnetic and acoustic data.
2. Receives an initial estimate of the Ego CAV's Spatial Attitude and Environment Data (e.g., temperature, pressure, humidity, etc.) from the Motion Actuation Subsystem.
3. Produces a sequence of Basic Environment Representations (BER) for the duration of the travel.
4. Passes the Basic Environment Representations to the Autonomous Motion Subsystem.

5.2 Reference Architecture of Subsystem

Figure 4 gives the Environment Sensing Subsystem Reference Model.

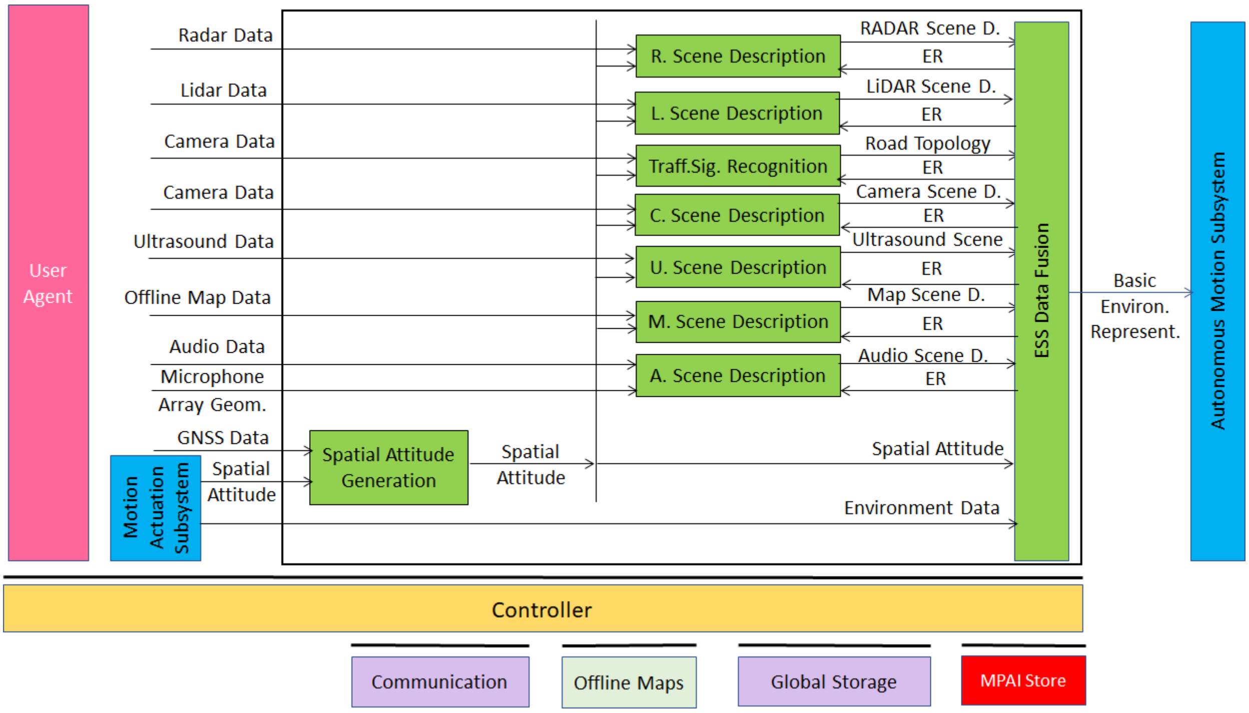

Figure 4 – Environment Sensing Subsystem Reference Model

The typical sequence of operations of the Environment Sensing Subsystem AIM is:

1. Compute the CAV's Spatial Attitude using the initial Spatial Attitude provided by the Motion Actuation Subsystem and the GNSS.
2. Produce Environment Sensing Technology (EST) Scene Descriptors, e.g., the RADAR Scene Descriptors, using Snapshots of information provided by an EST e.g., RADAR Data.
3. Integrate the Scene Descriptors from different Environment Sensing Technologies into the time-dependent Basic Environment Representation.

Figure 4 assumes that Traffic Signalisation Recognition produces the Road Topology by analysing Camera Data. The model of *Figure 4* can easily by extended to the case where Data from other ESTs is processed to compute or help compute the Road Topology.

5.3 Input/Output Data of Subsystem

The currently considered Environment Sensing Technologies (EST) are:

1. Global navigation satellite system or GNSS (~1 & 1.5 GHz Radio).
2. Geographical position and orientation, and their time derivatives up to 2nd order (Spatial Attitude).
3. Visual Data in the visible range, possibly supplemented by depth information (400 to 700 THz).
4. LiDAR Data (~200 THz infrared).
5. RaDAR Data (~25 & 75 GHz).
6. Ultrasound Data (> 20 kHz).
7. Audio Data in the audible range (16 Hz to 16 kHz).
8. Spatial Attitude (from the Motion Actuation Subsystem).
9. Other environmental data (temperature, humidity, ...).

Table 6 gives the input/output data of Environment Sensing Subsystem.

Table 6 – I/O data of Environment Sensing Subsystem

Input data	From	Comment
Spatial Attitude from AMS	Motion Actuation Subsystem	To be fused with GNSS data
Other Environment Data	Motion Actuation Subsystem	Temperature etc. to be added to Basic Environment Representation
Global Navigation Satellite System (GNSS)	~1 & 1.5 GHz Radio	Get Pose from GNSS
Radar	~25 & 75 GHz Radio	Capture Environment with Radar
Lidar	~200 THz infrared	Capture Environment with Lidar
Ultrasound	Audio (>20 kHz)	Capture Environment with Ultrasound
Cameras (2/D and 3D)	Video (400-800 THz)	Capture Environment with Cameras
Microphones	Audio (16 Hz-16 kHz)	Capture Environment with Microphones
Output data	**To**	**Comment**
Alert	Autonomous Motion Subsystem	Critical last minute Environment Description from EST
Basic Environment Representation	Autonomous Motion Subsystem	Locate CAV in the Environment

5.4 Functions of AI Modules

Table 7 gives the functionality of all Environment Sensing Subsystem AIMs.

