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

|  |  |
| --- | --- |
|  | |
| **N336** | 2021/08/25 |
| **Source** | Requirements (CAV) |
| **Title** | MPAI-CAV Use Cases and Functional Requirements wd0.3 |
| **Target** | MPAI Community |

[1 Introduction 3](#_Toc80712993)

[2 Terms and definitions 4](#_Toc80712994)

[3 Use Cases 5](#_Toc80712995)

[3.1 Human-CAV interaction (HCI) 6](#_Toc80712996)

[3.1.1 Use Case description 6](#_Toc80712997)

[3.1.2 Reference architecture 7](#_Toc80712998)

[3.1.3 Input and output data 8](#_Toc80712999)

[3.1.4 AI Modules 8](#_Toc80713000)

[3.2 Environment Sensing Subsystem (EDS) 9](#_Toc80713001)

[3.2.1 Use Case description 9](#_Toc80713002)

[3.2.2 Reference architecture 9](#_Toc80713003)

[3.2.3 Input and output data 10](#_Toc80713004)

[3.2.4 AI Modules 10](#_Toc80713005)

[3.3 Autonomous Motion Subsystem (AMS) 11](#_Toc80713006)

[3.3.1 Use Case description 11](#_Toc80713007)

[3.3.2 Reference architecture 11](#_Toc80713008)

[3.3.3 Input and output data 12](#_Toc80713009)

[3.3.4 AI Modules 12](#_Toc80713010)

[3.4 CAV-to-Everything (V2X) 13](#_Toc80713011)

[3.4.1 Use Case description 13](#_Toc80713012)

[3.4.2 Reference architecture 13](#_Toc80713013)

[3.4.3 Input and output data 13](#_Toc80713014)

[3.4.4 AI Modules 14](#_Toc80713015)

[3.5 Motion Actuation Subsystem (MAS) 14](#_Toc80713016)

[3.5.1 Use Case description 14](#_Toc80713017)

[3.5.2 Reference architecture 15](#_Toc80713018)

[3.5.3 Input and output data 15](#_Toc80713019)

[3.5.4 AI Modules 15](#_Toc80713020)

[4 Technologies and Functional Requirements 16](#_Toc80713021)

[4.1 Human-CAV Interaction 16](#_Toc80713022)

[4.1.1 I/O Data summary 16](#_Toc80713023)

[4.1.2 Text 17](#_Toc80713024)

[4.1.3 Audio 17](#_Toc80713025)

[4.1.4 Video 17](#_Toc80713026)

[4.1.5 Object identifier 18](#_Toc80713027)

[4.1.6 Emotion 18](#_Toc80713028)

[4.1.7 Meaning 18](#_Toc80713029)

[4.1.8 Intention 18](#_Toc80713030)

[4.1.9 Reply (speech) 18](#_Toc80713031)

[4.1.10 Reply (face) 18](#_Toc80713032)

[4.1.11 Human commands 18](#_Toc80713033)

[4.1.12 AMS response 18](#_Toc80713034)

[4.1.13 Face identity 18](#_Toc80713035)

[4.1.14 Speaker identity 19](#_Toc80713036)

[4.1.15 Full World Representation 19](#_Toc80713037)

[4.1.16 FWR interaction 19](#_Toc80713038)

[4.2 Environment Sensing Subsystem 19](#_Toc80713039)

[4.2.1 I/O Data summary 19](#_Toc80713040)

[4.2.2 Global Navigation Satellite System (GNSS) 19](#_Toc80713041)

[4.2.3 Odometer 20](#_Toc80713042)

[4.2.4 Speedometer 20](#_Toc80713043)

[4.2.5 Accelerometer 20](#_Toc80713044)

[4.2.6 Radio Detection and Ranging (RADAR) 20](#_Toc80713045)

[4.2.7 Light Detection and Ranging (LIDAR) 21](#_Toc80713046)

[4.2.8 Cameras (2D and 3D) 22](#_Toc80713047)

[4.2.9 Ultrasound 22](#_Toc80713048)

[4.2.10 Microphones 22](#_Toc80713049)

[4.2.11 State 22](#_Toc80713050)

[4.2.12 Offline maps 22](#_Toc80713051)

[4.2.13 Basic World Representation 22](#_Toc80713052)

[4.3 Autonomous Motion Subsystem 23](#_Toc80713053)

[4.3.1 Summary of Autonomous Motion Subsystem data 23](#_Toc80713054)

[4.3.2 User input data 23](#_Toc80713055)

[4.3.3 Offline map 23](#_Toc80713056)

[4.3.4 State 24](#_Toc80713057)

[4.3.5 Goal 24](#_Toc80713058)

[4.3.6 Route 24](#_Toc80713059)

[4.3.7 Basic World Representation 24](#_Toc80713060)

[4.3.8 Traffic signals 24](#_Toc80713061)

[4.3.9 Traffic rules 24](#_Toc80713062)

[4.3.10 Pose 24](#_Toc80713063)

[4.3.11 Velocity 24](#_Toc80713064)

[4.3.12 Full World Description 24](#_Toc80713065)

[4.3.13 Path 24](#_Toc80713066)

[4.3.14 Trajectory 24](#_Toc80713067)

[4.4 CAV to Everything 24](#_Toc80713068)

[4.4.1 Summary of CAV to Everything data 24](#_Toc80713069)

[4.4.2 CAV identity 25](#_Toc80713070)

[4.4.3 Attitude-Path-Trajectory 25](#_Toc80713071)

[4.4.4 Basic World Representation 25](#_Toc80713072)

[4.4.5 Full World Representation 25](#_Toc80713073)

[4.4.6 Events 25](#_Toc80713074)

[4.5 Motion Actuation Subsystem 25](#_Toc80713075)

[4.5.1 Summary of Motion Actuation Subsystem data 25](#_Toc80713076)

[4.5.2 Command 26](#_Toc80713077)

[4.5.3 Road wheel motor control 26](#_Toc80713078)

[4.5.4 Road wheel direction control 26](#_Toc80713079)

[4.5.5 Brake control 26](#_Toc80713080)

[4.5.6 Feedback to Command 26](#_Toc80713081)

[4.5.7 Odometer data 26](#_Toc80713082)

[4.5.8 Speedometer data 26](#_Toc80713083)

[4.5.9 Accelerometer data 26](#_Toc80713084)

[4.5.10 Position 26](#_Toc80713085)

[4.5.11 Velocity 26](#_Toc80713086)

[4.5.12 Acceleration 26](#_Toc80713087)

[5 Data privacy 26](#_Toc80713088)

[5.1 Human-CAV Interaction (HCI) 26](#_Toc80713089)

[5.2 Environment Sensing Subsystem (EDS) 26](#_Toc80713090)

[5.3 Autonomous motion subsystem (AMS) 27](#_Toc80713091)

[5.4 CAV to Everything (V2X) 27](#_Toc80713092)

[5.5 Motion Actuation Subsystem (MAS) 27](#_Toc80713093)

[Annex 1 - Terminology 28](#_Toc80713094)

[Annex 2 – Datasets for CAV research 29](#_Toc80713095)

[Annex 3 – ETSI Technical Report 30](#_Toc80713096)

[Annex 4 – Some CAV Communication Technologies 31](#_Toc80713097)

[6 References 32](#_Toc80713098)

# Introduction

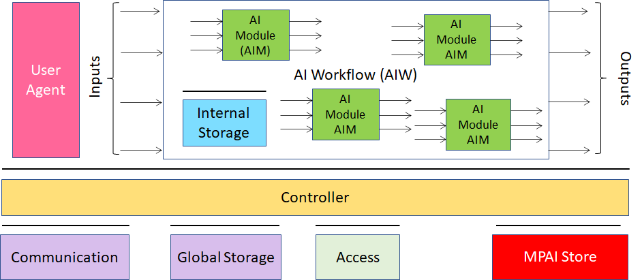
Moving Picture, Audio and Data Coding by Artificial Intelligence (MPAI) is an [international Standards Developing Organisation](http://mpai.community/) with the mission to develop *AI-enabled data coding standards*. Research has shown that data coding with AI-based technologies is generally *more efficient* than with existing technologies. Compression and feature-based description are notable examples of coding.

*Compression and understanding of industrial data* (MPAI-CUI) is an MPAI Standard, comprising the “AI-based Company Performance Prediction (CPP)” Use Case. CPP uses AI to extract the most relevant information from indus­trial data, with the aim of assessing the performance of a company and predicting the risk of bankruptcy long before it may happen.

The current version of MPAI-CUI has been developed by the MPAI Compression and Understan­ding of Industrial Data Development Committee (CUI-DC). Future versions of the standard may extend the scope of the Use Case and/or add new Use Cases in the scope of Compression and Understanding of Industrial Data.

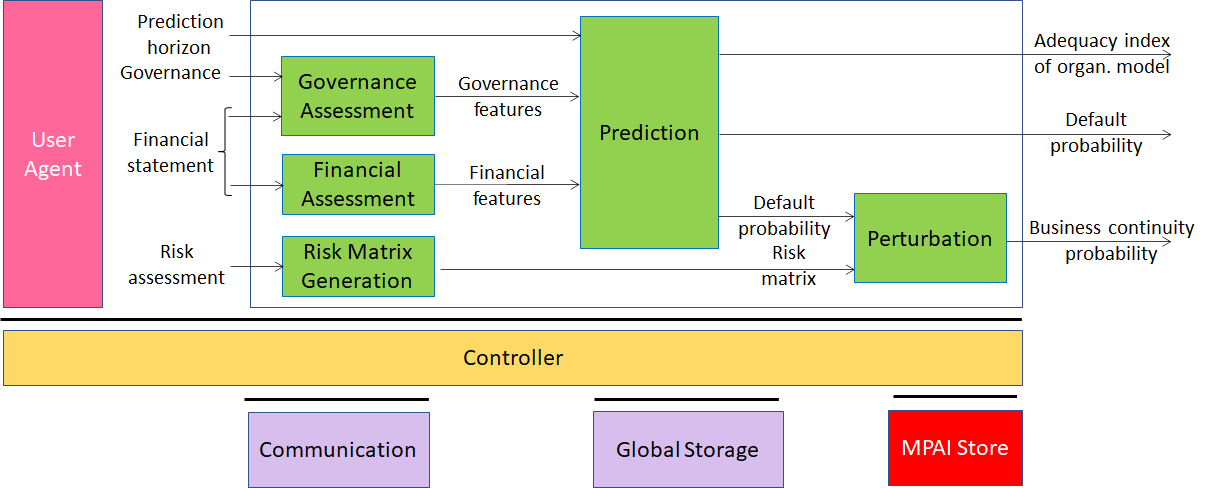
In the following Terms beginning with a capital letter are defined in *Table 1* if they are specific to this MPAI-CUI Standard and to *Table 10* if they are common to all MPAI Standards.

