Human and Machine Communication (MPAI-HMC) V1.1

Leonardo Chiariglione, MPAI

Online presentation, 2024/07/22T15UTC

.

Moving Picture, Audio, and Data Coding by Artificial Intelligence

MPAI: what it is

International, unaffiliated, non-profit SDO.

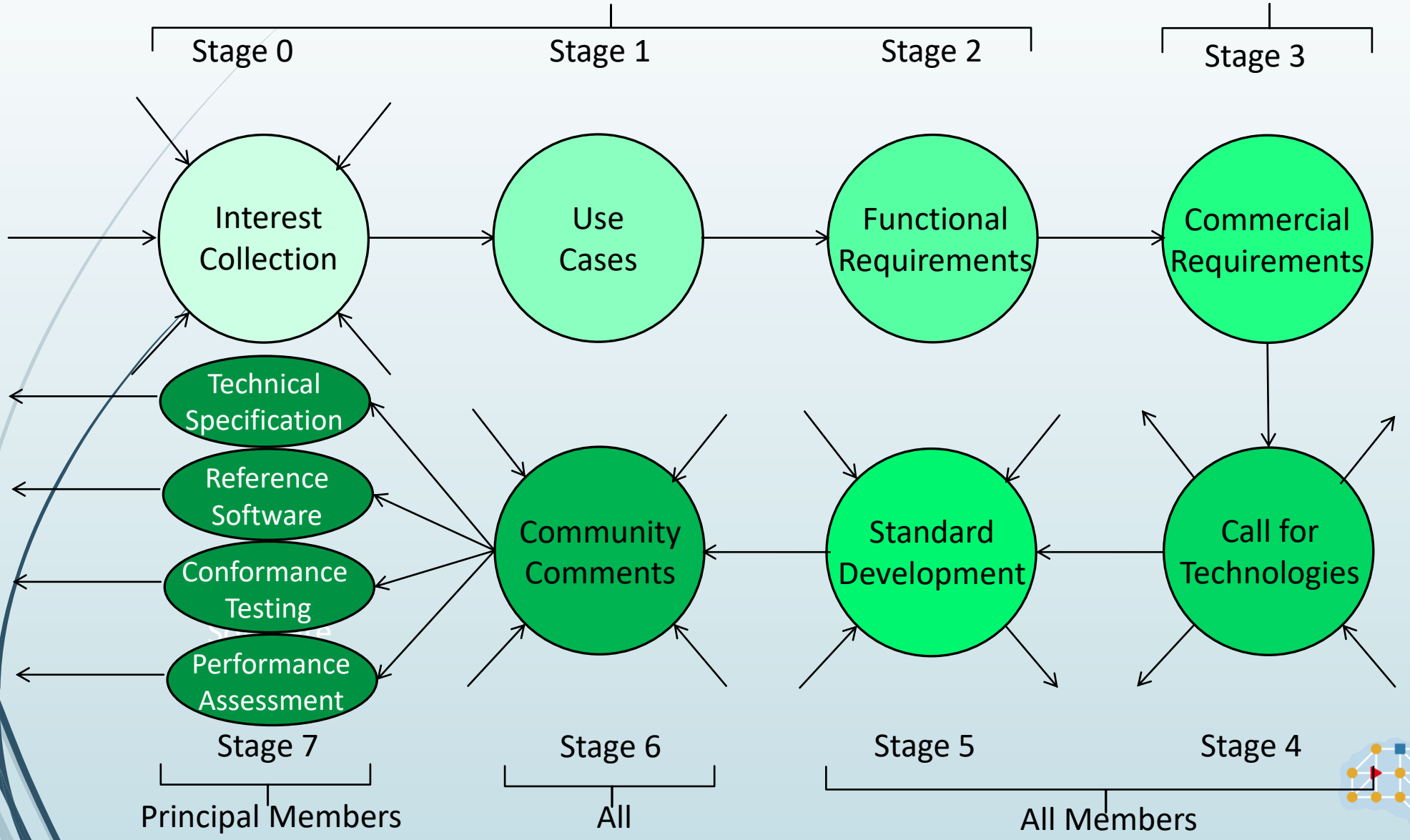
MPAI: what it does

Develops AI-based data coding standards.

MPAI: how it does it

With clear Intellectual Property Rights licensing frameworks.

Rigorous standards development process



MPAI. community

A standard for human and machine communication

- **Human and Machine Communication** enables advanced forms of communication between Entities – humans and Machines in Contexts.
- **Context** is information about a communicating Entity such as language and culture.
- Conditions for humans and machines to communicate:
  - A human is
    - Part of a real space – as when a human communicates with a human
    - Digitally represented in a Virtual Space as a Digital Human in an Audio-Visual Scene
  - A Machine is
    - Represented in a Virtual Space as a Virtual Human in an Audio-Visual Scene
    - Rendered in a real space

How MPAI-HMC does it

- MPAI-HMC enables a real human to communicate with a Machine by representing themselves as a Digitised Human in an Audio-Visual Scene.
- The Term “communication” includes the possibility for a Machine:
  - To understand the semantics of the communication at different layers of depth.
  - To produce a multimodal response congruent with the received information.
- The Machine may act as an “interpreter” by communicating with another Entity in a different Context.
- The Machine runs as a process represented as a speaking Virtual Human in a Virtual Space and rendered as a speaking humanoid in the real space.

Virtual Human

Digitised Human

Machine

In a Virtual Scene or rendered as a real scene

Virtual Human

#2 Virtual Scene

#3 Virtual Scene

Real Human

Real or Digitised Human

#1 Real scene or representation in a Virtual Scene

#4 Real or virtual scene

MPAI. community

What does MPAI-HMC V1.1 standardise?

- Digital representation of audio, visual, and audio-visual objects including humans.
- Digital representation of audio, visual, and audio-visual scenes.
- The format of Communication Items or Portable Avatar – the information exchanged between Entities.
- The interfaces of AI Modules – the bricks used to build HMC-capable Machines.

What is a Portable Avatar (MPAI-PAF)?