Table 7 – AI Modules of the *Environment Sensing Subsystem*

AIM	Function
Spatial Attitude Generation	Computes the CAV Spatial Attitude using information received from GNSS and Motion Actuation Subsystem with respect to a predetermined point in the CAV defined as the origin (0,0,0) of a set of (x,y,z) Cartesian coordinates with respect to the local coordinates.
RADAR Scene Description	Produces RADAR Scene Descriptors from RADAR Data
LiDAR Scene Description	Produces LiDAR Scene Descriptors from LiDAR Data
Camera Scene Description	Produces Camera Scene Descriptors from Camera Data
Traffic Signalisation Recognition	Produces Road Topology of the Environment from Camera and LiDAR Data.
Ultrasound Scene Description	Produces Ultrasound Scene Descriptors from Ultrasound Data.
Audio Scene Description	Produces Audio Scene Descriptors from Audio Data.
Online Map Scene Description	Produces Online Map Data Scene Descriptors from Online Map Data.
Environment Sensing Subsystem Data Fusion	Selects critical Environment Representation as Alert; produces CAV’s Basic Environment Representation by fusing the Scene Descriptors of the different ESTs, The Basic Environment Representation (BER) includes all available information from ESS and MAS that enables the

	CAV to define a Path in the Decision Horizon Time. The BER results from the *integration* of: 1. The different Scene Descriptors generated by the different EST-specific Scene Description AIMs. 2. Environmental data. 3. The Spatial Attitude of the Ego CAV as estimated by the Motion Actuation Subsystem.

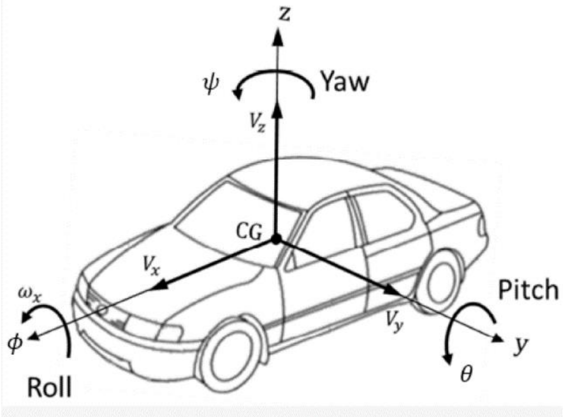

Figure 5 - Roll, Pitch, and Yaw in a vehicle [10]

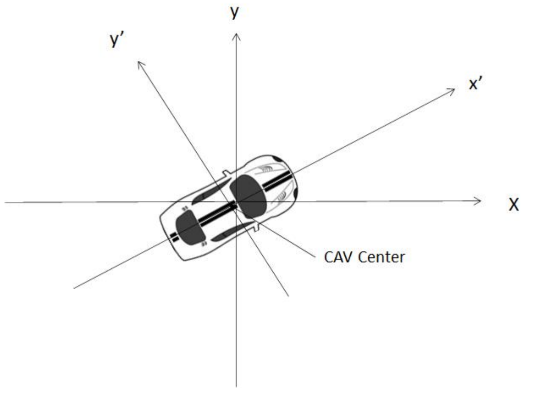

Figure 6 – Spatial Attitude in a CAV

5.5 Input/Output Data of AI Modules

For each AIM (1st column), *Table 8* gives the input (2nd column) and the output data (3rd column). The following 3-digit subsections give the requirements of the data formats in columns 2 and 3.

Table 8 – Environment Sensing Subsystem AIMs and Data

AIM	Input	Output
Spatial Attitude Generation	GNSS Data Spatial Attitude form MAS	Spatial Attitude
Radar Scene Description	Radar Data Basic Environment Representation	Radar Scene Descriptors
Lidar Scene Description	Lidar Data Basic Environment Representation	Lidar Scene Descriptors
Traffic Signalisation Recognition	Camera Data Basic Environment Representation	Road Topology
Camera Scene Description	Camera Data Basic Environment Representation	Lidar Scene Descriptors
Ultrasound Scene Description	Ultrasound Data Basic Environment Representation	Ultrasound Scene Descriptors
Audio Scene Description	Audio Data Basic Environment Representation	Audio Scene Descriptors
Environment Sensing Subsystem Data Fusion	RADAR Scene Descriptors LiDAR Scene Descriptors Road Topology Lidar Scene Descriptors Ultrasound Scene Descriptors	Basic Environment Representation Alert

	Audio Scene Descriptors Map Scene Descriptors Spatial Attitude Other Environment Data	

6 Autonomous Motion Subsystem (AMS)

6.1 Functions of Subsystem

The typical series of operations carried out by the Autonomous Motion Subsystem (AMS) is described below. Note that the sequential description does not imply that an operations can only be carried out after the preceding one has been completed.

1 Human-CAV Interaction requests Autonomous Motion Subsystem to plan and move the CAV to the human-selected destination. A dialogue between AMS-HCI-human may follow.
2 Computes the Route satisfying the human’s request.
3 Receives the current Basic Environment Representation from Environment Sensing Subsystem.
4 While moving, the CAV:
4.1 Broadcasts a subset of the Basic Environment Representation and other data to CAVs in range.
4.2 Receives subsets of Basic Environment Representations and other data from other CAVs.
4.3 Produces the Full Environment Representation by fusing its own Basic Environment Representation with those from other CAVs in range.
4.4 Plans a Path connecting Poses.
4.5 Selects behaviour and motion to reach the next Pose acting on information about the Poses other CAVs in range intend to reach and the Objects between the current Pose and the next Pose.
4.6 Defines a Trajectory that:
4.6.1 Complies with general traffic regulations and local traffic rules.
4.6.2 Preserves passengers’ comfort.
4.7 Refines Trajectory to avoid obstacles.
4.8 Sends Commands to the Motion Actuation Subsystem to take the CAV to the next Pose.
5 Stores the data resulting from a decision (Route Planner, Path Planner etc.)

The AMS should take into account that different levels of autonomy, e.g., those indicated by SAE International [12] are possible depending on the amount and level of available functionalities.

6.2 Reference Architecture of Subsystem

Figure 7 gives the Autonomous Motion Subsystem Reference Model.

