



Multimodal Conversation (MPAI-MMC)

08 and 15 UTC, 5 September 2023 (Tuesday)

Contents of presentation

- [About MPAI](#)
- [About MPAI-MMC](#)
- [Use Cases and AI Modules](#)
- [MPAI-MMC Data Types](#)
- [What's next](#)



About MPAI

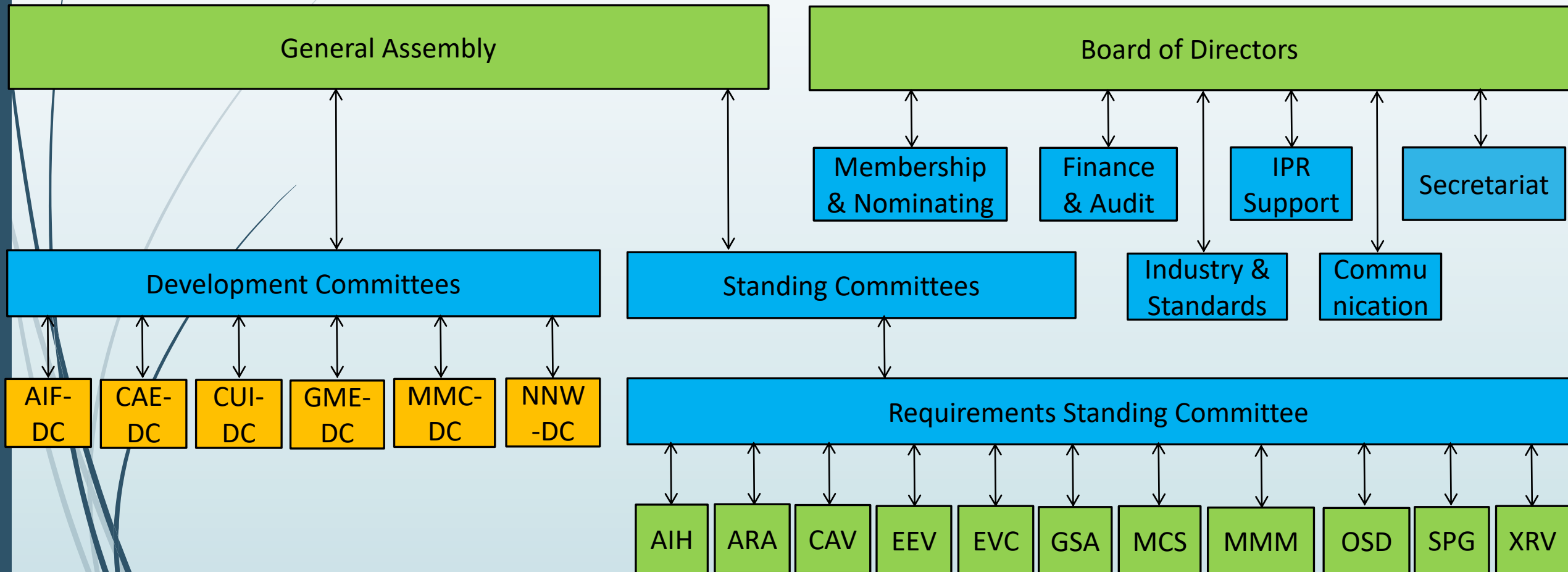
MPAI stands for Moving Picture, Audio, and Data Coding by Artificial Intelligence.

International, unaffiliated, non-profit SDO.