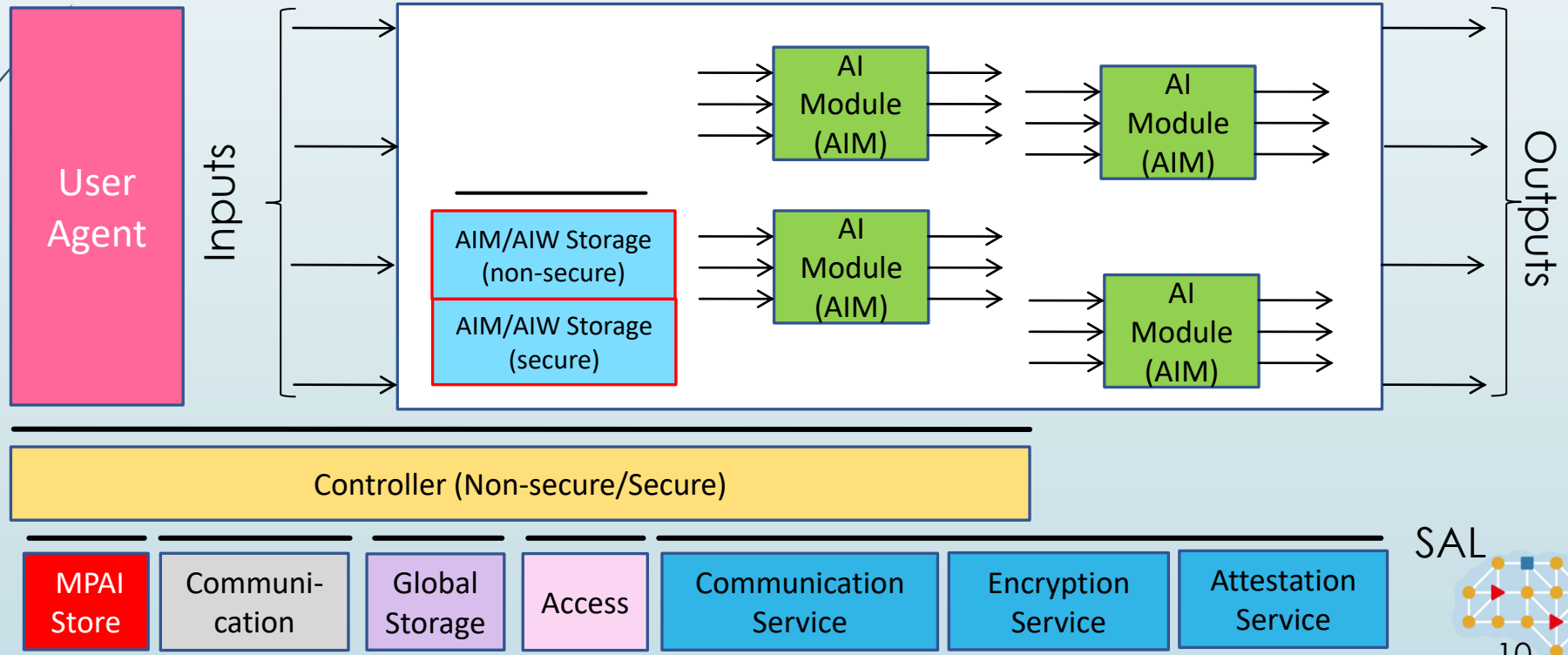
Portable Avatar	
Header	Portable Avatar Header: PAF-PAV-V1.1
MInstanceID	ID of the Virtual Space
PortableAvatarID	ID of the Portable Avatar
AvatarData	Set of Avatar Data
- Avatar	Model – Face Descriptors – Body Descriptors
- Portable Avatar Space-Time	Space-Time information of Avatar in Virtual Space
- Language Selector	Avatar's Language capabilities
- Text Object	Avatar's Text
- Speech Object	Avatar's Speech
- Speech Model	Avatar's Speech Model
- Personal Status	Avatar's Personal Status
Audio-Visual Scene	Set of data related to Audio-Visual Scene
- Audio-Visual Scene Space-Time	Space-Time information referred to Virtual Space
- Audio-Visual Scene Descriptors	Virtual Space's Scene Descriptors

What is an MPAI-HMC V1.1 capable Machine?

- A computing device able to:
  - Capture a human and their Context as Audio-Visual Scene Descriptors.
  - Produce/exchange/interpret Communication Items (Machine-to-Machine)
  - Produce audio-visual scenes that include a rendered Virtual Human (Machine-to-human).
- A Communication Item is a Portable Avatar

How is MPAI-HMC implemented?

- As an AI Workflow of interconnected AI Modules executed in an MPAI AI Framework.
- AIMs are specified by their interfaces and I/O data.
- Only the format of the data is specified.



Structure of MPAI-HMC

- Introduction
- Scope
- Definitions
- References
- Usage Scenarios
- AI Workflows
  - Functions
  - Reference Model
  - Input/Output Data
  - Functions of AI Modules
  - Input/Output Data of AI Modules
  - AIW, AIMs, and JSON Metadata
- AI Modules
- Data Types