The AI Framework (AIF) execution environment (MPA-AIF) [2] depicted in *Figure 1* enables Interoperable AI applications and services. Further details of the AIF Reference Model can be found in Annex 3.



*Figure 1 –Architecture and Components of the AI Framework (AIF)*

MPAI-CUI normatively specifies the technologies required to support the AI-based Company Performance Prediction Use Case (CPP). *Figure 2* is the instantiation of the general AIF Architec­ture to the CPP Use Case. Note: the Internal Storage and Access of *Figure 1* are non represented.

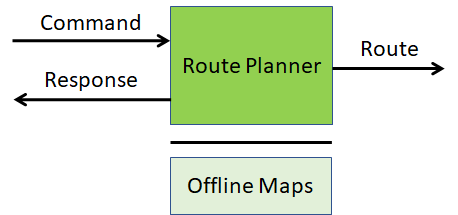


*Figure 2 – AI-based Company Performance Prediction Reference Model*

The rectangle including green boxes in *Figure 2* is called AI Workflow (AIW). The AIW is charac­terised by the following normative elements:

1. The format and semantics of the input data, i.e., “Prediction Horizon”, “Governance”, “Finan­cial Statement” and “Risk Assessment”.
2. The function of the AIW, i.e., “Compute company performance indicators in a given prediction horizon”.
3. The format and semantics of the output data, i.e., “Adequacy index of organisational data”, “Default probab­ility” and “Business continuity probability”.
4. The Connections of the green boxes – called AI Modules (AIM).

Each AIM, like the one depicted in *Figure 3* called Governance Assessment,



*Figure 3 – An example of AI Module (AIM)*

is characterised by the following normative elements:

1. The format and semantics of the input data, i.e., “Governance” and “Financial Statement”.
2. The function, i.e., “Compute the Governance Features”.
3. The format and semantics of the output data, i.e., “Governance Features”.

AIMs are defined by their functions and interfaces, not by their internal architectures, which may need not be known and may be based on AI or data processing, and implemented in software, hardware or hybrid software and hardware technologies.

MPAI defines Interoperability as the ability to replace an AIM in an AIW with another AIM having the same normative elements and Interoperability Level and obtain a functionally equivalent AIM.

An Implementation may have 3 different Interoperability Levels:

|  |  |
| --- | --- |
| Level 1 | An AIF Implementation running an AIW composed of AIMs performing any propri­etary function and exposing any proprietary interface but exposing the interfaces requ­ired to be executed in the AIF. |
| Level 2 | An AIF Implementation running an AIW composed of AIMs whose functions and interfaces are specified by an MPAI Application Standard. |
| Level 3 | An Implementation running an AIW composed of AIMs certified to possess the attributes of Reliability, Robustness, Replicability and Fair­ness – collectively called Performance. |

Moving Picture, Audio and Data Coding by Artificial Intelligence (MPAI) is an [international standards developing organisation](http://mpai.community/) with the mission to develop *AI-enabled data coding standards*. Research has shown that data coding with AI-based technologies is generally *more efficient* than with existing technologies. Compression and feature-based description are notable examples of coding.

MPAI defines 3 different levels of interoperability offered to implementers of based on MPAI Standards:

|  |  |
| --- | --- |
| Level 1 | Implementations running workflows of data processing elements – referred to AI Modules (AIM). These can have any function and interface but retain those that make AIMs executable in an MPAI standard environment called AI Framework (AIF). |
| Level 2 | Implementations that have the additional feature of execuring workflows and AIMs having functions and interfaces specified by MPAI Standards executed in an AIF. |
| Level 3 | Implementations with the additional feature of having Level 2 workflows and AIMs certified to have the attributes of Reliability, Robustness, Replicability and Fairness, collectively called Performance. |

MPAI normatively specifies the following aspects of an AI Module (AIM):

1. The format and semantics of the input data, e.g., “video of a talking human face”.
2. The function, e.g., “identification of the emotion on the face of and the meaning of the sentence uttered by a speaking human”.
3. The format of the output data, e.g., “emotion” and “meaning”.

AIMs are solely defined by their functions and interfaces, but not by their internal architecture which can be based on AI or data processing elements, software, hardware or hybrid software and hardware.

The 2 basic elements of the MPAI standardisation are represented in *Figure 1* and *Figure 2*.

*Figure 1* shows an AIM (deep green block) which has the function of detecting the emotion on the face and the meaning of the sentence the human is uttering using the video coming from a camera shooting a human face. The AIM can be implemented with a neural network or with DP technologies. In the latter case, the AIM accesses a knowledge base external to the AIM (light green box).

|  |  |
| --- | --- |
|  | Diagram  Description automatically generated |
| *Figure 1 – An MPAI AI Module (AIM)* | *Figure 2 – The MPAI AI Framework (AIF)* |

The MPAI approach to developing AI data coding standards is based on the definition of *standard interfaces* of *AI Modules (AIM)* *combined* and *executed* in an MPAI-specified *AI-Framework* (MPAI-AIF). AIMs operate on input data and produce output data, both with standard formats. MPAI is silent on how an AIM produces output data from input data, with the constraint that an MPAI-standardised AIM must execute the normatively specified function.

By exposing standard interfaces, AIMs can interoperate in the MPAI AI Framework. However, their performance may differ depending on the technologies used to implement them.

MPAI believes that *competing* developers striving to provide more performing *proprietary* while still *interoperable* AIMs will naturally create *horizontal markets* of *AI solutions* that build on and further promote AI *innovation*.

Each Use Case normatively defines:

1. *Data format*: any type of static (time independent) or dynamic (time dependent) data that is used in an AI system, e.g., video, emotion and meaning in *Figure 1*.
2. *AI Module*: a subsystem that is characterised by:
   1. the function performed by the AIM (e.g., extraction of emotion and meaning).
   2. the data entering and leaving the AIM, e.g., video, emotion and meaning in *Figure 1*.
3. *Use Case*: an AI system that implements an MPAI-specified Use Case characterised by
   1. the function performed by the AI system.
   2. the data entering and leaving the AI system, e.g., those of *Figure 2*.
   3. the topology and connection of the AIMs in the AI system.

The word *normatively* is to be interpreted to mean that if an implementer claims conformance to

1. A Use Case, the implementation shall perform the functions specified by the Use Case, its AIMs and their data formats.
2. An AIM, the implementation shall perform the functions specified by the AIMs, and receive as input and produce as output the data in the specified formats.
3. A data format, the data shall have the format specified in the referenced AIM.

# Terms and definitions

*Table 1* defines the terms used in this document. Terms are organised by the CAV Subsystems identified in *Figure 3*. The general MPAI Terms are defined in

*Table 1 – Definition of Terms used in this document organised by sybsystems*

|  |  |  |
| --- | --- | --- |
| **SubS** | **Term** | **Definition** |
| AMS | Command | High-level instructions whose execution allows a CAV to reach a Goal. |
| AMS | Decision horizon | The estimated time between the current state and the Goal. |
| AMS | Full World Representation | A description of Environment using the CAV’s and other CAVs’ Basic World Representation. |
| AMS | Goal | The planned State at a future time. |
| AMS | Path | A sequence of Poses 𝑝𝑖 = (𝑥𝑖,𝑦𝑖,zi,𝜃𝑖) in the Offline Map. |
| AMS | Pose | Coordinates and orientation of the CAV in the Offline Map *p* = (𝑥,𝑦,z,𝜃) |
| AMS | Route | A sequence of Way Points |
| AMS | State | Pose, Velocity and Acceleration of a CAV at a given time. |
| AMS | Way Point | A point 𝑤𝑖 given as a coordinate pair (𝑥𝑖, 𝑦𝑖), in an Offline Map |
| CAV | Connected Autonomous Vehicle | A vehicle capable to autonomously reach an assigned target by understan­ding human utterances, planning a route, sensing and interpreting the environment, exchanging information with other CAVs and acting on the CAV’s motion subsystem. |
| ESS | Basic World Representation | A description of the local CAV environment based on the CAV sensors and an Offline Map. |
| ESS | Environment | The portion of the world of interest to the CAV. |
| ESS | Global Navigation Satellite System | (GNSS) includes GPS, Galileo, Glonass. BeiDou, Quasi Zenith Satellite System (QZSS) and Indian Regional Navigation Satellite System (IRNSS). |
| ESS | Inertial Measurement Unit | An inertial positioning device, e.g., accelerometer, speedometer, gyros­cope, odometer etc. |
| ESS | Offline Map | An offline-created map of a location and associated metadata. |
| HCI | Command | High-level instructions whose execution allows a CAV to reach a Goal. |
| HCI | Full World Representation | A description of Environment using the CAV’s and other CAVs’ Basic World Representation. |

# Use Cases

fully autonomous, but exceprional human interventions

The MPAI-CAV Use Cases relate to the 5 main subsystems in a Connected Autonomous Vehicle:

*Human-CAV interaction (HCI)* recognises the human CAV rights holder, responds to humans’ commands and queries, provides extended environment representation (Full World Repres­entation) for humans to enjoy, senses human activities during the travel and may activate other subsystems as required by humans or as deemed necessary by the identified conditions.