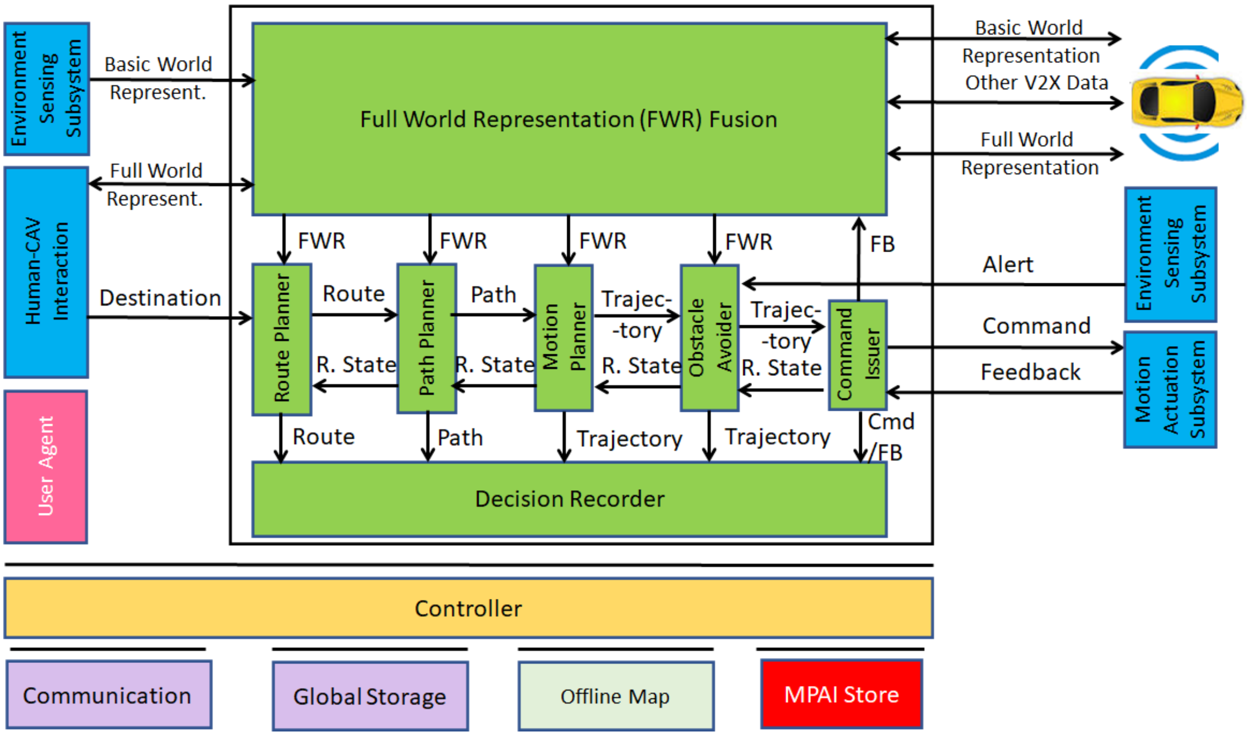

Figure 7 – Autonomous Motion Subsystem Reference Model

This is the operation according to the Reference Model:

1. A human requests the Human-CAV Interaction to be transported to a destination.
2. This request is interpreted and passed to the AMS.
3. The AMS activates the Route Planner to generate a set of Waypoints starting from the current Pose, obtained from the Full Environment Representation, up to the destination.
4. The Waypoints enter the Path Planner which generates a set of Poses to reach the next Waypoint.
5. For each Path, the Motion Planner generates a Trajectory to reach the next Pose.
6. The Obstacle Avoider AIM receives the Trajectory and checks if there is a last-minute change, detected from Alert.
7. If an Alert was received, the Obstacle Avoider AIM checks whether the implementation of the Trajectory creates a collision, especially with the Object creating the Alert.
    a. If a collision is indeed detected, the Obstacle Avoider AIM requests a new Trajectory from the Motion Planner.
    b. If no collision is detected, Obstacle Avoider AIM issues a Command to the Motion Actuation Subsystem.
8. The Motion Actuation Subsystem sends Feedback about the execution of the Command.
9. The AMS, based on the MAS-AMS Responses received and the potential discovery of changes in the Environment, can decide to discontinue the execution of the earlier Command and issue another AMS-MAS Command instead.
10. The decision of each element of the said chain may be recorded.

6.3 Input/Output Data of Subsystem

Table 9 gives the input/output data of Autonomous Motion Subsystem.

Table 9 – I/O data of Autonomous Motion Subsystem

Input data	From	Comment

Command from HCI	Human-CAV Interaction	Human commands, e.g., “take me home”
Basic Environment Representation	Environment Sensing Subsystem	CAV’s Environment representation.
Other V2X Data	Other CAVs	Other CAVs and vehicles, and road-side units.
Feedback from MAS	Motion Actuation Subsystem	CAV’s response to Command.
Output data	**To**	**Comment**
Response to HCI	Human-CAV Interaction	MAS’s response to AMS Command
Command to MAS	Motion Actuation Subsystem	Macro-instructions, e.g., “in 5s assume a given State”.
Full Environment Representation	Other CAVs	For information to other CAVs

6.4 Functions of AI Modules

Table 10 gives the AI Modules of the Autonomous Motion Subsystem.

Table 10 – AI Modules of Autonomous Motion Subsystem

AIM	Function
Route Planner	Computes a Route, through a road network, from the current to the target destination.
Path Planner	Generates a set of Paths, considering: 1. The Route. 2. Spatial Attitude. 3. Full Environment -Representation. 4. Traffic Rules.
Motion Planner	Defines a Goal and a Trajectory to reach the Goal using the Spatial Attitude satisfying the CAV’s kinematic and dynamic constraints and considering passengers’ comfort.
Obstacle Avoider	Checks that the Trajectory is compatible with any Alert information. If it is, it passes the Trajectory to the Command Issuer. If it is not, it requests a new Trajectory. If Command Issuer informs Obstacle Avoider that there is an anomalous situation, Obstacle Avoider may issue a “discontinue previous Command” and forward to the next appropriate upstream AIM, possibly including the Route Planner. This may decide to communicate the Road State to the Human-CAV Interaction Subsystem.
Command Issuer	Instructs the MAS to execute the Trajectory considering the Environment conditions and receives MAS-AMS Responses about the execution.
Full Environment Representation Fusion	Creates an internal representation of the Environment by fusing information from itself, CAVs in range and other transmitting units.

6.5 Input/Output Data of AI Modules

Table 11 gives, for each AIM (1st column), the input data (2nd column) and the output data (3rd column).