A collection of usage scenarios

real human1

AV Scene

AV Scene

real human2

Digitised Human in AV Scene

AV Scene

Virtual Human

AV Scene

Machine1

Machine2

Machine3

Communication Items

AV Scene

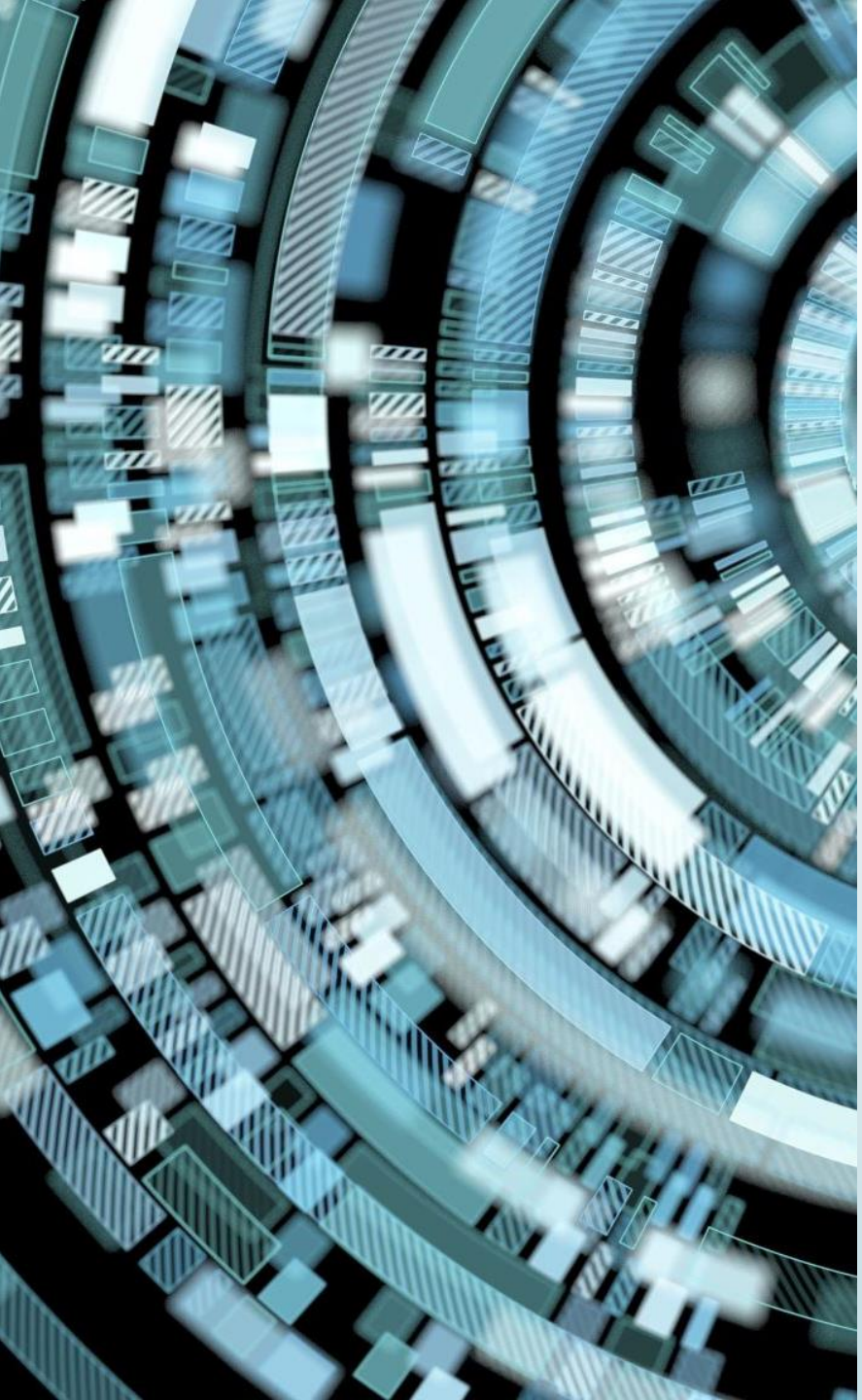
real human3

AV Scene

Communication Items

Machine4

MPAI. community

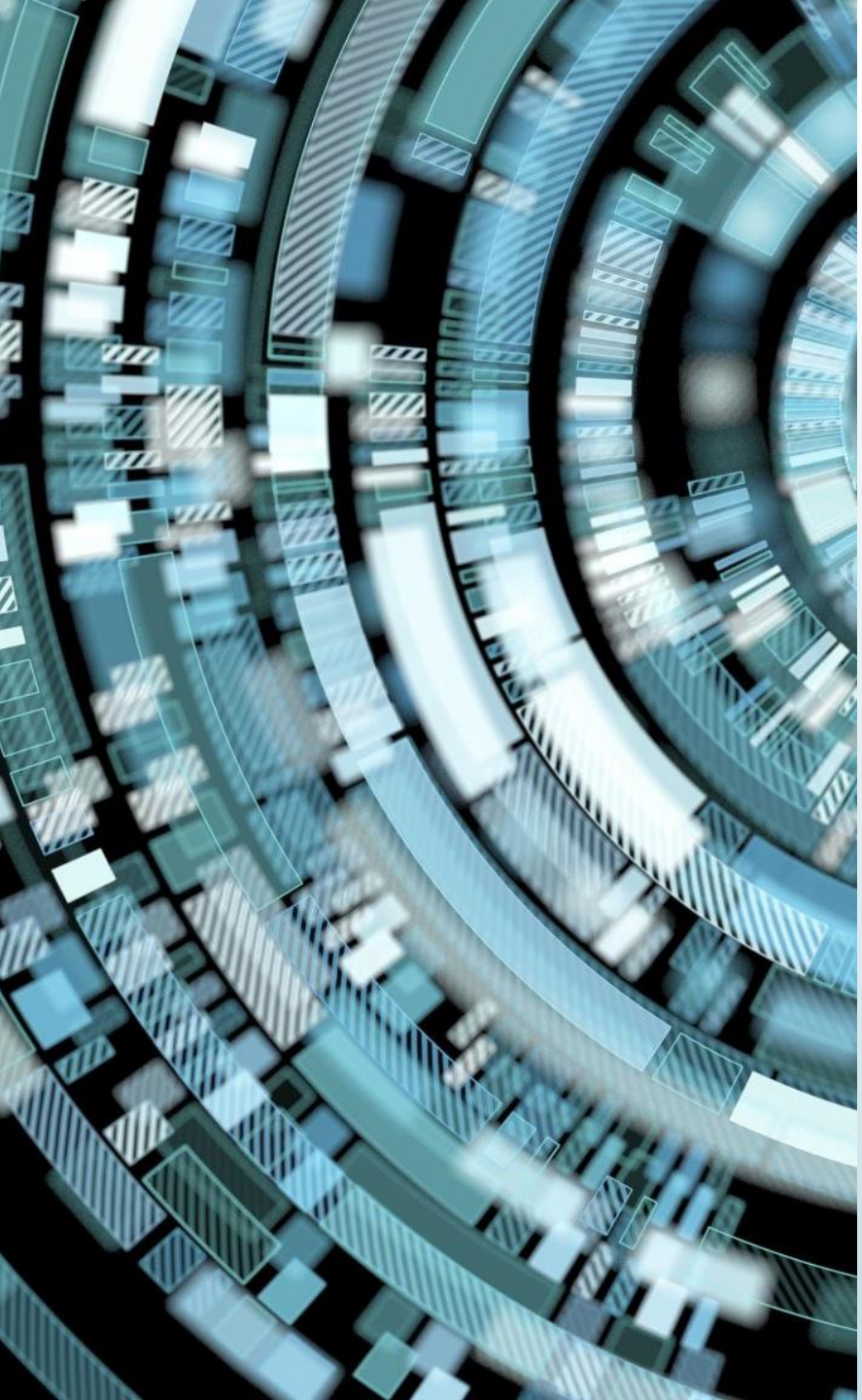


Communicating Entities in Context

Enables Machines to communicate with Entities in different Contexts:

- **Machine** is Software embedded in a device implementing HMC-CEC.
- **Entity** refers to one of:
  1. A human in a real audio-visual scene.
  2. A human in a real scene represented as a Digitised Human in an Audio-Visual Scene.
  3. A Machine represented as a Virtual Human in an Audio-Visual Scene.
  4. A Digital Human in an Audio-Visual Scene rendered as a real audio-visual scene.
- **Digital Human** is either a Digitised Human or a Virtual Human.