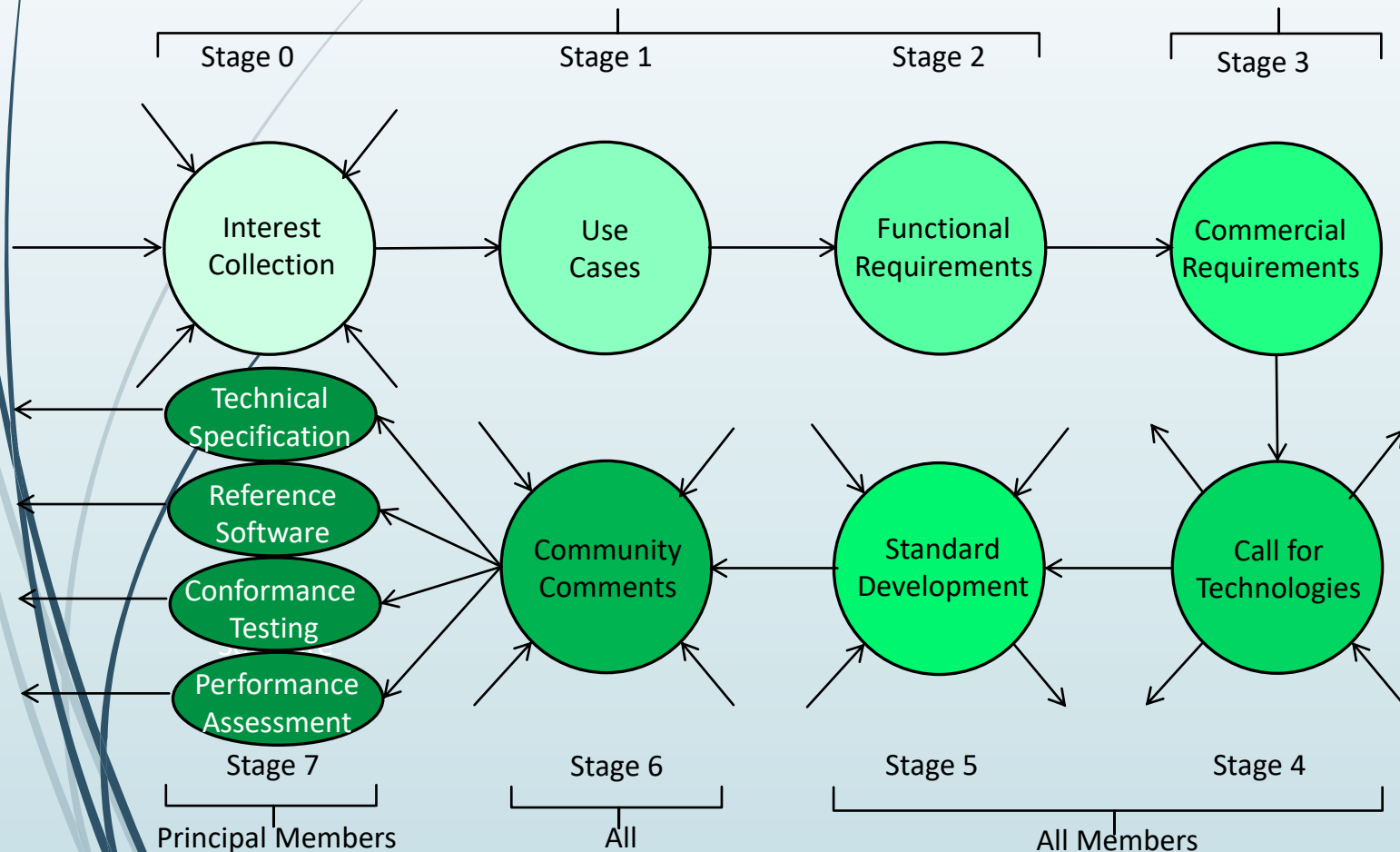
Developing AI-based data coding standards.

With clear Intellectual Property Rights licensing frameworks.

The MPAI organisation



The MPAl standard development process



- Develop Use Cases and Functional Requirements.
- Develop Commercial Requirements (Framework Licence).
- Issue Call for Technologies with attached:
 - Functional Requirements.
 - Commercial Requirements.
- Develop standard (MPAl members only).
- SEP holders select patent pool administrator.

MPAI standards for a better AI

- MPAI's data coding standards make explicit the computing workflow of AI applications.
- An MPAI standard **breaks up monolithic AI applications** into a set of interacting components of known data semantics (as far as possible).
- **Developers compete** offering “improved” performance “standard” components.
- Humans can **select applications whose internal operation they can somehow understand**.