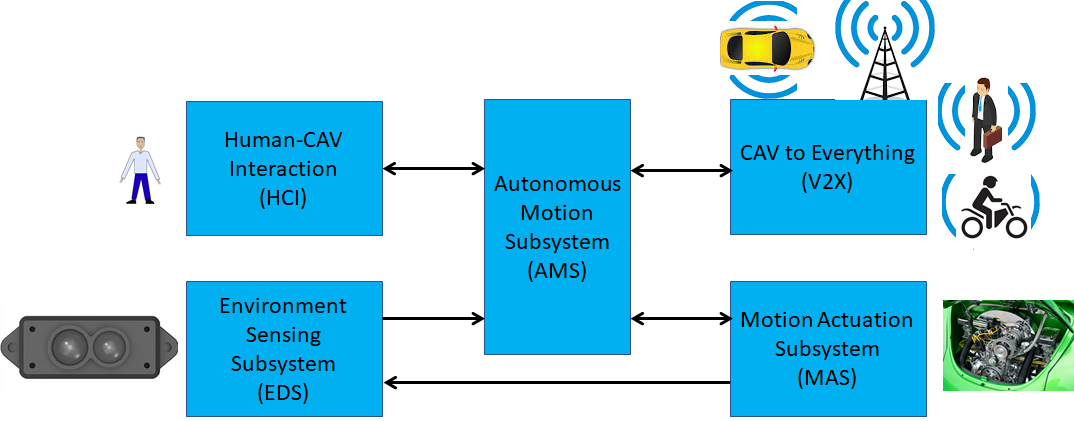
*Environment Sensing Subsystem (EDS)* acquires information from the physical environment via a variety of sensors and develops the best representation of the environment (Basic World Representation).

*Autonomous Motion Subsystem (AMS)* computes the Route to destination, uses dif­ferent sources of information –CAV sensors, other CAVs and transmitting units – to produce a Full World Representation and gives command that drive the CAV to the intended destination.

*CAV to Everything Subsystem (V2X)* receives information from external sources, including other CAVs, other vehicles, Roadside Units (RSU) and sends information to other CAVs.

*Motion Actuation Subsystem (MAS)* provides environment information¸ receives and actuates motion commands in the physical world.

The interaction of the 5 subsystems in depicted in *Figure 3*:



*Figure 3 – The CAV subsystems*

The following high-level workflow illustrates the operation of the CAV envisaged by this docum­ent.

A *human* with appropriate credentials requests the CAV, via Human-CAV Interaction, to take the human to a given place.

*Human-CAV Interaction* authenticates the human, interprets the request and passes a command to the *Autonomous Motion Subsystem*.

*Autonomous Motion Subsystem*:

Requests *Environment Sensing Subsystem* to provide the current Pose.

Computes the Route.

Issues the start command.

*Environment Sensing Subsystem* computes and sends Basic World Representations to the *Autonomous Motion Subsystem.*

*CAV to Everything*

Becomes aware of other CAVs and external sources (CAVs, RSU etc.).

Shares the CAV’s Basic World Representation with CAVs in range.

*Autonomous Motion Subsystem*:

Receives and processes data broadcasted by external sources (CAVs, RSU etc.).

Computes the Full World Representation.

Shares the CAV’s Full World Representation with CAVs in range.

Computes a Path.

Issues commands to the *Motion Actuation Subsystem* to move the CAV accordingly.

While the CAV moves, humans

Interact and hold conversation with the Human-CAV Interaction and possibly other humans on board.

Issue commands.

Request views of the environment (Full World Representation) etc.

## Human-CAV interaction (HCI)

### Use Case description

A human and a CAV interact in several ways:

1. CAV authenticates human.
2. Human-CAV commands, e.g.,
   1. Commands to Autonomous Motion Subsystem, e.g.: go to a Way point, display Full World Representation (see 5.3), etc.
   2. Other commands, e.g.: turn off air conditioning, turn on radio, call a person, open window or door, search for information etc.
3. Human-CAV dialogue, e.g..
   1. Information requests, e.g.: time to destination, route conditions, weather at destination etc.
   2. Casual conversation.
4. Compartment monitoring, e.g.,
   1. Physical conditions, e.g.: temperature level, media being played, sound level, noise level, anomalous noise, etc.
   2. Passenger data, e.g.: number of passengers, ID, estimated age, destination of passengers.
   3. Passenger activity, e.g.: level of passenger activity, level of passenger-generated sound, level of passenger movement, emotion on face of passengers.
   4. Passenger-to-passenger dialogue, two passengers shake hands, or passengers hold everyday conversation.

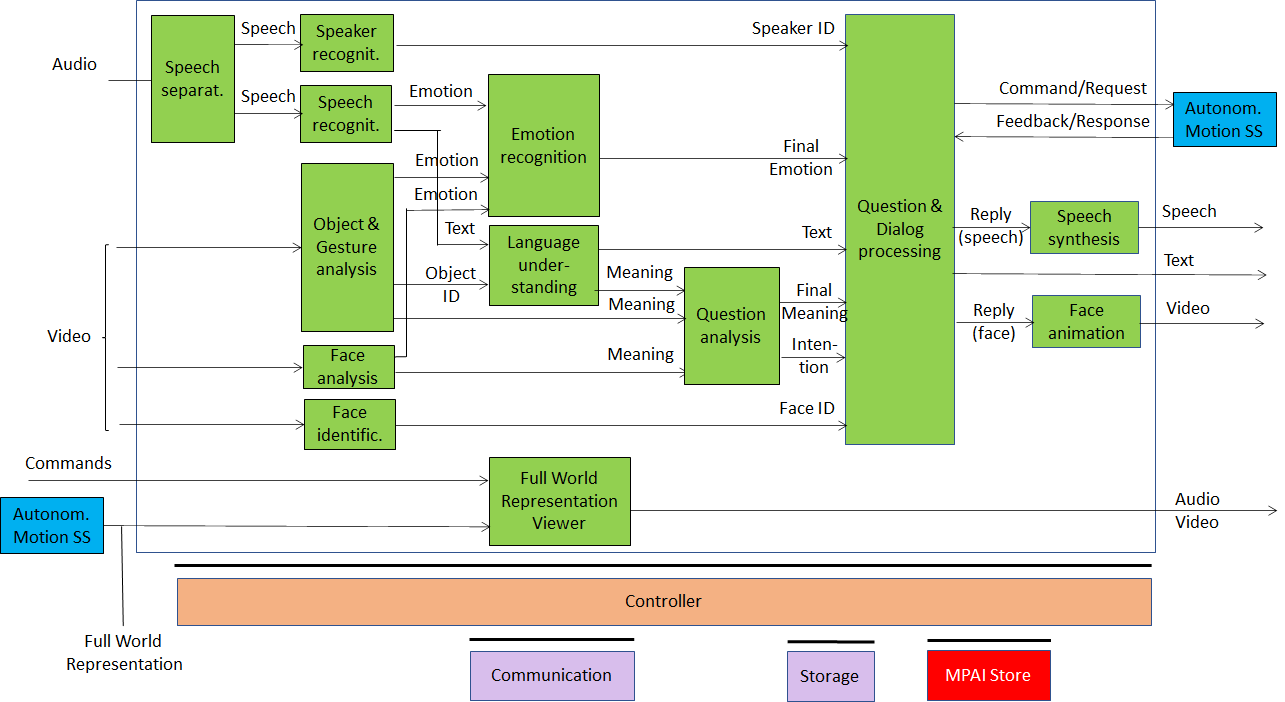
Why we use emotion. this is just the first step in the implementation of a fully natural H-Cav dialogue. CAV getting angry about another driver. cav to be reactive to the behavious in the Environment. human makes a joke and the cav shows an appropriate face

The CAV collects data generated by humans inside the vehicle for possible action. In general, such data shall be anonymised, if meant for later, e.g., statistical, use. Otherwise, any data related to an identifiable human shall be deleted at the end of the travel.

### Reference architecture

*Figure 4* is the reference model of Human-CAV Interaction (HCI). The following is noted

1. A combination of Conversation with Emotion and Multimodal Question Answering AIMs with gesture recognition capabilities covers most Human-CAV Interaction needs.
2. Additional AIMs can be added should new HCI interactions be required.



*Figure 4 – Human-CAV Interaction Reference Model*

Depending on the technology used (data processing or AI), the AIMs in *Figure 4* may need to access ex­ternal information, such as Knowledge Bases, to perform their functions. While not represented in *Figure 4*, they will be identified, if required, in the AI Modules subsection.

### Input and output data

|  |  |  |
| --- | --- | --- |
| **Input data** | **From** | **Comment** |
| Audio | User outdoor | User authentication  User command |
| Text | User outdoor | User authentication  User command |
| Text | Passenger compartment | Social life of user  Commands or interaction with CAV |
| Audio | Passenger compartment | Social life of user  Commands or interaction with CAV |
| Video | Passenger compartment | Social life of user  Commands or interaction with CAV |
| Full World Representation | Autonomous Motion SS | For processing by FWR Viewer |
| **Output data** | **To** | **Comments** |
| Text | Autonomous Motion Subsystem | Commands to be executed |
| Synthetic speech | Passenger compartment | CAV’s response to passengers |
| Synthetic face | Passenger compartment | CAV’s response to passengers |
| Full World Representation | Passenger compartment | For passengers to view external world |

### AI Modules

The AI Modules of the Human-CAV Interaction depicted in *Figure 4* are given in *Table 2*.