MPAI. community



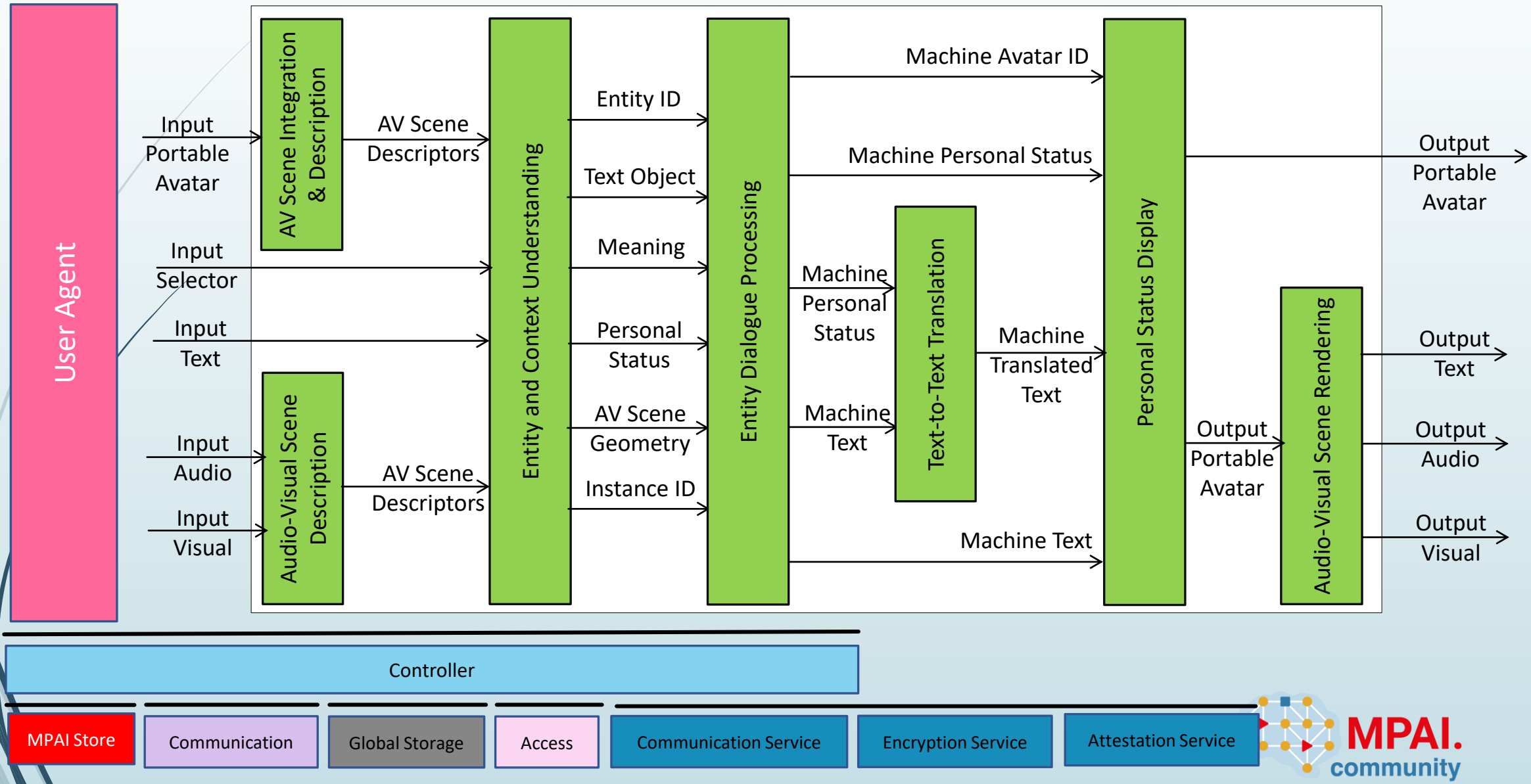
Communicating Entities in Context

Entities communicate in one of the following ways:

1. To **humans** by using:
    1. Entities' body, speech, Context, and audio-visual scenes that the Entities are immersed in.
    2. HMC-CEC-enabled Machines emitting Communication Items.
2. To **Machines** by:
    1. Rendering Entities as speaking humanoids in audiovisual scenes, as appropriate.
    2. Emitting Communication Items.

MPAI. community

Communicating Entities in Context (HMC-CEC)



MPAI. community

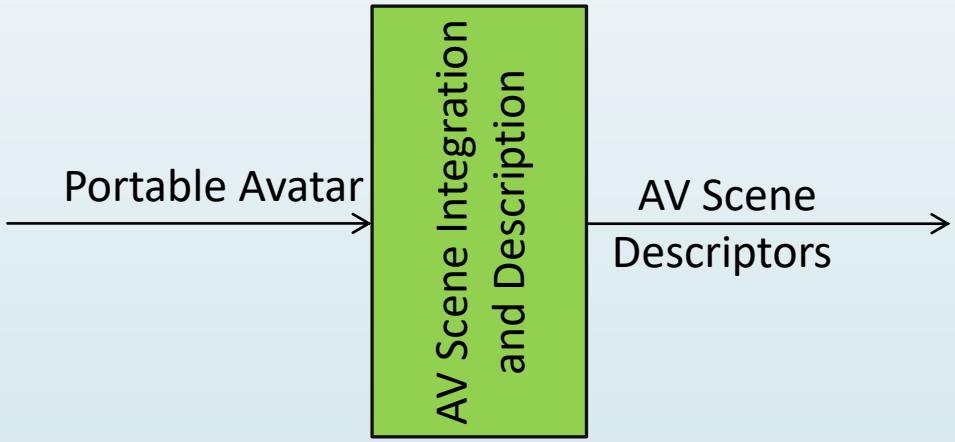
MPAI-HMC Operation

AIM	Receives/Produces
AV Scene Integration & Description	Receives a Portable Avatar that includes a speaking avatar and an AV Scene. Produces AV Scene Descriptors by adding speaking avatar.
Audio-Visual Scene Description	Captures Input Audio and Input Visual. Produces the Audio-Visual Scene Descriptors.
Entity and Context Understanding	Processes information from the Digital Human and its Context. Produces data for downstream processing.
Entity Dialogue Processing	Processes the understood context (scene, words, body etc.) Produce Machine's response (Text and Personal Status).
Personal Status Display	Receives Machine's response. Produces a PA including its speaking avatar in an AV Scene.
Audio-Visual Scene Rendering	Receives Portable Avatar. Renders the Portable Avatar as Output Audio and Output Visual.

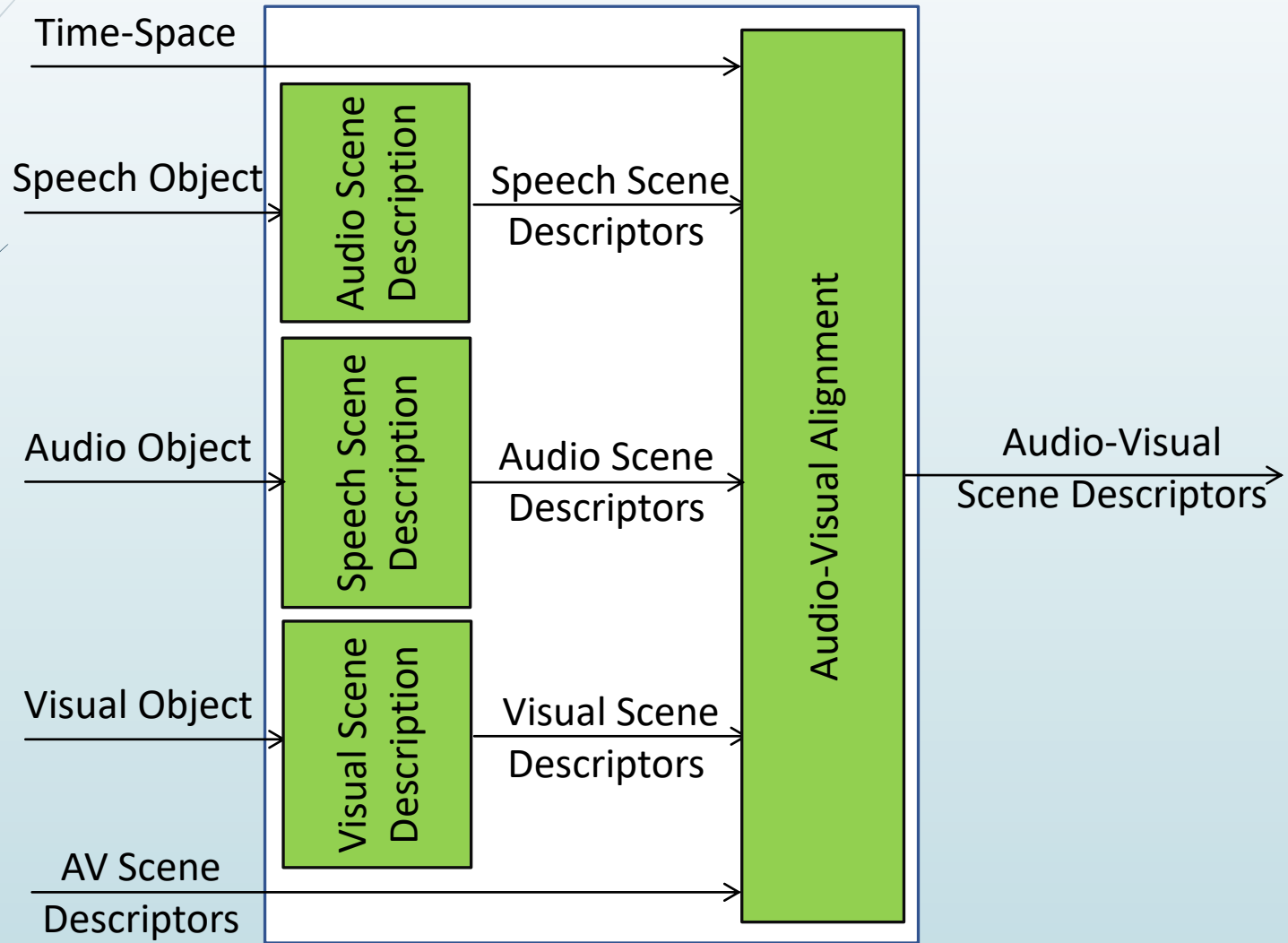
The power of the MPAI AI Framework

- Complex HMC workflow (AIW) broken up into simpler standard Components (AIM).
- AIMs can be Basic or Composite (can include Composite and Basic AIMs).
- Implementer decides whether an AIM is Basic or Composite.
- MPAI-HMC Conformance depends on conforming AIW/AIM and AIW topology
- MPAI-HMC Performance depends on how smart the data processing technologies.
- Enhanced functionalities in future MPAI-HMC versions.