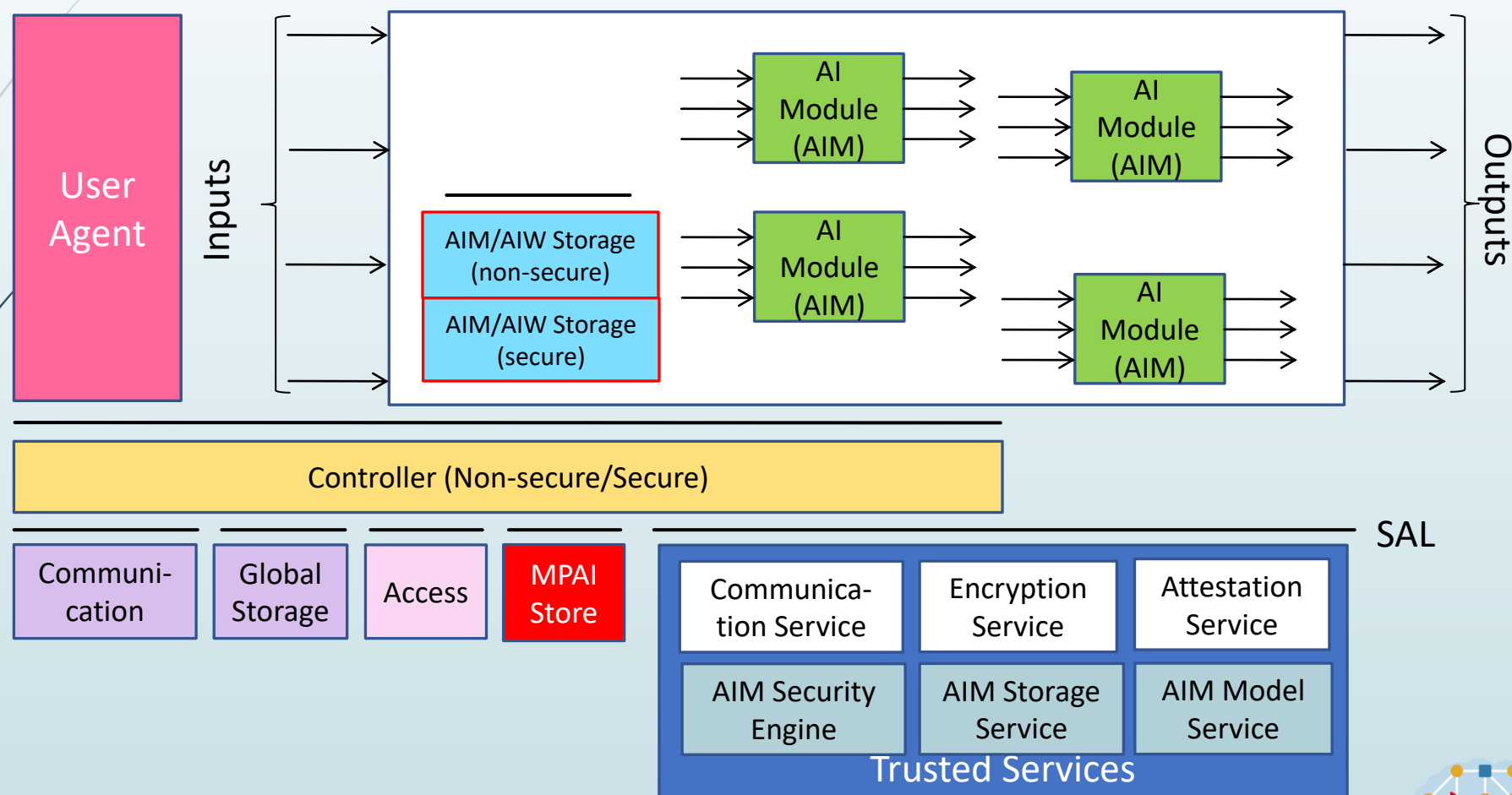
*MPAI's AI
standardisation is
“component-based”.*

An AI application is:

- *Subdivided in smaller components: AI modules (AIM).*
- *Aggregated in one or more AI workflows (AIW).*
- *Executed in a standard environment (AIF).*

*1 foundational Technical Specification
AI Framework (MPAI-AIF)*

The MPAI AI Framework



A sustainable MPAI Ecosystem

- **MPAI standards** create an ecosystem composed of:
 - **Developers:** develop components
→ require interoperability to bring their components to the market.
 - **Integrators:** assemble components
→ require ability to assemble third party components.
 - **Consumers:** use assembled components
→ require that the assembled components be trusted.
- The MPAI Store guarantees that AIMs/AIW's are:
 - Interoperable.
 - Trusted.
 - Available.

***1 system Technical Specification:
Governance of the MPAI Ecosystem (MPAI-GME).***

MPAI standards for a better AI

- MPAI's data coding standards make explicit the computing workflow of AI applications.
- An MPAI standard **breaks up monolithic AI applications** into a set of interacting components of known data semantics (as far as possible).
- **Developers compete** offering “improved” performance “standard” components.
- Humans can **select applications whose internal operation they can somehow understand**.