*Table 2* *– AI Modules of* *Human-CAV interaction*

|  |  |
| --- | --- |
| **AIM** | **Function** |
| **Speech detection and separation** | 1. Separates relevant speech vs non-speech signals 2. Detects request for dialogue. |
| **Speaker identification** | Recognises speaker. |
| **Speech recognition** | 1. Analyses the speech input 2. Generates text and emotion output. |
| **Object and gesture analysis** | 1. Analyses video to identify object 2. Produces the ID of the object in focus 3. Analyses video 4. Produces motion and mean­ing of gesture. |
| **Face recognition** | 1. Analyses the video of the face of a human 2. Recognise the human’s identity. |
| **Face analysis** | 1. Analyses the video of the face of a human 2. Extracts emotion and meaning. |
| **Language understanding** | 1. Uses a language model (embedded in AIM) 2. Analyses natural language expressed as text 3. Produces the meaning of the text. 4. Produces text related to Object ID |
| **Emotion recognition** | 1. Fuses Emotions from Speech, Face and Gesture. 2. Produces Final Emotion. |
| **Question analysis** | 1. Fuses Meanings of Speech, Face and Gesture 2. Analyses the meaning of the sentence 3. Determines the Intention. 4. Outputs Final Meaning |
| **Question & dialog processing** | 1. Receives Speaker ID and Face ID 2. If speaker ID and face ID match, then    1. Produces a command to Autonomous Motion SS    2. Analyses user’s emotion, intention, meaning and/or ques­tion, text    3. Produces Reply (speech) and Reply (face). 3. Else, responds appropriately |
| **Speech synthesis** | Converts Reply (speech) to speech. |
| **Face animation** | Converts Reply (face) to animated face |
| **Full World Representation Viewer** | 1. Receives Full World Representation 2. Presents a view as instructed by human via Commands. |

## Environment Sensing Subsystem (EDS)

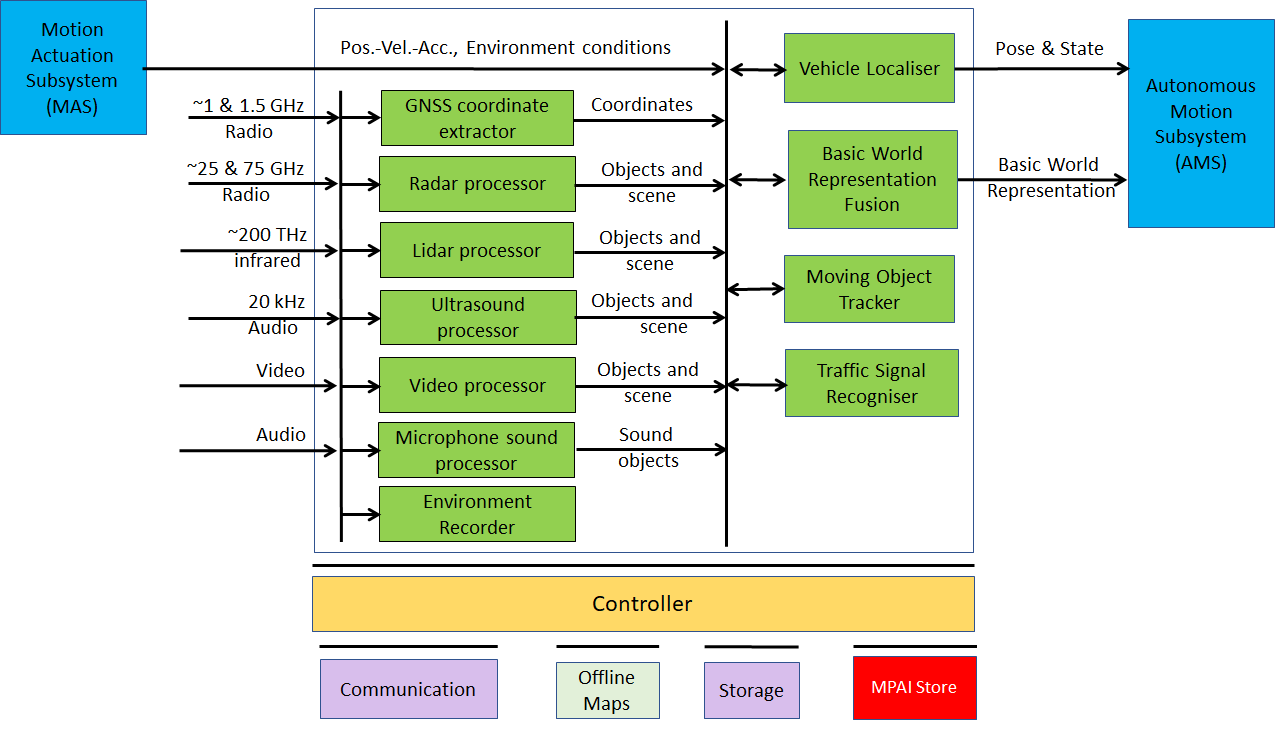
### Use Case description

The typical series of operations carried out by the Environment Sensing Subsystem (ESS) is given below. The sequential description of steps does not imply that an action is only carried out after the preceding one has been completed.

1. The CAV gets its Pose and other environment data from
   1. Global Navigation Satellite System (GNSS).
   2. Pose Data Generation in Environment Sensing Subsystem (EDS).
   3. Other sensor data converter (e.g., weather, air pressure etc.)
2. The CAV creates a Basic World Representation (BWR) by:
   1. Acquiring available Offline maps of its current Pose.
   2. Updating the Offline maps with
      1. Other static objects.
      2. All moving objects.
      3. All traffic signals.
3. The CAV compresses a subset of the sensor data and stores them on board.

### Reference architecture

*Figure 5* gives the Reference Model.



*Figure 5 – Environment Sensing Subsystem Reference Model*

### Input and output data

|  |  |  |
| --- | --- | --- |
| **Input data** | **From** | **Comment** |
| Global Navigation Satellite System (GNSS) | ~1 & 1.5 GHz Radio | Get Pose from GNSS |
| Radio Detection and Ranging (RADAR) | ~25 & 75 GHz Radio | Ger RADAR view of environment |
| Light Detection and Ranging (LIDAR) | ~200 THz infrared | Ger LiDAR view of environment |
| Ultrasound | 20 kHz Audio | Ger 20 kHz view of environment |
| Cameras (2/D and 3D) | Video (400-800 THz) | Get visible view of environment |
| Microphones | 16 Hz-16 kHz sound | Get audible view of environment |
| **Output data** | **To** | **Comment** |
| CAV’s Pose and State | Autonomous Motion Subsystem | For Route, Path and Trajectory |
| Basic World Representation | Autonomous Motion Subsystem | To locate CAV in environment |

### AI Modules

The AI Modules of Environment Sensing Subsystem are given in *Table 3*.

*Table 3 – AI Modules of* *CAV-Environment Interaction*

|  |  |
| --- | --- |
| **AIM** | **Function** |
| **GNSS coordinate extractor** | Computes global coordinates of CAV. |
| **Radar Processor** | Extracts electromagnetic scene and objects. |
| **Lidar Processor** | Extracts electromagnetic scene and objects. |
| **Ultrasound Processor** | Extracts ultrasound scene and objects. |
| **Cameras Processor** | Extracts visual scene and objects. |
| **Microphone Sound Processor** | Extracts audible audio scene and objects. |
| **Environment Recorder** | Compresses/records a subset of data produced by CAV sensors at a given time. |
| **Vehicle Localiser** | Estimates the current CAV State in the Offline Maps. |
| **Moving Objects Tracker** | Detects, tracks and represents position and velocity of Environment moving objects. |
| **Traffic Signal Recogniser** | Detects and recognises traffic signs to enable the CAV to correctly move in conformance with traffic rules. |
| **Basic World-Representation Fusion** | 1. Integrates Offline map, moving and traffic objects, and other sensor data 2. Creates Basic World-Representation. |

## Autonomous Motion Subsystem (AMS)

### Use Case description

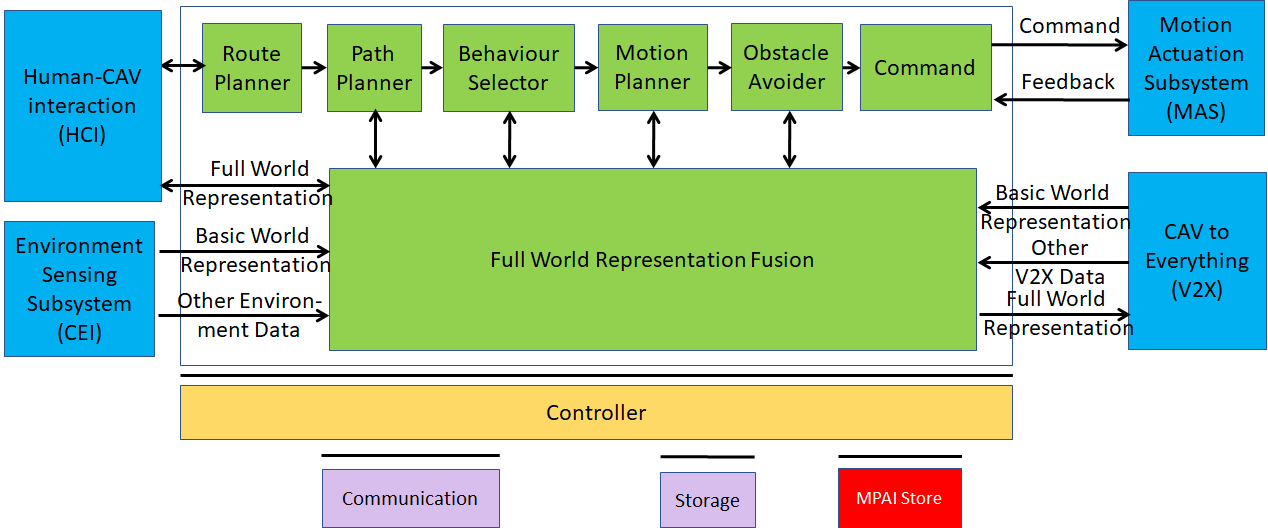
The typical series of operations carried out by the Autonomous Motion Subsystem (AMS) is described below. Note that the sequential description does not imply that an action can only be carried out after the preceding one has been completed.

1. Human-CAV Interaction requests Autonomous Motion Subsystem to plan and move the CAV to the Pose indicated by the human.
2. The CAV requests Environment Sensor Subsystem to provide the current Basic World Representation
3. While moving. the CAV
   1. Transmits the Basic World Representation and other data to CAVs in range.
   2. Receives Basic World Representations and other data from CAVs in range.
   3. Fuses its own Basic World Representation with those from other CAVs to produce the Full World Representation.
   4. Plans a Path connecting several Poses.
   5. Selects a behaviour to reach intermediate goals.
   6. Defines a trajectory, complying with general traffic rules and local traffic regulations and taking passengers’ comfort into account.
   7. Avoids obstacles.
   8. Sends the Motion Actuation Subsystem the commands to take the CAV to the next intermediate goal.

The AMS should be designed in such a way that different levels of autonomy, e.g., those indicated by SAE International [1], are possible depending on the amount and level of available func­tionalities.

### Reference architecture

The Autonomous Motion Subsystem reference model is given by *Figure 6*.