AV Scene Integration & Description (HMC-SID)

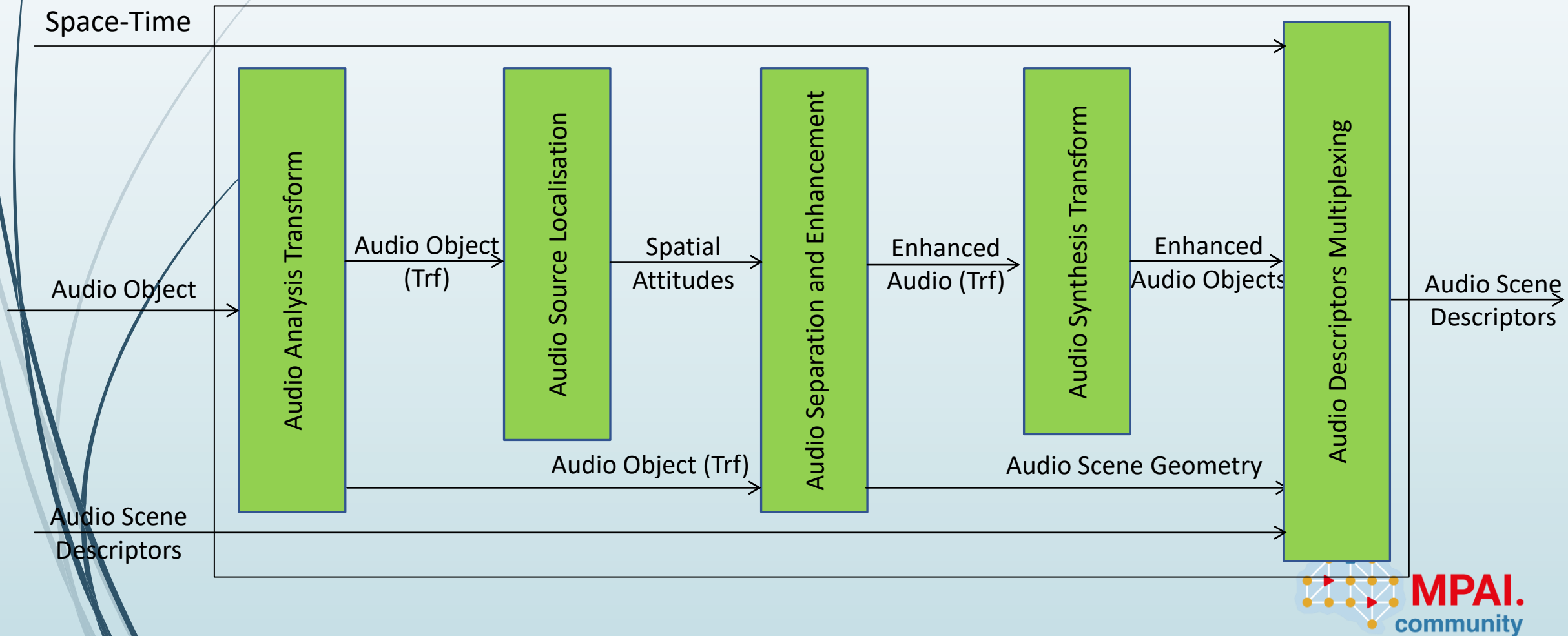


Audio-Visual Scene Description (OSD-VSD) V1.1



MPAI. community

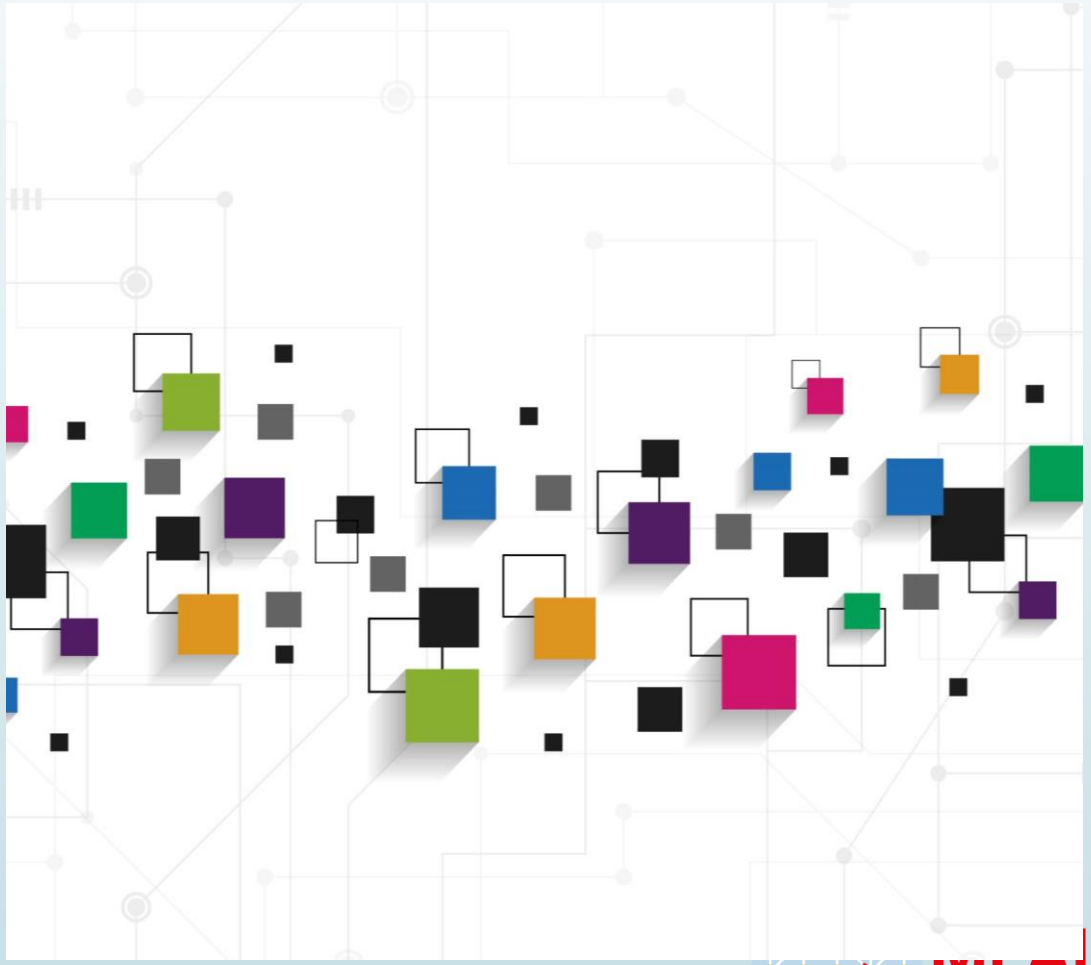
Audio Scene Description (CAE-ASD) V2.2



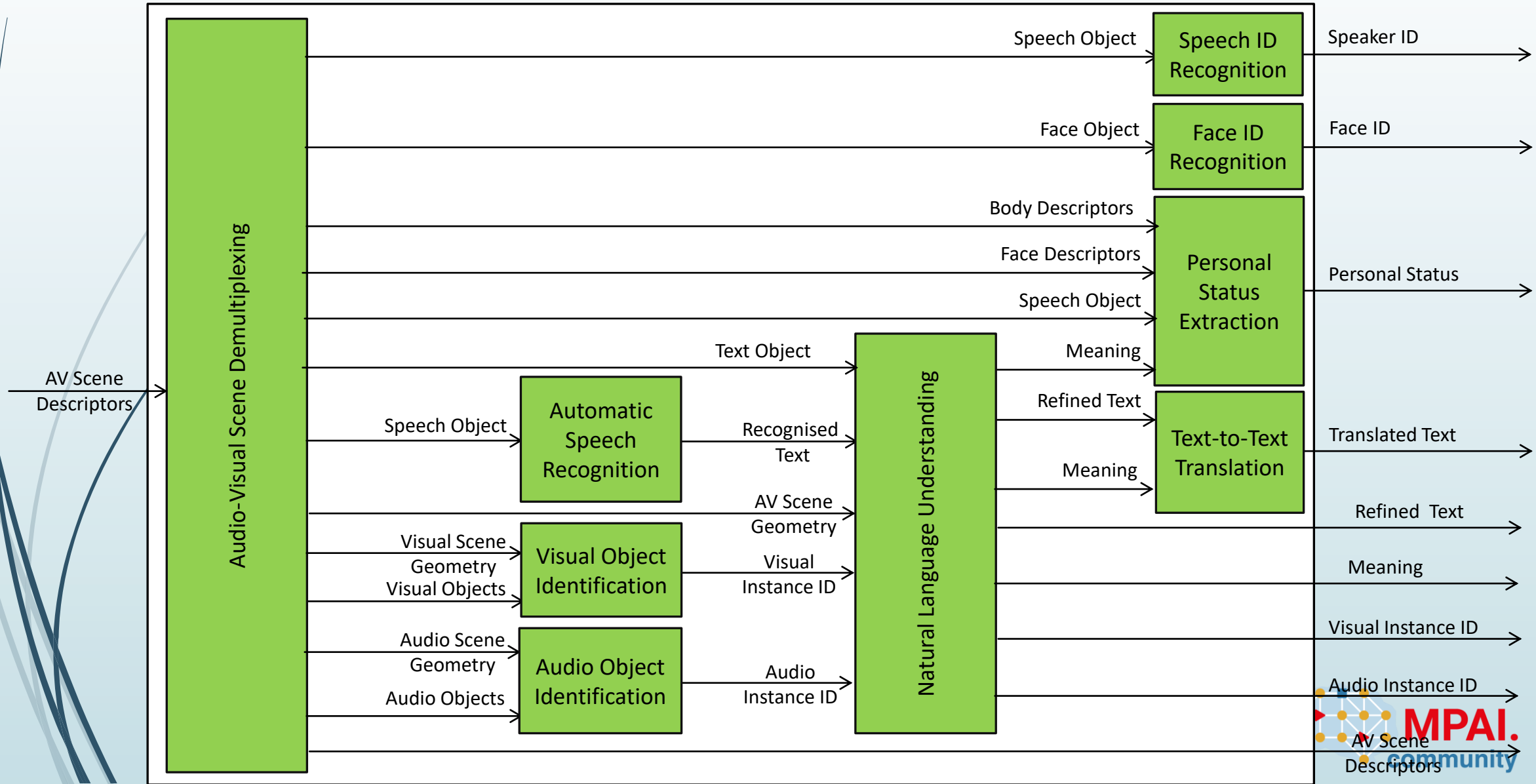
MPAI. community

About Types, Formats, and Attributes (MPAI-TFA)

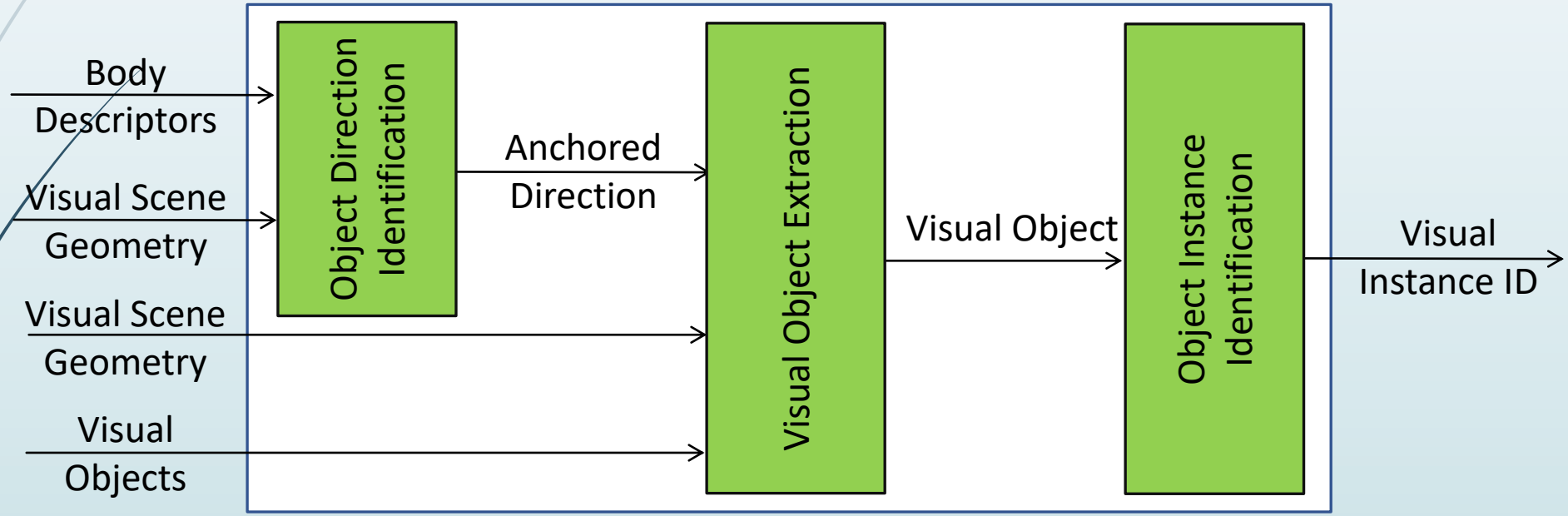
- A new standard for Data Type Qualifiers, a specialised type of metadata supporting the operation of AIMs receiving data from other AIMs and conveying information on:
  - Sub-Types: the Video Data Type has several Sub-Types (e.g., Colour, Brightness, ...)
  - Format: the Visual Data Sub-Type can be represented as PCM, JPEG, AVC…
  - Attribute: the Visual Data Sub-Types can have Face ID as an Attribute
- Qualifiers can be defined for a variety of Data Types, currently only for Text, Speech, Audio, and Visual.
- More qualifiers in the future.