Table 11 – CAV Autonomous Motion Subsystem data

CAV/AIM	Input	Output
Route Planner	Pose Destination Offline maps	Route Estimated time
Full Environment Representation	Alert Basic Environment Representations Other V2X Data	Full Environment Representation
Path Planner	Offline maps Route Full Environment Representation	Set of Paths
Motion planner	Path Full Environment Representation	Trajectory
Obstacle Avoider	Trajectory Full Environment Representation	Trajectory Route State
Command to AMS	Trajectory Environment Data Feedback	Command
Ego and Other CAVs	CAV identity and model	CAV identity and model
Ego and Other CAVs	Basic Environment Representation	Basic Environment Representation
Ego and Other CAVs	Full Environment Representation	Full Environment Representation
Ego and Other CAVs	Messages	Messages

The AMS may harvest available bandwidth and utilise it to send a version of the Full Environment Representation that is compatible with the available mobile bandwidth.

7 Motion Actuation Subsystem (MAS)

7.1 Functions of Subsystem

The Motion Actuation Subsystem:

1. Transmits spatial and environmental information gathered from its sensors and its mechanical subsystems to the Environment Sensing Subsystem.
2. Receives AMS-MAS Commands from the Autonomous Motion Subsystem.
3. Translates Commands into specific Commands to its own mechanical subsystems, e.g., brakes, wheels directions, and wheel motors.
4. Receives Responses from its mechanical subsystems.
5. Packages Responses into high-level information.
6. Sends MAS-AMS Responses to AMS Command contain the value of a Spatial Attitude at an intermediate Pose and Time.

7.2 Reference Architecture of Subsystem

Figure 8 represents the Motion Actuation Subsystem Reference Model.

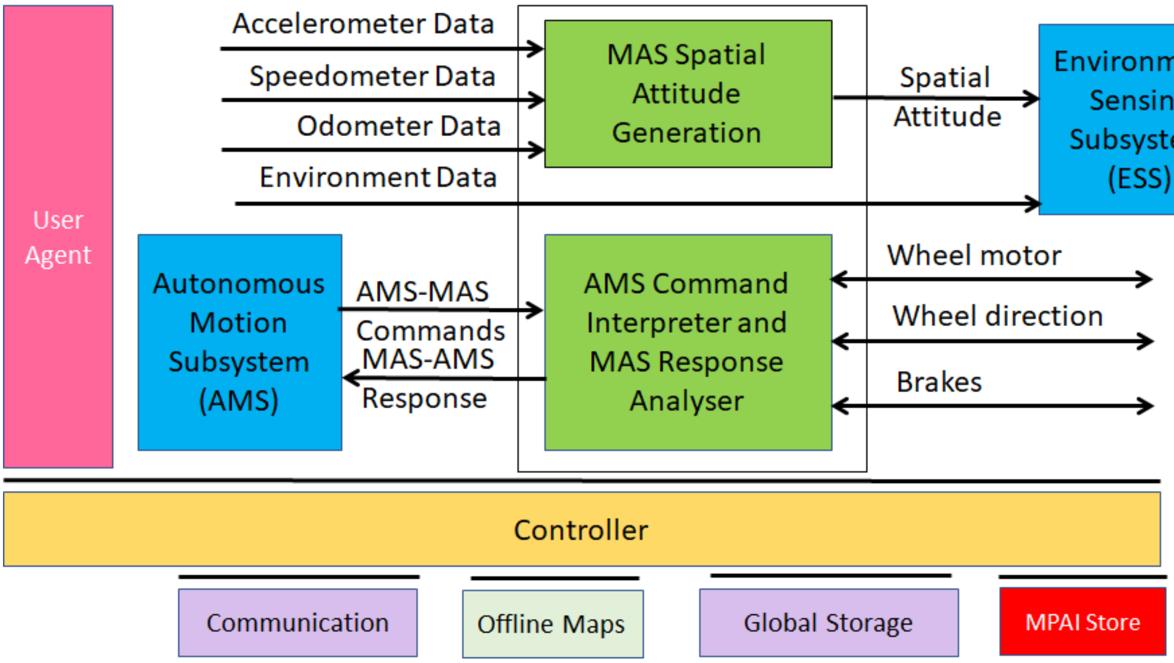

Figure 8 – Motion Actuation Subsystem Reference Model

7.3 Input/Output Data of Subsystem

Table 12 gives the input/output data of Motion Actuation Subsystem.

Table 12 – I/O data of Motion Actuation Subsystem

Input	**Comments**
Odometer	Provides distance data.
Speedometer	Provides instantaneous velocity.
Accelerometer	Provides instantaneous acceleration.
Other Environment data	Provide other environment data, e.g., humidity, pressure, temperature.
AMS-MAS Command	High-level motion command.
Brakes	Acts on brakes, gives feedback.
Wheel Directions	Moves wheels by an angle, gives feedback.
Wheel Motors	Forces wheels rotation, gives feedback.
Output	**Comments**
Motion data	Position, velocity, acceleration.
Other data	Other environment data.
MAS-AMS Response	Feedback from Response Analyser during and after Command execution.

7.4 Functions of AI Modules

Table 13 gives the AI Modules of Autonomous Motion Subsystem.

Table 13 – AI Modules of Motion Actuation Subsystem

AIM	**Function**
Spatial Attitude Generation	Computes Ego CAV’s Spatial Attitude using GNSS, odometer, speedometer, and accelerometer data.

AMS-MAS Command Interpreter and MAS-AMS Response Analyser	Converts AMS Commands into specific actuation commands to Brakes, Wheel directions, and Wheel motors. Forwards MAS-AMS Feedbacks to AMS.

7.5 Input/Output Data of AI Modules

Table 14 gives, for each AIM (1st column), the input data (2nd column) from which AIM (column) and the output data (3rd column).

Table 14 –Motion Actuation Subsystem data

CAV/AIM	Input	Output
AMS Command Interpreter and MAS Response Analyser	AMS-MAS Command Brakes Response Wheel Directions Response Wheel Motors Response	MAS-AMS Response Brakes Command Wheel Motors Command Wheel Directions Command
MAS Spatial Attitude Generation	Odometer Speedometer Accelerometer	Spatial Attitude

8 CAV-to-Everything (V2X)

V2X is the CAV component that allows the CAV Subsystems to communicate to entities external to the Ego CAV. For instance, the HCI of a CAV may send/request information to/from the HCI of another CAV or an AMS may send/request the BER to/from the AMS of another CAV.