*Figure 6 – Autonomous Motion Subsystem Reference Model*

### Input and output data

|  |  |  |
| --- | --- | --- |
| **Input data** | **From** | **Comment** |
| Human | Human-CAV Interaction | Human commands, e.g., “take me home” |
| Basic World Representation | Environment Sensing Subsystem | Internal BWR |
| Other Environment Data | Environment Sensing Subsystem | E.g., tempteratire, air pressure |
| Other V2X Data | CAV to Everything | Roadside units, other vehicles |
| Feedback | Motion Actuation Subsystem | CAv’s response to command |
| **Output data** | **To** | **Comment** |
| Response to human | Human-CAV Interaction | CAV’s response to command |
| Command | Motion Actuation Subsystem | Micro-instructions |
| Full World Representation | CAV to Everything | Informing other CAVs |

### AI Modules

The AI Modules of Autonomous Motion Subsystem are given in *Table 4*.

*Table 4 – AI Modules of Autonomous Motion Subsystem*

|  |  |
| --- | --- |
| **AIM** | **Function** |
| **Route Planner** | Computes a Route 𝑊, through a road network, from the CAV’s current State to the target Pose. |
| **Path Planner** | Generates a set of Paths, considering 1) the current Route, 2) the CAV State, 3) the Full World-Representation, and 4) the traffic rules. |
| **Behaviour Selector** | Sets a Goal to be reached with a Driving Behaviour, avoiding collisions with static and moving objects within the decision horizon time frame. |
| **Motion Planner** | Defines a Trajectory, from the current CAV State to the current Goal fol­lowing the Behaviour Selector’s Path as close as possible, satisfying the CAV’s kinematic and dynamic constraints, and passengers’ comfort. |
| **Obstacle Avoider** | Defines a new Trajectory that avoids obstacles. |
| **Command** | Instructs the CAV to execute the Trajectory as best as the Environment allows. |
| **Full World-Representation Fusion** | Creates an internal representation of the Environment by fusing its own information, that of CAVs in range and of other transmitting units.. |

## CAV-to-Everything (V2X)

### Use Case description

CAVs can exchange information via radio with other entities, e.g., nearby CAVs and other infor­mation sources such as roadside units, thereby improving their perception capabilities:

CAVs within range.

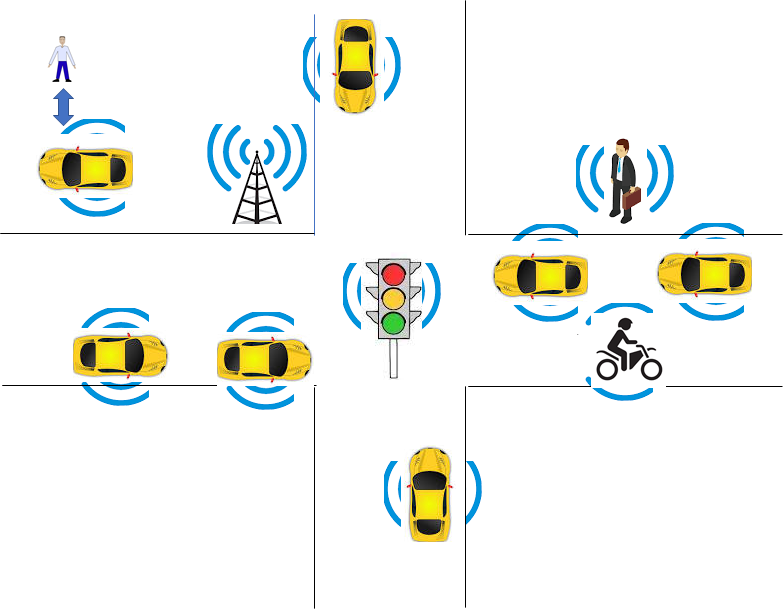
Other vehicles (not CAVs), such as electric scooters, bicycles.

Pedestrians whose personal device ransmits their Pose.

Fixed equipment (e.g., traffic light, bus stop, roadside units)

### Reference architecture

*Figure 7* depicts the environment with which a CAV interacts.



*Figure 7 – The CAV-to-Everything environment*

### Input and output data

#### CAVs within range

*Table 5* gives the data types a CAV broadcasts to nearby CAVs. The Simple and Full World Representation data are exchanged to enable all relevant CAVs to share a common volumetric model of the relevant portion of the world.

*Table 5 – MPAI-CAV Interaction with Environment data*

|  |  |
| --- | --- |
| **Data type** | **Description** |
| CAV identity | Digital equivalent of today’s plate number |
| CAV route data | Pose, velocity, acceleration, trajectory etc. |
| Basic World Representation | Objects with IDs, local CAV coordinates of the objects seen by CAV and bounding boxes, traffic signalisation (e.g., traffic signals, roundabouts). |
| Misalignment of Basic World Representation | IDs and coordinates of objects whose existence is not shared. |
| Full World Representation | CAV’s fused world representation of the CAV and CAVs in range. |
| Emergency signal | E.g., passenger health, mechanical problem, weather, environ­ment evacuation. |
| Events | E.g., works, traffic jams etc. |

#### Other vehicles (not CAVs)

Other vehicles can be scooters, motorcycles, bicycles, other non-CAV vehicles.

They transmit their position as derived from GNSS.

#### Pedestrians

Their smartphones can transmit their coordinates as available from GNSS.

#### Fixed equipment

Fixed equipment are traffic lights, bus stops, roadside units.

**Traffic lights** can broadcast:

geographic coordinates.

state (Green-Yellow-Red), time to change state.

lane markings.

speed limits.

pedestrian crosswalks

general traffic information.

**Roadside units** can broadcast:

1. Identity and coordinates (exact coordinate reference)
2. Full World Description (without moving objects) recorded in a roadside unit and regularly updated via download (this could be part of the offline map).

### AI Modules

There is only one AIM sending and receiving the data described above.

## Motion Actuation Subsystem (MAS)

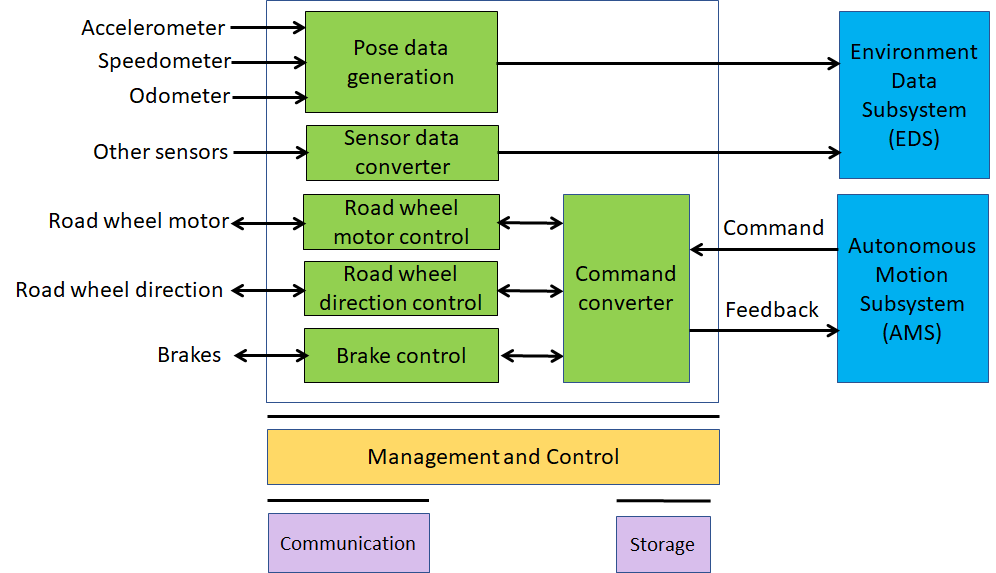
### Use Case description

The Motion Actuation Subsystem is in charge to

1. Receive instructions from Autonomous Motion Subsystem.
2. Translate instructions into specific commands to its own subsystems, e.g., road wheels, accelerator, brakes.
3. Receive feedbacks from its subsystems.
4. Package feedbacks into high-level information.
5. Send packaged information to Autonomous Motion Subsystem.
6. Transmit information gathered from its subsystems to the Environment Sensing Subsystem.

### Reference architecture

The Motion Actuation Subsystem reference model is given by *Figure 8*.



*Figure 8 – The Motion Actuation Subsystem Reference Model*

*Figure 8* includes data from other sensors that are relevant to the motion of the CAV in the environment, e.g., air humidity, temperature etc.

### Input and output data

**Input**

1. Position data generation
2. Other sensor data
3. Command control
4. Road wheel motor data
5. Road wheel direction data
6. Brake data

Output

1. Odometer data
2. Velocity data
3. Acceleration data
4. Road wheel motor data
5. Road wheel direction data
6. Brake data

### AI Modules

The AI Modules of Autonomous Motion Subsystem are given in *Table 4*.

*Table 6 – AI Modules of Motion Actuation Subsystem*

|  |  |
| --- | --- |
| **AIM** | **Function** |
| **Position data generator** | Delivers data produced by odometer |
| **Velocity data generator** | Delivers data produced by Speedometer |
| **Acceleration data generation** | Delivers data produced by Accelerometer |
| **Data converter** | Provides other data from the environment in a standard usable form |
| **Command converter** | Converts commands into specific actuation commands to Road wheel motor control, Road wheel direction control, Brake control |
| **Road wheel motor control** | Actuates the commands from Command & Control, provides feedback |
| **Road wheel direction control** | Actuates the commands from Command & Control, provides feedback |
| **Brake control** | Actuates the commands from Command & Control, provides feedback |

# Technologies and Functional Requirements

The Functional Requirements refer to the individual technologies identified as necessary to implement MPAI-CAV Use Cases belonging using AIMs operating in an MPAI-AIF AI Framework. The Functional Requirements developed in this document adhere to the following guidelines:

AIMs are defined to allow implementations by multiple technologies (AI, ML, DP)

DP-based AIMs need interfaces, e.g., to a Know­ledge Base. AI-based AIM will typically require a learning process, however, support for this process is not included in the document. MPAI may develop further requirements covering that process in a future document.

AIMs can be aggregated in larger AIMs. Consequently, some data flows of aggregated AIMs may no longer be accessible.

## Human-CAV Interaction

### I/O Data summary

For each AIM (1st column), *Table 7* gives the input (2nd column) and the output data (3rd column).