Entity and Context Understanding (HMC-ECU) V1.1

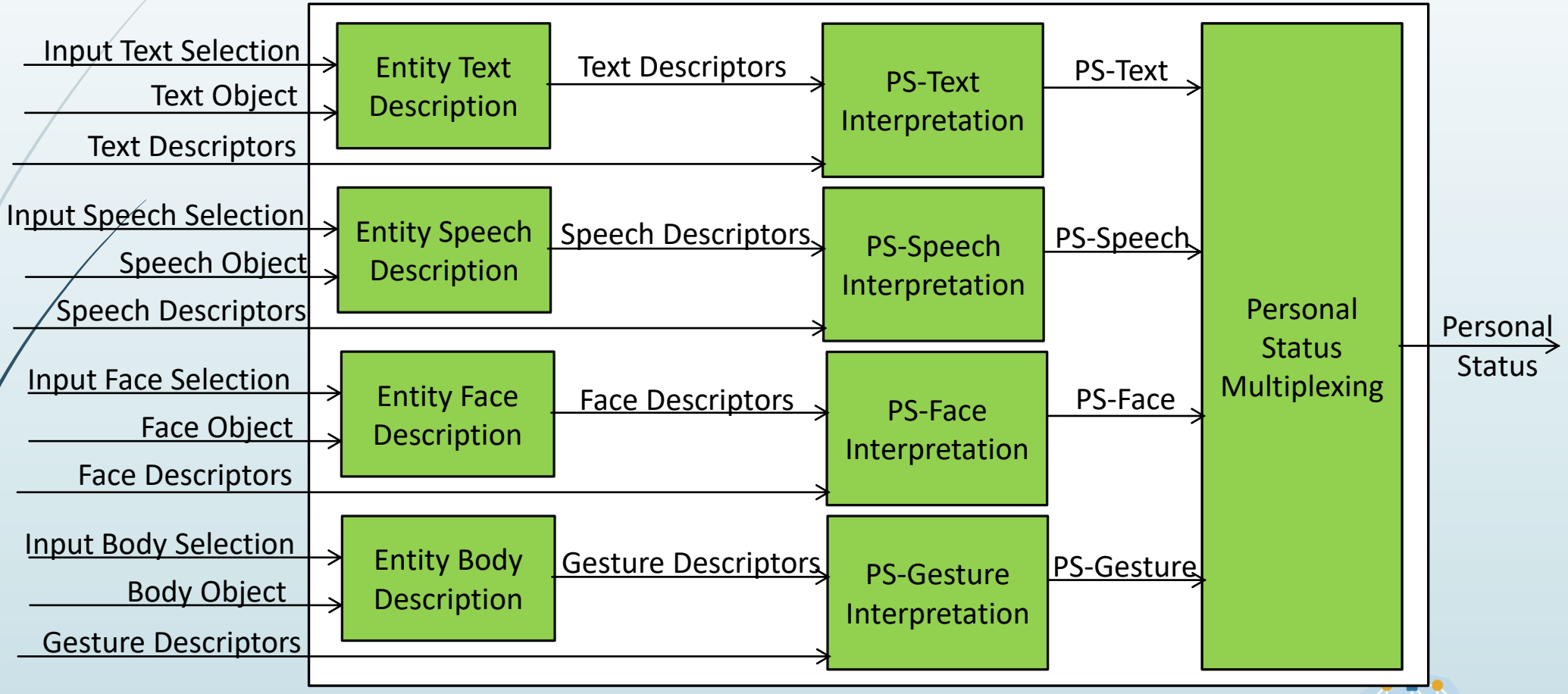


Visual Object Identification (OSD-VOI)