8.1 Description

To improve its own capabilities to perceive the Environment, a CAV exchanges information via radio with other entities, e.g., CAVs in range and other CAV-aware communication devices such as Roadside Units and Traffic Lights. Communication may be achieved via secure channels.

A CAV may broadcast to CAVs in range information it has become aware of by using the V2X communication interface. For instance, while executing a Command, the MAS of CAV may become aware of ice on the road. The AMS may decide to broadcast that information to CAVs in range.

Multicast Communication may be used when a CAV broadcasts its identity or in case the data exchanged entail the transmission of large amounts of data such as the Basic Environment Representation (BER). Unicast mode may be used in other cases.

A Communication Device outside of the MPAI-AIF Trusted Zone of the Autonomous Motion Subsystem (AMS) manages communication. The Device communicates with any Ego CAV AIF which has communication needs or from AIFs which the Communication Device has received data.

The operation flow of the Communication Device when handling communication with other CAVs or with devices having CAV functionality (e.g., a traffic light or a roadside unit) can be described as:

1. Receive identities broadcasted by CAVs in range.
2. Establish unicast sessions with CAVs in range.

3. Create a list of CAVs in range with which it has established a session.
4. Send the list with Basic Environment Representations (BER) received via broadcast to the Autonomous Motion Subsystem (AMS).
5. Sends the CAV's BER to CAVs in range.
6. Communication Device may broadcast BER in encrypted form using a key that is only known to CAVs in range that have an open unicast session with the Communication Device.

The Communication Device may also be made aware of the nature of CAVs and CAV-like devices, e.g.:

1. Traffic light.
2. Fire Truck.
3. Police.
4. Ambulance.
5. Flock Leader.

CAVs should communicate using a protocol that assigns a slice of the available transmission rate to each CAV based on the number of CAVs.

8.2 Input and output data

8.2.1 CAVs within range

Table 15 gives the Data Types a CAV broadcasts to CAVs in range via its Communication Device.

Table 15 – I/O data of CAV's Communication Device

Input Data	From	Comments
Basic Environment Representation	Other CAVs	A digital representation of the Environment created by the ESS.
CAV Identity	Other CAVs	In principle, this should be the digital equivalent of today's plate number including Manufacturer and Model information.
CAV Intention	Other CAVs	The Path and other motion data relevant to other CAVs
Full Environment Representation	Other CAVs	A digital representation of the Environment created by fusing all available Basic Environment Representations.
Information Messages	Other CAVs	Typical messages CAV can broadcast. 1. CAV is an ambulance. 2. CAV carries an authority. 3. CAV carries a passenger with critical health problem. 4. CAV has a mechanical problem of an identified level. 5. Works and traffic jams ahead 6. Environment must be evacuated 7. …
Output Data	**To**	**Comments**
Basic Environment Representation	Other CAVs	Same as input for all other input data.
Full Environment Representation	Other CAVs	A digital representation of the Environment obtained from the fusion of all available Basic Environment Representations.

8.2.2 CAV-aware equipment

Examples of such equipment are traffic lights, roadside units, vehicles with CAV communication capabilities. The following data may be exchanged:

1. Identity and coordinates (exact coordinate reference).
2. Static Full Environment Representation regularly updated via download (may be part of the Offline Map).
3. Current objects in Environment.
4. State (Green-Yellow-Red) of traffic light and time to change state.
5. Lane markings.
6. Speed limits.
7. Pedestrian crosswalks
8. General information on the Environment (e.g., one way street etc.)
9. Etc.

Such equipment can:

1. Act as any other CAV in range.
2. Have the authority to organise motion of CAVs in range.

8.2.3 Other non-CAV vehicles

Other vehicles can be scooters, motorcycles, bicycles, etc. possibly transmitting their position as derived from GNSS. No response capability is expected. Vehicles may also have the capability to transmit additional information, e.g., identity, model, speed.

8.2.4 Pedestrians

Their smartphones can transmit their coordinates as available from GNSS. No response capability is expected.

9 Summary of the contributions requested by this Call for Technologies

The Call for Technologies – Connected Autonomous Vehicle (MPAI-MMM) – Architecture specifically requests comments on, modification of, and additions to the following:

1. MPAI-CAV Reference Model (see Chapter 1 - Introduction).
2. Terminology (see Chapter 2 – Terms and definitions).
3. Functions of Subsystems.
4. Input and Output Data of Subsystems.
5. Input and Output Data of AI Modules.
6. Functions of AI Modules.

Annex 1 - General MPAI Terminology

The Terms used in this standard whose first letter is capital and are not already included in *Table 1* are defined in *Table 16.*

Table 16 – 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 AIM-specific inputs and producing AIM-specific inputs according to its Function.
AIF Metadata	The data set describing the capabilities of an AIF set by the AIF Implementer.
AIM Metadata	The data set describing the capabilities of an AIM set by the AIM Implementer.
Application Programming Interface (API)	A software interface that allows two applications to talk to each other
Application Standard	An MPAI Standard specifying AIWs, AIMs, Topologies and Formats suitable for a particular application domain.
Channel	A physical or logical connection between an output Port of an AIM and an input Port of an AIM. The term "connection" is also used as a synonym.
Communication	The infrastructure that implements message passing between AIMs.
Component	One of the 9 AIF elements: Access, AI Module, AI Workflow, Communication, Controller, Internal Storage, Global Storage, MPAI Store, and User Agent.
Conformance	The attribute of an Implementation of being a correct technical Implementation of a Technical Specification.
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 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 2 - The Governance of the MPAI Ecosystem

Level 1 Interoperability

With reference to *Figure 10*, MPAI issues and maintains a standard – called MPAI-AIF – whose components are:

1. An environment called AI Framework (AIF) running AI Workflows (AIW) composed of interconnected AI Modules (AIM) exposing standard interfaces.
2. A distribution system of AIW and AIM Implementation called MPAI Store from which an AIF Implementation can download AIWs and AIMs.

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

Implementers' benefits	Upload to the MPAI Store and have globally distributed Implementations of - AIFs conforming to MPAI-AIF. - AIWs and AIMs performing proprietary functions executable in AIF.
Users' benefits	Rely on Implementations that have been tested for security.
MPAI Store	- Tests the Conformance of Implementations to MPAI-AIF. - Verifies Implementations' security, e.g., absence of malware. - Indicates unambiguously that Implementations are Level 1.