*Table 7 – I/O data of Human-CAV Interaction AIMs*

|  |  |  |
| --- | --- | --- |
| **AIM** | **Input Data** | **Output Data** |
| **Object and gesture analysis** | Video | Object ID  Gesture ID  Emotion |
| **Face analysis** | Video | Emotion  Meaning |
| **Face identification** | Video | Face ID |
| **Speech separation** | Audio | Speech |
| **Speech recognition** | Audio | Text  Emotion |
| **Speaker identification** | Audio | Speaker ID |
| **Language understanding** | Object identifier  Text | Text  Meaning |
| **Emotion recognition** | Emotion (speech)  Emotion (face)  Emotion (gesture) | Final emotion |
| **Question analysis** | Meaning (speech)  Meaning (face)  Meaning (gesture) | Intention |
| **Question and dialogue processing** | Speaker ID  Face ID  Intention  Text  Final emotion | Synthetic speech  Face animation  Text |
| **Speech synthesis** | Reply (speech) | Speech |
| **Face animation** | Reply (face) | Face animation |

MPAI has acquired a set of first-generation technologies related to the data types listed below in a previous Call for Technologies for the MPAI-MMC standard [13]. MPAI is ready, however, to consider new technologies related to the data below to support new requirements and/or to enhance capabilities.

### Text

As there is a need to support most languages in use, Text representation conforms to ISO/IEC 10646, Information technology – Universal Coded Character Set (UCS).

**To respondents**

Respondents are invited to comment on this choice.

### Audio

Audio is sampled from an analogue source (passenger compartment) at a frequency in the 44.1-96 kHz range with at least 16 and at most 24 bits/sample.

**To respondents**

Respondents are invited to comment on this choice.

### Video

Video is intended for use in the passenger compartment.

Video has adopted the following characteristics of 2D Video.

1. Pixel shape: square
2. Bit depth: 8-10 bits/pixel
3. Aspect ratio: 4/3 and 16/9
4. 640 < # of horizontal pixels < 1920
5. 480 < # of vertical pixels < 1080
6. Frame frequency 50-120 Hz
7. Scanning: progressive
8. Colorimetry: ITU-R BT709 and BT2020
9. Colour format: RGB and YUV
10. Compression: uncompressed, if compressed AVC, HEVC

**To respondents**

Respondents are invited to comment on MPAI’s choice for 2D Video.

Respondents are requested to propose a data format for video+depth.

### Object identifier

Report choice of MQA.

### Emotion

Report choice of CAE and MMC.

### Meaning

Report choice of MMC.

### Intention

Report choice of MMC.

### Reply (speech)

Report choice of MMC.

### Reply (face)

MPAI has already adopted a format for lips animation (MMC).

Here MPAI is looking for a technology that enables head and face animation capable to represent

1. Motion of head when speaking
2. Motion of face muscles and eyeballs
3. Turning of gaze to a particular person

### Human commands

The commands given to the Autonomous Motion Subsystem are:

1. Go to a Waypoint
2. Park close to a Waypoint
3. Go faster
4. Drive slowly
5. Display FWR

**To respondents**

Respondents are requested to propose a coded representation of the above commands. The addition of coded representation of new commands is welcome.

### AMS response

The responses of the Autonomous Motion Subsystem are:

1. The desired Waypoint has been reached

**To respondents**

Respondents are requested to propose a coded representation of the above commands. The addition of coded representation of new responses is welcome.

### Face identity

The Face Identity AIM shall be able to recognise the identity of a limited number of faces.

**To respondents**

Respondents are requested to propose a face identification system suitable for a limited number of faces.

### Speaker identity

The Speaker Identity AIM shall be able to recognise the identity of a limited number of speakers.

**To respondents**

Respondents are requested to propose a face identification system suitable for a limited number of speakers.

### Full World Representation

The requirements of the FWR AIM are developed in the context of Autonomous Motion Subsys­tem requirements.

### FWR interaction

The requirements of FWR interaction should be developed once the FWR requirements will be defined.

## Environment Sensing Subsystem

### I/O Data summary

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 data*

|  |  |  |
| --- | --- | --- |
| **CAV AIM** | **Input** | **Output** |
| **Vehicle** | GNSS | State |
| Odometer |
| Offline Maps |
| **Environment Recorder** | State | -- |
| Volumetric data (TBD) |
| Environment conditions (TBD) |
| **GNSS coordinate extractor** | GNSS data | Global coordinates |
| **Odometer coordinate extractor** | Odometer data | Differential coordinates |
| **Radar Processor** | Radar data | Visual Objects and Scene |
| **Lidar Processor** | Lidar data | Visual Objects and Scene |
| **Ultrasound Processor** | Ultrasound data | Visual Objects and Scene |
| **Cameras Processor** | Camera data | Visual Objects and Scene |
| **Microphone Sound Processor** | Microphone data | Sound Objects and Scene |
| **Traffic Signalisation Detector** | State | Traffic signals  Traffic rules |
| **Moving Objects Tracker** | State | Objects’ states |
| Basic World Representation |
| **Basic World Representation Fusion** | State | Basic World Representation |
| Offline maps |
| (Static and) moving objects |
| Traffic signals |

### Global Navigation Satellite System (GNSS)

GNSSs provide spatial information with different accuracies. GNSS can only be relied on when reception conditions are above a certain level. This excludes GNSS in tunnels or urban canyons.

**To Respondents**

Some data formats are:

1. GPS Exchange Format (GPX) provides an XML schema providing a common GPS data format that can be used to describe waypoints, tracks, and routes.
2. World Geodetic System (WGS) includes the definition of the coordinate system's fundamental and derived constants, the ellipsoidal (normal) Earth Gravitational Model (EGM), a description of the associated World Magnetic Model (WMM), and a current list of local datum transfor­mations.
3. International GNSS Service (IGS) SSR is a format used to disseminate real-time products to support the IGS (igs.org) Real-Time Service. The messages support multi-GNSS and include corrections for orbits, clocks, DCBs, phase-biases and ionospheric delays. Extensions are planned to also cover satellite attitude, phase centre offsets and variations and group delay variations.

Respondents are requested to propose a single GNSS data format.

### Odometer

An odometer converts the number of wheel rotations times the tire circumference (π x tire diameter) from the start up to the point being considered as the distance travelled.

**To Respondents**

Respondents are requested to propose a single Odometer data format.

### Speedometer

A speedometer is an electronic sensor that measures the instantaneous speed of a CAV.

**To Respondents**

Respondents are requested to propose a single Speedometer data format.

### Accelerometer

An accelerometer is an electronic sensor that measures the acceleration forces acting on a CAV. An accelerometer measures proper acceleration, i.e., the acceleration of a body in its own instantaneous rest frame, not to be confused with coordinate acceleration, i.e., acceleration in a fixed coordinate system. Therefore, an accelerometer at rest on the surface of the Earth measures an acceleration straight upwards of g ≈ 9.81 m/s2. In free fall (falling toward the centre of the Earth at ≈ 9.81 m/s2) measures zero.

**To Respondents**

Respondents are requested to propose a single Accelerometer data format.

### Radio Detection and Ranging (RADAR)

RADAR, LiDAR and ultrasound are based on “time-of-flight”, i.e., they measure distance and speed based on the time it takes for a signal to hit an object and be reflected back.

Radar operates in the mm range. Radar can detect vehicles (CAVs and trucks) because they typically reflect Radar signals while objects that are smaller and have less reflectance, e.g., pedestrians and motorcycles have a poor reflectance. In a busy environment, vehicles’ reflections can swamp that from a motorcycle; a child next to a vehicle can go undetected, while a can may produce an image out of proportion to its size.

Main features

1. Measure distance.
2. Independent of environment.
3. Low resolution (objects detected, not classified).
4. Short range radar in the 25 GHz band, distance is computed.
5. Long range radar in the 76-77 GHz, detects objects and measures speed @ ≤ 250 m. Typical ranges of long-range radar (LRR) systems are 80-200 m. The antenna is small because the wavelength is ~3.5-4 mm. Atmospheric absorption limits interference with other systems. A multitask 94-GHz pulse Doppler radar has 25-cm radial and 1.5 degrees angular resolution

Frequency Modulated Continuous Wave (FMCW) is the dominant radar technology. A radar TX sends a sinusoidal carrier with a frequency that increases then decreases periodically over time (chirp). The difference in frequency between the RX signal (rep­resenting an object) and the TX signal is proportional to travel time and a measure of the distance. The relative velocity of an object is obtained by comparing the Doppler frequency shifts of the chirp’s increasing and decreasing frequency portions.

Radar sensors build a representation of the environment based on the observation of complex, scattered radio waves, from which information of an object’s distance and velocity can be derived.

**To Respondents**

Known Radar data formats include [19]:

1. OPERA BUFR format (Paulitsch et al., 2010).
2. hdf5 formats (Michelson et al., 2011).
3. NetCDF files generated by the commercial EDGE software.
4. hdf5 files generated by the commercial GAMIC software.
5. German Weather Services quantitative local scan format (DX).
6. Quantitative composite format (RADOLAN, see German Weather Service, 2004).

As CNNs can be used to predict 3D object orientation of a single vehicle using multichannel radar data, respondents are invited to propose a format of Radar images that facilitates identification, tracking and representation of objects.

### Light Detection and Ranging (LIDAR)

Like Radar, LiDAR is an active sensor. Unlike Radar, however, it operates in the µm range. It sends an electromagnetic signal and receives the reflected signal back. A typical eye-safe LiDAR:

1. Has a frequency of ~200 THz and a wavelength ~1.5 µm (the visible range is 0.4 to 0.75 µm).
2. Measures the range in each pixel (called also voxels).
3. Pixel grayscale is measured by the intensity variation of the reflected light.
4. The colour of an object can be measured by using more than one wavelength.
5. Velocity can be measured using the Doppler shift in frequency due to motion, or by measuring the position at different times.
6. Micro-motion can be measured using the Doppler shift measured with a coherent LiDAR.
7. Produces 100 kpoints/frame or 1.35 Mbytes: 32\*3 bits (coordinates) +16 bits (ref­lectance). Today 200 kpoints/frame are reasonable.
8. Angular resolution is 0.1º and the vertical field is 40º.
9. A Lidar scan captured at 25 fps is 270 Mbit/s or 33.75 Mbytes/s.

**To Respondents**

The LAS (LASer) format is a binary file format for LiDAR point cloud data specified by the American Society for Photogrammetry and Remote Sensing (ASPRS) [15].

Pcap is a well-established data format for Lidar scans [16, 17, 18]. Other formats are listed in [20].