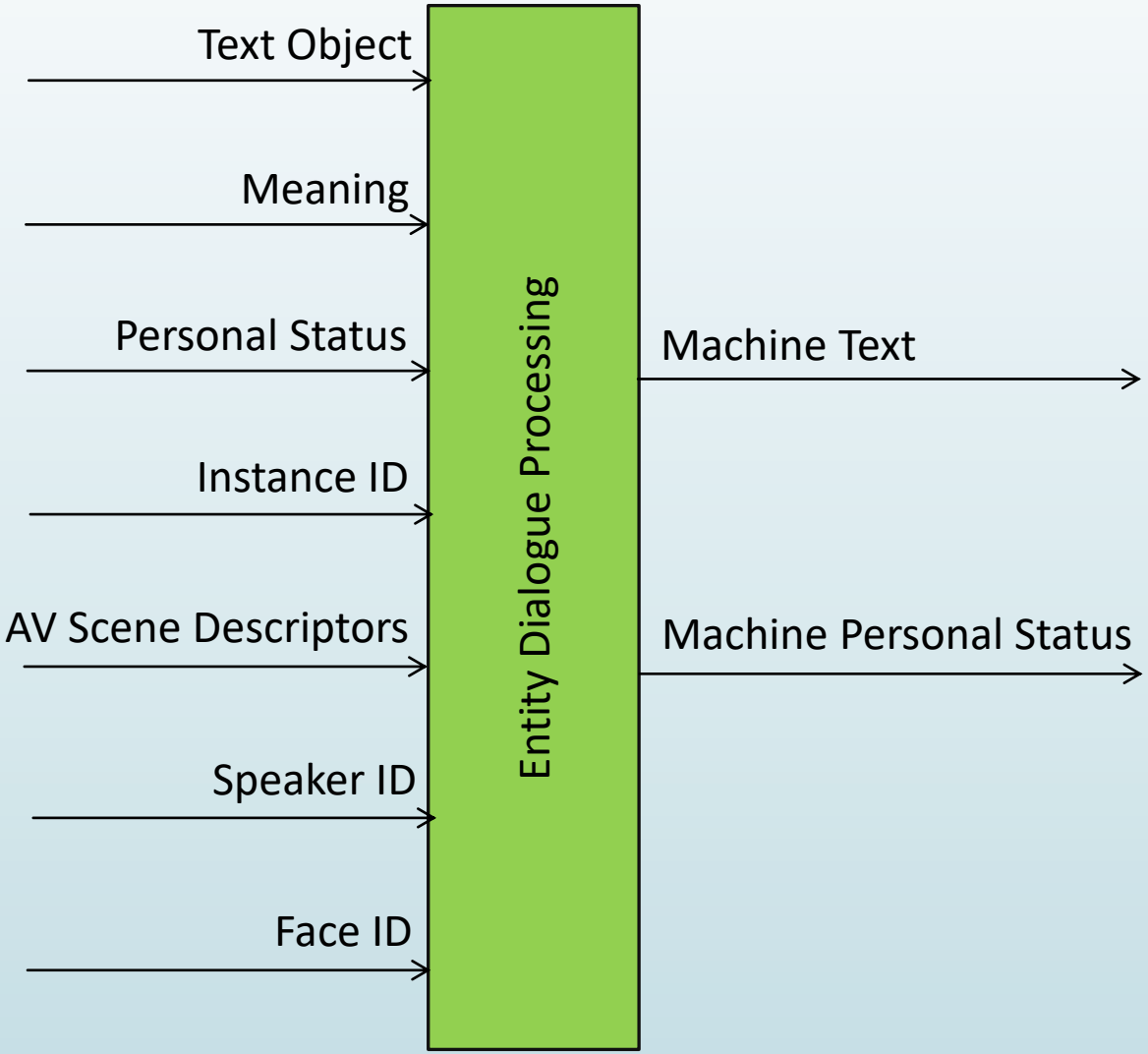


MPAI. community

Personal Status Extraction (MMC-PSE) V2.2

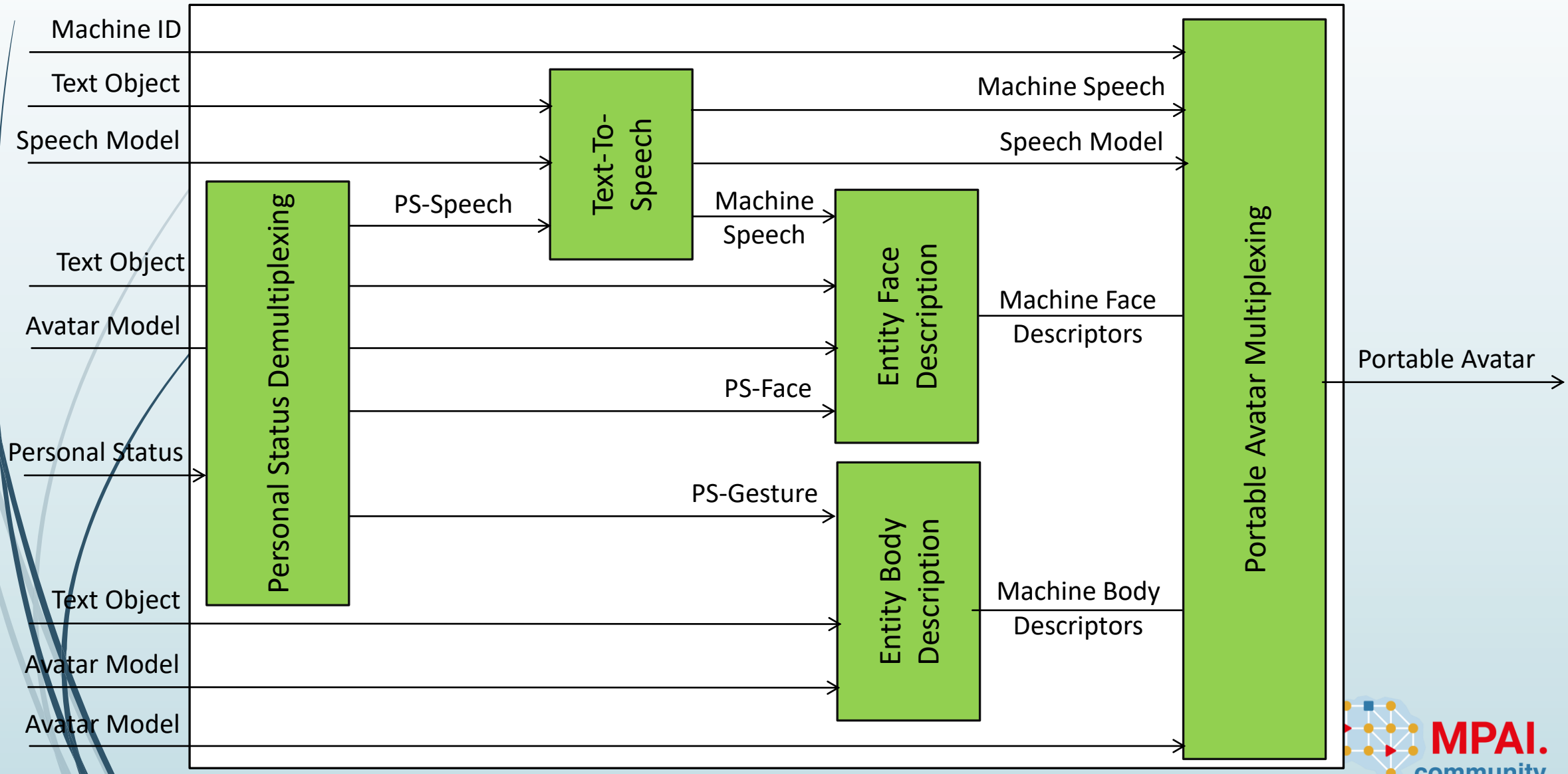


Entity Dialogue Processing (MMC-EDP) V2.2



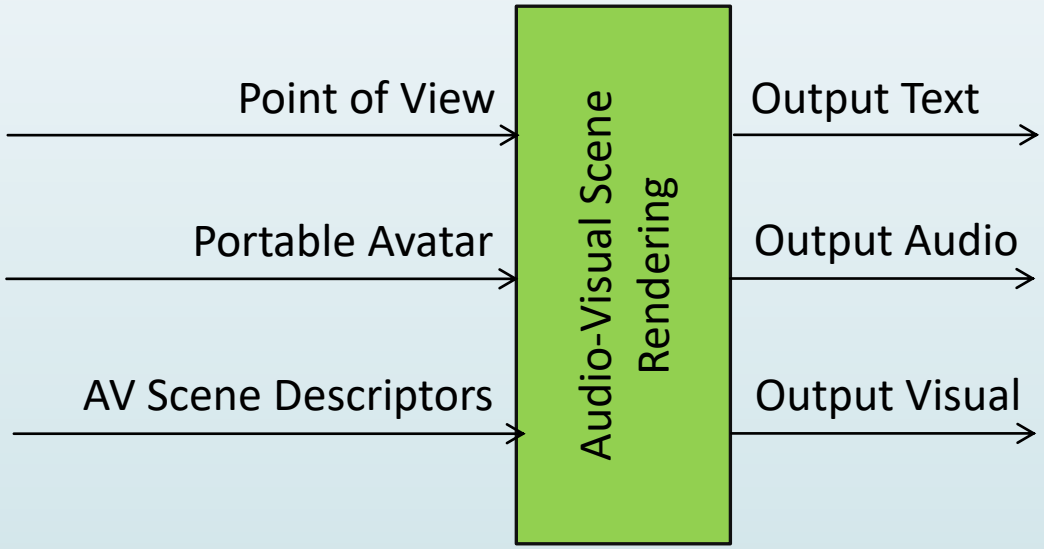
MPAI. community

Personal Status Display (PAF-PSD) V1.2



MPAI. community

Audio-Visual Scene Rendering (PAF-AVR)



HMC-CEC Profiles/1

HMC-CEC uses six groups of technologies to process a Communication Item:

Group does	Functions
Receives	Communication Item directly or Audio-Visual Scene.
Extracts	Personal Status from the Modalities (Text, Speech, Face, or Gesture) in the Communication Item received.
Understands	The Communication Item from the Modalities and the extracted Personal Status with or without use of the spatial information embedded in the Communication Item.
Translates	Using the set of Modalities available to the Machine.
Generates	Response.
Displays	The response using available Modalities.

HMC-CEC Profiles/2

Attributes	Codes	Sub-Attribute Codes				
Audio-Visual Scene	AVS	TXO	SPO	AUO	VIO	PAF
Personal Status	EPS	TXO	SPO	FCO	GSO	
Understanding	UND	TXO	SPO	FCO	GSO	SPC
Translation	TRN	TXO	SPO	FCO	GSO	
Display Response	RES	TXO	SPO	FCO	GSO	

https://schemas.mpai.community/HMC/V1.1/data/CECProfiles.json

MPAI. community

A disclaimer

- MPAI-HMC shows how new applications can be developed from AIWs, AIMs, and Data Types specified by 12 MPAI standards.
- MPAI-HMC uses technologies from:
  - AI Framework (MPAI-AIF)
  - Context-based Audio Enhancement (MPAI-CAE)
  - Multimodal Conversation (MPAI-MMC)
  - Object and Scene Description (MPAI-OSD)
  - Portable Avatar Format (MPAI-PAF)
- MPAI-HMC only specifies 2 AIMs, no Data Type

We look forward to working
with you
on this other MPAI projects!

https://mpai.community/

Join MPAI
Share the fun
Build the future

MPAI.
community