Level 2 Interoperability

In a Level 2 Implementation, the AIW must be an Implementation of an MPAI Use Case and the AIMs must conform with an MPAI Application Standard.

Implementers' benefits	Upload to the MPAI Store and have globally distributed Implementations of - AIFs conforming to MPAI-AIF. - AIWs and AIMs conforming to MPAI Application Standards.
Users' benefits	- Rely on Implementations of AIWs and AIMs whose Functions have been reviewed during standardisation. - Have a degree of Explainability of the AIW operation because the AIM Functions and the Data Formats are known.
Market's benefits	- Open AIW and AIM markets foster competition leading to better products. - Competition of AIW and AIM Implementations fosters AI innovation.
MPAI Store's role	- Tests Conformance of Implementations with the relevant MPAI Standard. - Verifies Implementations' security. - Indicates unambiguously that Implementations are Level 2.

Level 3 Interoperability

MPAI does not generally set standards on how and with what data an AIM should be trained. This is an important differentiator that promotes competition leading to better solutions. However, the performance of an AIM is typically higher if the data used for training are in greater quantity and more in tune with the scope. Training data that have large variety and cover the spectrum of all cases of interest in breadth and depth typically lead to Implementations of higher "quality".

For Level 3, MPAI normatively specifies the process, the tools and the data or the characteristics of the data to be used to Assess the Grade of Performance of an AIM or an AIW.

Implementers' benefits	May claim their Implementations have passed Performance Assessment.
Users' benefits	Get assurance that the Implementation being used performs correctly, e.g., it has been properly trained.
Market's benefits	Implementations' Performance Grades stimulate the development of more Performing AIM and AIW Implementations.
MPAI Store's role	- Verifies the Implementations' security. - Indicates unambiguously that Implementations are Level 3.

The MPAI ecosystem

The following *Figure 9* is a high-level description of the MPAI ecosystem operation applicable to fully conforming MPAI implementations as specified in the Governance of the MPAI Ecosystem Specification [1]:

1. MPAI establishes and controls the not-for-profit MPAI Store.
2. MPAI appoints Performance Assessors.
3. MPAI publishes Standards.
4. Implementers submit Implementations to Performance Assessors.
5. If the Implementation Performance is acceptable, Performance Assessors inform Implementers and MPAI Store.
6. Implementers submit Implementations to the MPAI Store
7. MPAI Store verifies security and Tests Conformance of Implementation.
8. Users download Implementations and report their experience to MPAI.

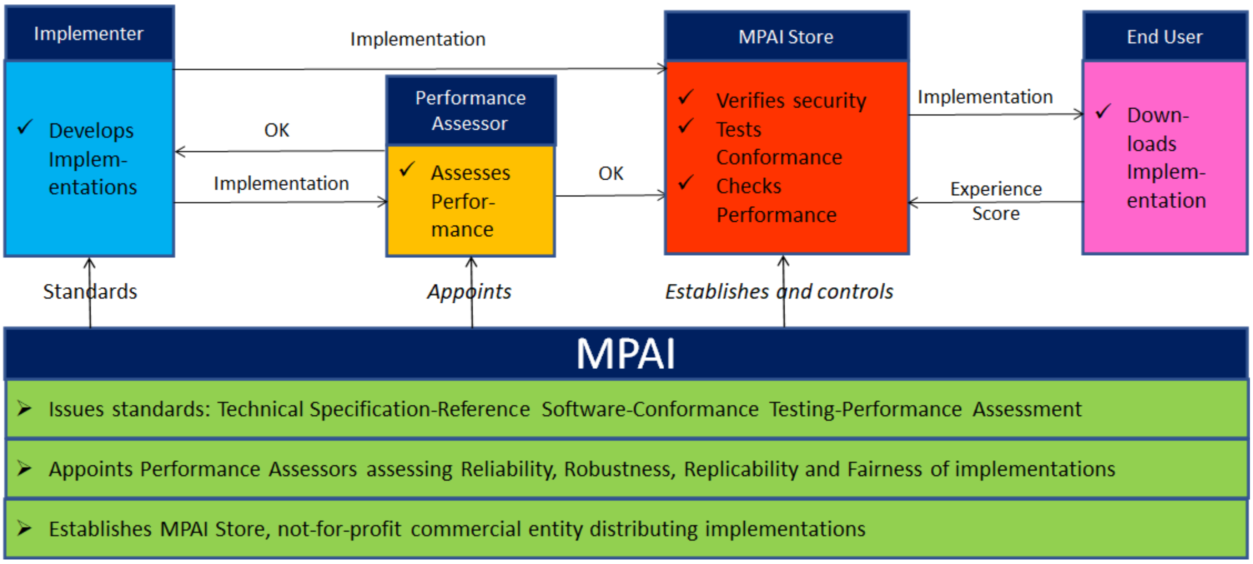
Implementer
✓ Develops Implem-entations
Implementation
OK
Implementation
Performance Assessor
✓ Assesses Perfor-mance
OK
MPAI Store
✓ Verifies security
✓ Tests Conformance
✓ Checks Performance
Implementation
Experience Score
End User
✓ Down-loads Implem-entation
Standards
Appoints
Establishes and controls
MPAI
➢ Issues standards: Technical Specification-Reference Software-Conformance Testing-Performance Assessment
➢ Appoints Performance Assessors assessing Reliability, Robustness, Replicability and Fairness of implementations
➢ Establishes MPAI Store, not-for-profit commercial entity distributing implementations

Figure 9 – The MPAI ecosystem operation

Annex 3 - An overview of relevant MPAI Standards

1 AI Framework

In recent years, Artificial Intelligence (AI) and related technologies have been introduced in a broad range of applications, have started affecting the life of millions of people and are expected to do so even more in the future. As digital media standards have positively influenced industry and billions of people, so AI-based data coding standards are expected to have a similar positive impact. Indeed, research has shown that data coding with AI-based technologies is generally *more efficient* than with existing technologies for, e.g., compression and feature-based description.

However, some AI technologies may carry inherent risks, e.g., in terms of bias toward some classes of users. Therefore, the need for standardisation is more important and urgent than ever.

The international, unaffiliated, not-for-profit MPAI – Moving Picture, Audio and Data Coding by Artificial Intelligence Standards Developing Organisation has the mission to develop *AI-enabled data coding standards*. MPAI Application Standards enable the development of AI-based products, applications, and services.