Respondents are invited to provide a LiDAR data format that facilitates identification, tracking and representation of objects.

### Cameras (2D and 3D)

1. Passive measuring detects objects only if they are illuminated.
2. Do not provide distance information
3. 3D camera signals can be processed to provide stereo information.

**To Respondents**

Respondents are invited to provide a data format for RGB-D cameras.

### Ultrasound

1. Operates at 20 kHz
2. Independent of environment
3. Low resolution
4. limited range (≤ 10 m)

**To Respondents**

The Ultrasound File Format initiative has defined the Ultrasound File Format (UFF) format [14].

Respondents are invited to propose an ultrasound format that facilitates identification, tracking and representation of sound objects.

### Microphones

Microphones are used to capture the external sound, (e.g., for noise suppression inside the passen­ger compartment, but also to add the sound dimension to the Full World Representation.

**To Respondents**

Respondents are requested to propose a format of sound from multiple microphones that is suitable for processing the captured sound field for CAV purposes.

### State

State is the set of the following CAV attributes at a given time:

1. Pose, Velocity and Acceleration
2. Orientation, Angular Velocity and Angular Acceleration.

**To Respondents**

Respondents are requested to propose a State data format suitable for use in CAVs.

### Offline maps

Offine maps or HD maps or 3D maps are roadmaps with cm-level accuracy and a high environmental fidelity. They report the exact positions of pedestrian crossings, traffic lights/signs, barriers and more.

Navigation Data Standards [22] calls itself “The worldwide standard for map data in automotive eco-systems”. The NDS specification covers data model, storage format, interfaces, and protocols.

SharedStreets [26] Referencing System is a global non-proprietary system for describing streets.

**To Respondents**

Respondents are requested to propose an Offline data format.

### Basic World Representation

Data from different information sources, e.g., CAV’s Environment sensors, CAVs in range and Offline maps are combined to one comprehensive Basic World Representation (BWR) [23]. The BWR ensures that all CAV functions base their decisions on the same knowledge base, thus ensuring system consistency.

The requirements of the BWR are:

1. The BWR shall be represaented as
2. All real objects that impact the path decision process in the Decision Horizon Time shall be represented in the BWR
3. Each object in the BWR shall be described by
   1. Its ID.
   2. Its State.
   3. Its physical characteristics, e.g., static or dynamic.
   4. Its bounding box (as a minimum).
   5. Its semantics (e.g., other CAVs).
4. Each element of the BWR shall have an associated accuracy estimate.
5. The ground (roads etc.) shall be describes with all traffic signalisations, including roads and lanes lane geometry, topology, and lane-specific traffic rules.
6. The BWR shall have the ability to adapt to the level of structuredness of the Environment.

**To Respondents**

Respondents are requested to propose a Basic World Representation data format satisfying the requirements.

## Autonomous Motion Subsystem

### Summary of Autonomous Motion Subsystem data

*Table 9* gives, for each AIM (1st column), the input (2nd column) and the output data (4th column).

*Table 9 – CAV Autonomous Motion Subsystem data*

|  |  |  |
| --- | --- | --- |
| **CAV AIM** | **Input** | **Output** |
| **Route Planner** | State  Destination | Route  Estimated time |
| **Full World Representation Fusion** | State | Full World Representation |
| Offline Maps |
| Basic World Representations |
| Other Environment Data |
| **Path Planner** | State | Set of Paths |
| Route |
| Traffic Rules |
| **Behaviour Selector** | State | Path |
| Route |
| Full World Representation |
| **Motion planner** | Path | Trajectory |
| **Obstacle Avoider** | Full World Representation Trajectory | Trajectory |
| **Command** | Trajectory | Actuations |

### User input data

Text. To be further discussed.

### Offline map

See above

### State

See above

### Goal

Is a particular Pose

### Route

Find how a route is formatted

### Basic World Representation

See above

### Traffic signals

Format to represent traffic signal on a road

### Traffic rules

The traffic rules should be digitally represented to realise a route [25].

### Pose

See above

### Velocity

See above

### Full World Description

The elements of the FWD are:

1. appropriate portion of the offline map.
2. Physics of the environment: weather, temperature, air pressure, ice and water on the road).
3. For each object: ID, position, velocity, acceleration bounding box (more than a box, if available), semantics, flags (e.g., warning).
4. For CAVs, the path, i.e., the attitude and a series of future poses, the shape of the body.
5. Road structure.
6. Traffic signalisation.
7. Local traffic rules

### Path

Segment or sequence of segments

### Trajectory

Start (coordinates, velocity, acceleration) end (coordinates, velocity, acceleration)

## CAV to Everything

### Summary of CAV to Everything data

*Table 9* gives, for each AIM (1st column), the input data (2nd column) from which AIM (3rd column) and the output data (4th column).

*Table 10 –CAV to Everything data*

|  |  |  |  |
| --- | --- | --- | --- |
| **CAV AIM** | **Input** | **From** | **Output** |
| V2X | Transmission request | Store | CAV identity |
| V2X | Transmission request | AMS | Attitude-Path-Trajectory |
| V2X | Transmission request | ESS | Basic World Representation |
| V2X | Transmission request | AMS | Full World Representation |
| V2X | Transmission request |  | Events |

### CAV identity

### Attitude-Path-Trajectory

### Basic World Representation

As in Autonomous Motion Subsystem?

### Full World Representation

As in Autonomous Motion Subsystem?

### Events

## Motion Actuation Subsystem

### Summary of Motion Actuation Subsystem data

*Table 9* gives, for each AIM (1st column), the input data (2nd column) from which AIM (3rd column) and the output data (4th column).

*Table 11 –Motion Actuation Subsystem data*

|  |  |  |  |
| --- | --- | --- | --- |
| **CAV AIM** | **Input** | **From** | **Output** |
| Command converter | Command | Command and Control (AMS) | Road wheel motor control  Road wheel direction control  Brake control |
| Command converter | Feedback | Road wheel motor control  Road wheel direction control  Brake control | Command and Control (AMS) |
|  | Position | Position data generation | Environment Sensing Subsystem |
|  | Velocity | Velocity data generation | Environment Sensing Subsystem |
|  | Acceleration | Acceleration data generation | Environment Sensing Subsystem |

### Command

### Road wheel motor control

### Road wheel direction control

### Brake control

### Feedback to Command

### Odometer data

### Speedometer data

### Accelerometer data

### Position

### Velocity

### Acceleration

# Data privacy

A CAV can generate or acquire data for which privacy is an important characteristic. here are some of the functions potentially affected by data privacy or that are liable to become accessible to authorities. e.g., police, judiciary.

## Human-CAV Interaction (HCI)

By having interactions with humans, HCI becomes aware of potentially sensitive information, e.g.:

1. Result of monitoring the passenger compartment.
2. Minute requests from humans, e.g., go to a way point, display Full World Representation, turn off air conditioning, etc.
3. Dialogue with human,

## Environment Sensing Subsystem (EDS)

EDS collects large amoung of environment data for the purpose of creating instantaneous Basic World Representations, e.g.:

1. GNSS gives the position of the CAV and of whatever is perceived by the CAV that is approximate, but sufficiently precise for my uses.
2. Radar, Lidar, Ultrasound give variously defined information about what is in the environment surrounding the CAV.
3. Cameras give a 360° panoramic view of the environment where all objects, save those occluded, are visible.
4. External microphones give a complete representation of the external sound field.

A user could create a permanet and certified recording of important data acquired by EDS.

The environment recorder could compress and record all data acquired for a limited amount of time. Some data could be recorded for a longer time.

## Autonomous motion subsystem (AMS)

AMS knows the exact waypoints the CAV has passed through and all the commands given to the Motion Actuation Subsystem.

By integrating the Basic World Representations of all CAVs in range and its own, a CAV can create a pretty detailed and extended map of the environment.

## CAV to Everything (V2X)

V2X acquires the identity of the CAVs in range and communicates appropriate subsets of the Basic and Full Worlds Representations.

## Motion Actuation Subsystem (MAS)

MAS acquires position information through its Inertial Measurements Unit.

# Annex 1 – General MPAI Terminology

*Table 10 – 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 Workflow (AIW) | An organised aggregation of AIMs implementing a Use Case receiving AIM-specific Inputs and producing AIM-specific Outputs according to its Function. |
| AI Module (AIM) | A processing element receiving AIM-specific Inputs and producing AIM-specific Outputs according to according to its Function. |
| Application | A usage domain target of an Application Standard |
| Channel | A connection between an output port of an AIM and an input port of an AIM. The term “connection” is also used as synonymous. |
| Communication | The infrastructure that implements message passing between AIMs |
| Component | One of the 7 AIF elements: Access, Communication, Controller, Internal Storage, Global Storage, 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 format | The standard digital representation of data and their semantics. |
| 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. |
| 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. |
| Internal Storage | A Component to store data of the individual AIMs. |
| Identifier | A name that uniquely identifies an Implementation. |
| Implementation | 1. An embodiment of the MPAI-AIF Technical Specification, or 2. An AIW or AIM of a particular Level (1-2-3) conforming with a Use Case of an MPAI Applic­ation Standard. |
| Interoperability | The ability to functionally replace an AIM with another AIM 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 proprietary (Level 1) or to pass the Conformance Tes­ting (Level 2) or the Perform­ance Testing (Level 3) of an MPAI Applic­ation Standard. |
| Knowledge Base | Structured and/or unstructured information made accessible to AIMs via MPAI-specified interfaces |
| Message | A sequence of Records transported by Communication through Channels. |
| Normativity | The set of attributes of a technology or a set of technologies specified by the applicable parts of an MPAI standard. |
| Performance | The attribute of an Implementation of being Reliable, Robust, Fair and Replicable. |
| 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 |
| Profile | A particular subset of the technologies used in MPAI-AIF or an AIW of an Application Standard and, where applicable, the classes, other subsets, options and parameters relevant to that subset. |
| Record | A data structure with a specified structure |
| Reference 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 Performamce Assessor, can be replicated, within an agreed level, by another Performamce Assessor. |
| Robustness | The attribute of an Implementation that copes with data outside of the stated application scope with an estimated degree of confidence. |
| Service Provider | An entrepreneur who offers an Implementation as a service (e.g., a recommendation service) to Users. |
| 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 AI Framework.  (Application) the normative specification of the set of Use Cases belonging to an Application Domain along with the AIMs required to Implement the Use Cases. the collection of Use Cases relevant to the Applic­ation Domain that include:   1. The formats of the Input/Output data of the AIWs implementing the Use Cases. 2. The Topology of the AIMs of the AIWs. 3. The formats of the Input/Output data of the AIMs belonging the AIW. |
| 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 MPAI 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 |  |