As a rule, MPAI standards include four documents: Technical Specification, Reference Software Specifications, Conformance Testing Specifications, and Performance Assessment Specifications. The last type of Specification includes standard operating procedures to enable users of MPAI Implementations to make informed decision about their applicability based on the notion of Performance, defined as a set of attributes characterising a reliable and trustworthy implementation.

In the following, if a Term begins with a small letter, it has the commonly used meaning and if with a capital letter, it has either the meaning defined in *Table 2* if it is specific to this Technical Report and in *Table 16* if it is common to all MPAI Standards.

In general, MPAI Application Standards are defined as aggregations – called AI Workflows (AIW) – of processing elements – called AI Modules (AIM) – executed in an AI Framework (AIF). MPAI defines Interoperability as the ability to replace an AIW or an AIM Implementation with a functionally equivalent Implementation.

MPAI also defines 3 Interoperability Levels of an AIF that executes an AIW. The AIW and its AIMs may have 3 Levels:
Level 1 – Implementer-specific and satisfying the MPAI-AIF Standard.
Level 2 – Specified by an MPAI Application Standard.
Level 3 – Specified by an MPAI Application Standard and certified by a Performance Assessor.

MPAI offers Users access to the promised benefits of AI with a guarantee of increased transparency, trust and reliability as the Interoperability Level of an Implementation moves from 1 to 3. Additional information on Interoperability Levels is provided in [1].

Figure 10 depicts the MPAI-AIF Reference Model under which Implementations of MPAI Application Standards and user-defined MPAI-AIF Conforming applications operate [2].

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.

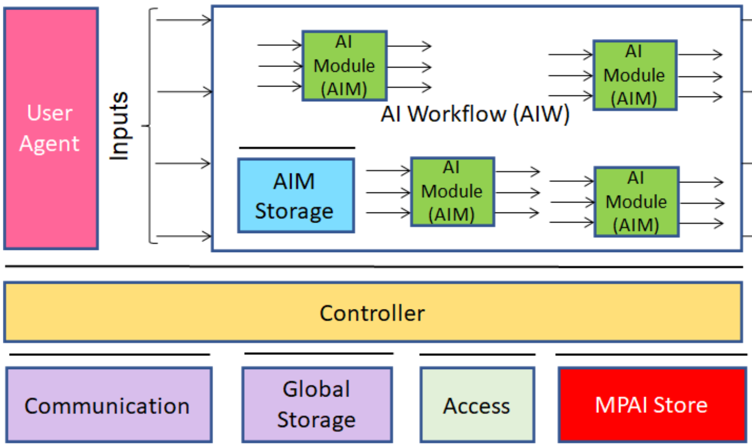

Figure 10 – The AI Framework (MPAI-AIF) V2 Reference Model

It should be noted that an AIM is defined by its Function and data, but not by its internal architecture, which may be based on AI or data processing, and implemented in software, hardware or hybrid software and hardware technologies.

MPAI Standards are designed to enable a User to obtain, via standard protocols, an Implementation of an AIW and of the set of corresponding AIMs and execute it in an AIF Implementation. The MPAI Store in *Figure 10* is the entity from which Implementations are downloaded. MPAI Standards assume that the AIF, AIW, and AIM Implementations may have been developed by independent implementers. A necessary condition for this to be possible, is that any AIF, AIW, and AIM implementations be uniquely identified. MPAI has appointed an ImplementerID Registration Authority (IIDRA) to assign unique ImplementerIDs (IID) to Implementers.[1]

A necessary condition to make possible the operations described in the paragraph above is the existence of an ecosystem composed of Conformance Testers, Performance Assessors, the IIDRA and an instance of the MPAI Store. Reference [1] provides an example of such ecosystem.

2 Personal Status

2.1 General

Personal Status is the set of internal characteristics of a human and a machine making a conversation. Reference [4] 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.

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

2.2 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 11 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.

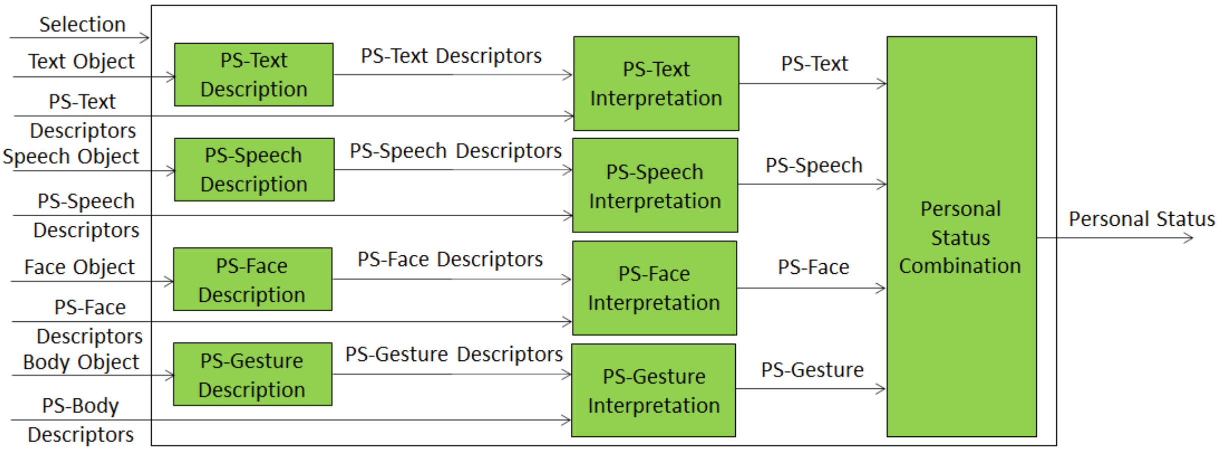

Figure 11 – Reference Model of Personal Status Extraction

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.

2.3 Personal Status Display

A Personal Status Display (PSD) is a Composite AIM receiving Text and Personal Status and generating an avatar producing Text and uttering Speech with the intended Personal Status while the avatar's Face and Gesture show the intended Personal Status. Instead of a ready-to-render avatar, the output can be provided as Compressed Avatar Descriptors. The Personal Status driving the avatar can be extracted from a human or can be synthetically generated by a machine as a result of its conversation with a human or another avatar. Reference Architecture.

Figure 12 represents the AIMs required to implement Personal Status Display.

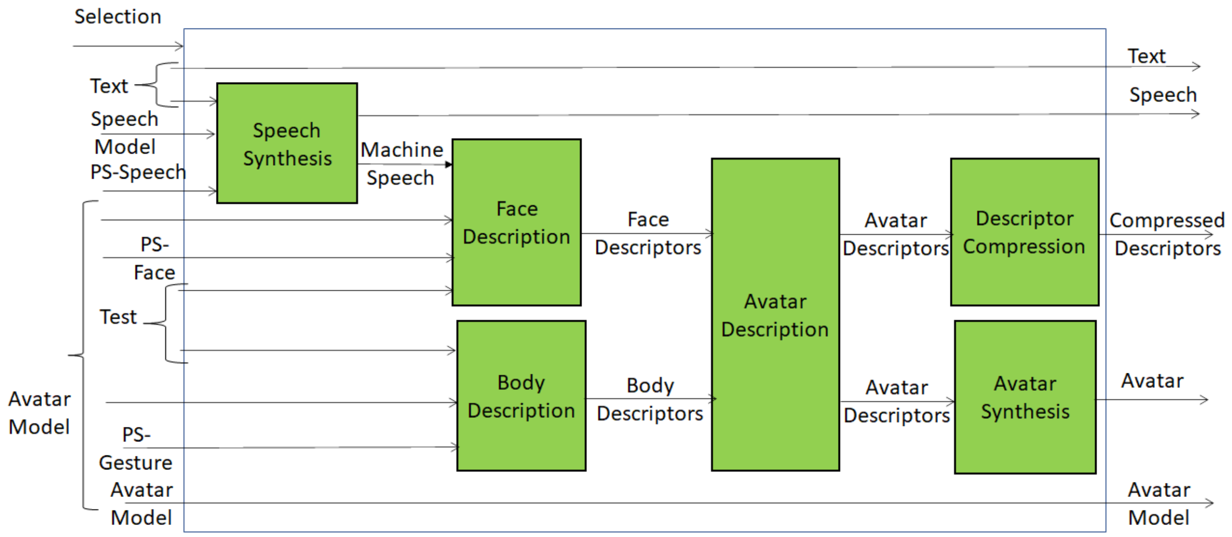

Figure 12 – Reference Model of Personal Status Display

The Personal Status Display operates as follows:

1. Selection determines the type of avatar output – ready-to-render avatar or compressed avatar descriptors.
2. Text is passed as output and synthesised as Speech using the Personal Status provided by PS (Speech).
3. Machine Speech and PS (Face) are used to produce the Face Descriptors.
4. PS (Gesture) and Text are used for Body Descriptors using the Avatar Model.
5. Avatar Description produces a complete set of Avatar Descriptors.
6. Descriptor Compression produces Compressed Avatar Descriptors.
7. Avatar Synthesis produces a ready-to-render Avatar.

3 Human-machine dialogue

Figure 13 depicts the model of the MPAI Personal-Status-based human-machine dialogue.

Audio Scene Description and Visual Scene Description are two front-end AIMs. The former produces 1) Physical Objects, Face and Body Descriptors of the humans, and Visual Scene Geometry; the latter produces Audio Objects and Audio Scene Geometry.

Body Descriptors, Physical Objects and Visual Scene Geometry are used by the Spatial Object Identification AIM. This provides the identifier of the Physical Object the human body is indicating by using the Body Descriptors and the Scene Geometry. The Speech extracted from the Audio Scene Descriptor is recognised and passed to the Language Understanding AIM together with the Physical Object ID. The AIM provided a refined text (Text (Language Understanding)) and Meaning (semantic, syntactic, and structural information extracted from input data).

Face and Body Descriptors, Meaning and Speech are used by Personal Status Extraction to extract the Personal Status of the human. Dialogue Processing produces a textual response with an associated machine Personal Status that is congruent with the input Text (Language Understanding) and human Personal Status. The Personal Status Display AIM produces a synthetic Speech and an avatar representing the machine.

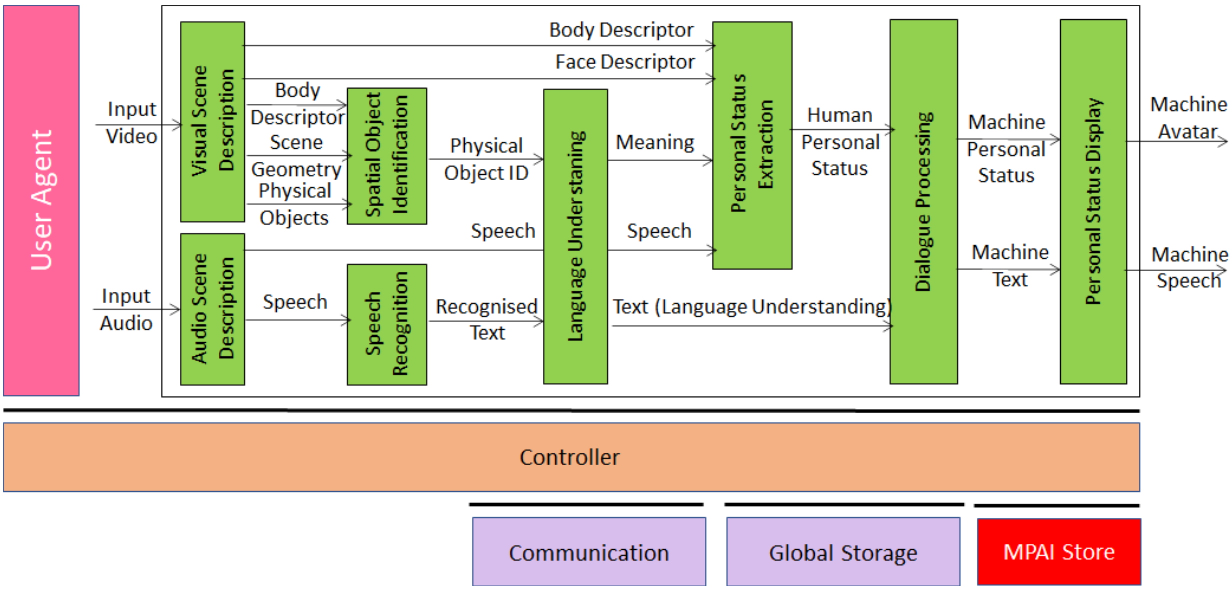

Figure 13 - Personal Status-based Human-Machine dialogue