# Annex 2 – Datasets for CAV research

**nuScenes**

The nuScenes dataset (https://nuscenes.org/) is a large-scale autonomous driving dataset with 3d object annotations. It features:

* Full sensor suite (1x LIDAR, 5x RADAR, 6x camera, IMU, GPS)
* 1000 scenes of 20s each
* 1,400,000 camera images
* 390,000 lidar sweeps
* Two diverse cities: Boston and Singapore
* Left versus right hand traffic
* Detailed map information
* 1.4M 3D bounding boxes manually annotated for 23 object classes
* Attributes such as visibility, activity and pose
* New: 1.1B lidar points manually annotated for 32 classes
* New: Explore nuScenes on SiaSearch
* Free to use for non-commercial use

For a commercial license contact nuScenes@motional.com

nuImages is a large-scale autonomous driving dataset with image-level 2d annotations. It features:

* 93k video clips of 6s each (150h of driving)
* 93k annotated and 1.1M un-annotated images
* Two diverse cities: Boston and Singapore
* The same proven sensor suite as in nuScenes
* Images mined for diversity
* 800k annotated foreground objects with 2d bounding boxes and instance masks
* 100k 2d semantic segmentation masks for background classes
* Attributes such as rider, pose, activity, emergency lights and flying
* Free to use for non-commercial use

**Road Hazard data**

Otonomo real-time Road Hazard data from connected passenger vehicles powers diverse road safety use cases, including mapping, accident predictions, smart cities and many more. The Otonomo Vehicle Data Platform secures, cleanses and normalizes the hazard data to make it more valuable and accessible for diverse use cases.

https://info.otonomo.io/hazard-data-datasheet-lp

# Annex 3 – ETSI Technical Report

ETSI specifies the Collective Perception Service (CPS) in its Technical Report [6]. The CPS includes the format and generation rules of the Collective Perception Message (CPM).

The CPM message format is (H=header, C=container, M=mandatory, O=optional).

*Table 12 – ETSI Collective Perception Message format*

|  |  |  |  |
| --- | --- | --- | --- |
| PDU header | H | M | protocol version, message ID and Station ID. |
| Management | C | M | transmitter type (e.g., vehicle or RSU) and position. |
| Station Data | C | O | transmitter heading, velocity, or acceleration etc. |
| Sensor Information | C | O | transmitter (e.g., speed, heading, or acceleration)  capabilities of the vehicle’s sensors. |
| Perceived Object | C | O | detected objects (e.g., distance, speed and dimensions)  time at which the measurements were done.  A CPM can report up to 128 detected objects |
| Free Space Addendum | C | O | free space areas/volume within the sensor detection areas |

Every 0.1s a CPM is generated if one of the 3 conditions is satified

no CPM has been generated in the last 1s

a new object has been detected

since last CPM sending info about a previously detected object (it must have an ID)

the following attributes have changed:

Absolute position ΔP > 4 m

Absolute speed ΔV > 0.5 m/s

more than 1s has passed (ΔT > 1 s).

ETSI makes use of a common coordinate system. A vehicle can communicate its absolute coordinates roll, pitch and yaw (Attitude).

Different CPM generation rules have been investigated [9].

# Annex 4 – Some CAV Communication Technologies

The following categories of vehicular communication are part of the literature or industry effort:

|  |  |  |
| --- | --- | --- |
| V2V | Vehicle-to-Vehicle | communication between vehicles to exchange information about the speed and position of surrounding vehicles |
| V2I | Vehicle-to-Infrastructure | communication between vehicles and road infrastructure. |
| V2X | Vehicle-to-Everything | communication between a vehicle and any entity that may affect, or may be affected by, the vehicle |
| V2R | Vehicle-to-Roadside | communication between a vehicle and Road Side Units (RSUs). |
| V2P | Vehicle-to-Pedestrian | communications between a vehicle and (multiple) pedestrian device(s) and to other vulnerable road users, e.g., cyclists, in close proximity |
| V2S | Vehicle-to-Sensors | communication between a vehicle and its sesnors on board |
| V2D | Vehicle-to-Device | communication between a vehicle and any electronic device that may be connected to the vehicle itself |
| V2G | Vehicle-to-Grid | communication with the power grid to sell demand response services by either returning electricity to the grid or by throttling their charging rate |
| V2N | Vehicle-to-Network | broadcast and unicast communications between vehicles and the V2X management system and also the V2X AS (Application Server) |
| V2C | Vehicle-to-Cloud | communication with data centers and other devices connected to the internet |

Technologies exist that support at least some aspects of the communivation types of the table:

Radio access, e.g., visible light communication, mmWave, Cellular-V2X, and 5G

Radio resource management (RRM) for vehicular communication using cellular technology

3GPP Release 14: air interfaces and core network technologies to support V2X communic­ation.

Vehicular ad hoc network (VANET)

Dedicated Short-Range Communication (DSRC): 5.9 GHz band with a range of ~300 metres.

Software defined vehicular networks (SDVN)

Internet of vehicles (IoV)

Protocol stack of the intelligent transportation system (ITS)

Cooperative Awareness Messages (CAMs) messages related to the status of CAV’s sent via wireless broadcast in VANETs.

Cooperative or collective perception improve CAV’s perception beyond the sensors’ detection range.

Traffic situation can be extracted from Local dynamic map (LDM) that aggregates CAMs.

# References

SAE International Releases Updated Visual Chart for Its “Levels of Driving Automation” Standard for Self-Driving Vehicles, https://www.sae.org/news/press-room/2018/12/sae-international-releases-updated-visual-chart-for-its-%E2%80%9Clevels-of-driving-automation%E2%80%9D-standard-for-self-driving-vehicles

ISO 8855: "Road vehicles -- Vehicle dynamics and road-holding ability -- Vocabulary"

Rodolfo W. L. Coutinho and Azzedine Boukerche, Guidelines for the Design of Vehicular Cloud Infrastructures for Connected Autonomous Vehicles, IEEE Wireless Communications - August 2019

Claudine Badue, Rânik Guidolini, Raphael Vivacqua Carneiro, Pedro Azevedo, Vinicius B. Cardoso, Avelino Forechi, Luan Jesus, Rodrigo Berriel, Thiago M. Paixão, Filipe Mutz, Lucas de Paula Veronese, Thiago Oliveira-Santos, Alberto F. De Souza; Self-driving cars: A survey; Expert Systems With Applications 165 (2021) 113816

D. Cireşan, U. Meier, J. Masci, and J. Schmidhuber, “Multi-column deep neural network for traffic sign classification,” Neural Netw., vol.32, pp.333–338, Aug. 2012

ETSI TR 103 562 V2.1.1 (2019-12), Analysis of the Collective Perception Service (CPS); Release 2.

CAR 2 CAR Communication Consortium, https://www.car-2-car.org/

Usman Ali Khan and Sang Sun Lee; Distance-Based Resource Allocation for Vehicle-to-Pedestrian Safety Communication; https://www.mdpi.com/2079-9292/9/10/1640/pdf

Gokulnath Thandavarayan, Miguel Sepulcre, and Javier Gozalvez; Generation of Cooperative Perception Messages for Connected and Automated Vehicles; IEEE Transactions on Vehicular Technology, Vol. 69, No. 12, December 2020

Pranav Kumar Singhab, Sunit Kumar Nandiac, Sukumar Nandi; A tutorial survey on vehicular communication state of the art, and future research directions; Vehicular Communications Volume 18, August 2019, 100164

https://phantom.ai/assets/uploads/PAI%20Renesas%20Partnership%20Announcement%20(1).pdf

Charles R., Qi Li, Yi Hao Su, Leonidas J. Guibas; PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space; <https://arxiv.org/pdf/1706.02413.pdf>

MPAI-MMC Call for Technologies, N154, https://mpai.community/standards/mpai-mmc/#CfT

Bernard, Olivier; Bradway, David; Hansen, Hendrik H.G.; Kruizinga, Pieter; Nair, Arun; Perdios, Dimitris; Ricci, Stefano; Rindal, Ole Marius Hoel; Rodriguez-Molares, Alfonso; Stuart, Matthias Bo; The Ultrasound File Format (UFF) - First draft, Proceedings of 2018 IEEE International Ultrasonics Symposium

LAS (LASer) File Format, Version 1.4 , https://www.loc.gov/preservation/digital/formats/fdd/fdd000418.shtml

https://www.lesliesikos.com/pcap/

https://pointclouds.org/documentation/tutorials/hdl\_grabber.html

https://www.mathworks.com/help/vision/ref/velodynefilereader.html

M. Heistermann, S. Jacobi and T. Pfaff,Technical Note: An open source library for processing weather radar data (wradlib), https://hess.copernicus.org/articles/17/863/2013/hess-17-863-2013.pdf

1. https://pro.arcgis.com/en/pro-app/latest/help/data/data-interoperability/supported-formats-with-the-data-interoperability-extension.htm

CDOT and Panasonic Take First Steps to Turn I-70 into Connected Roadway, <https://www.codot.gov/news/2018/july/cdot-and-panasonic-take-first-steps-to-turn-i-70-into-connected-roadway>

Navigation Data Standards, https://nds-association.org/

1. Matthias Schreier; Environment Representations for Automated On-Road Vehicles; https://www.researchgate.net/publication/323105152\_Environment\_Representations\_for\_Automated\_On-Road\_Vehicles
2. Toh Chai Keong, Juan Carlos Cano, Carlos-Javier Fernandez-Laguia, Pietro Manzoni; Wireless Digital Traffic Signs of the Future; IET Networks, September 2018
3. Vladimir Hahanov, Wajeb Gharibi, Eugenia Litvinova, Svitlana Chumachenko, Arthur Ziarmand, Irina Englesi, Igor Gritsuk, Vladimir Volkov, Anastasiia Khakhanova; Cloud-Driven Traffic Monitoring and Control Based on Smart Virtual Infrastructure; 2017-03-28
4. SharedStreets; https://sharedstreets.io/