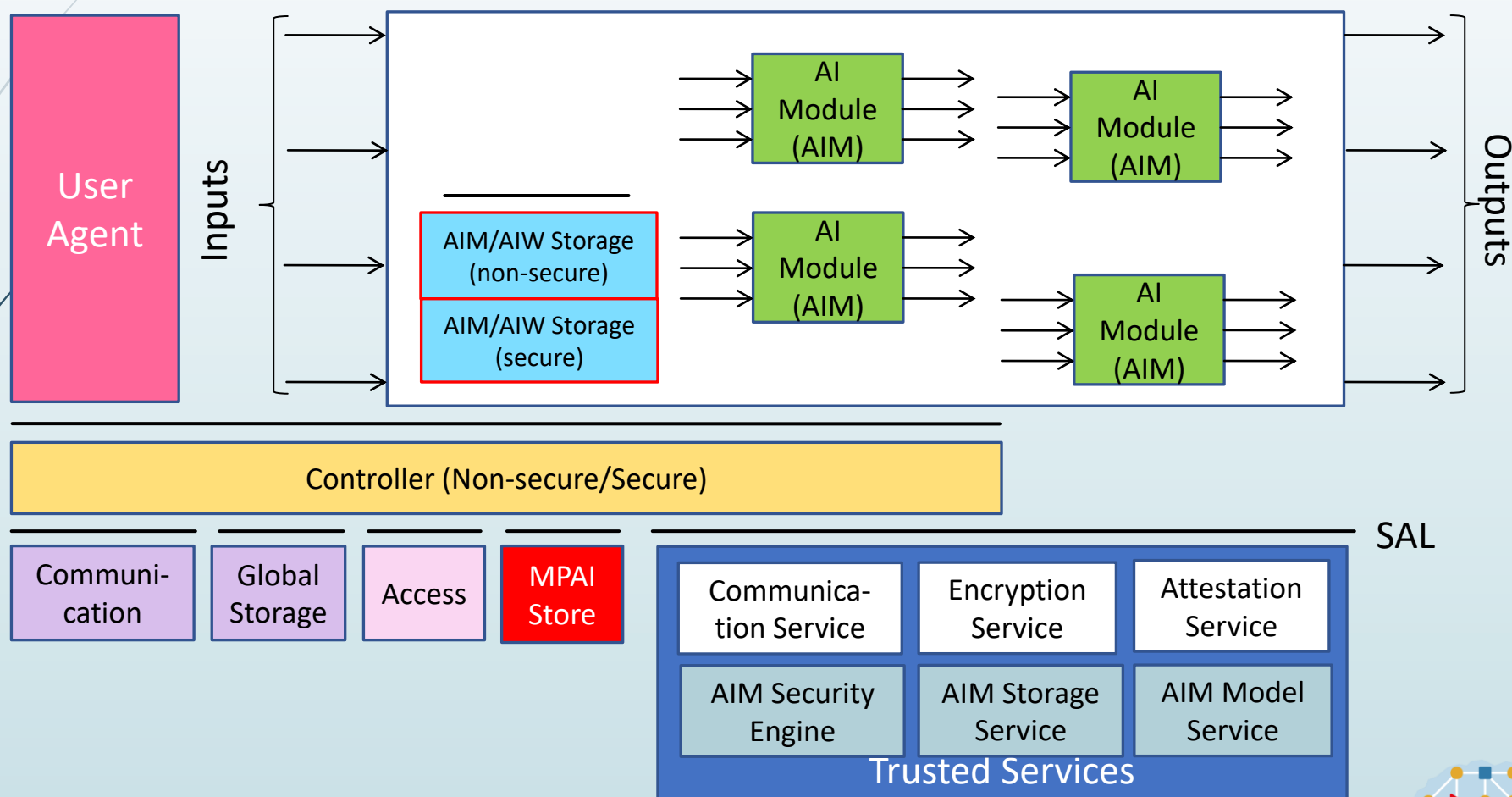
*MPAI's AI
standardisation is
“component-based”.*

An AI application is:

- *Subdivided in smaller components: AI modules (AIM).*
- *Aggregated in one or more AI workflows (AIW).*
- *Executed in a standard environment (AIF).*

1 foundational Technical Specification
AI Framework (MPAI-AIF)

The MPAI AI Framework

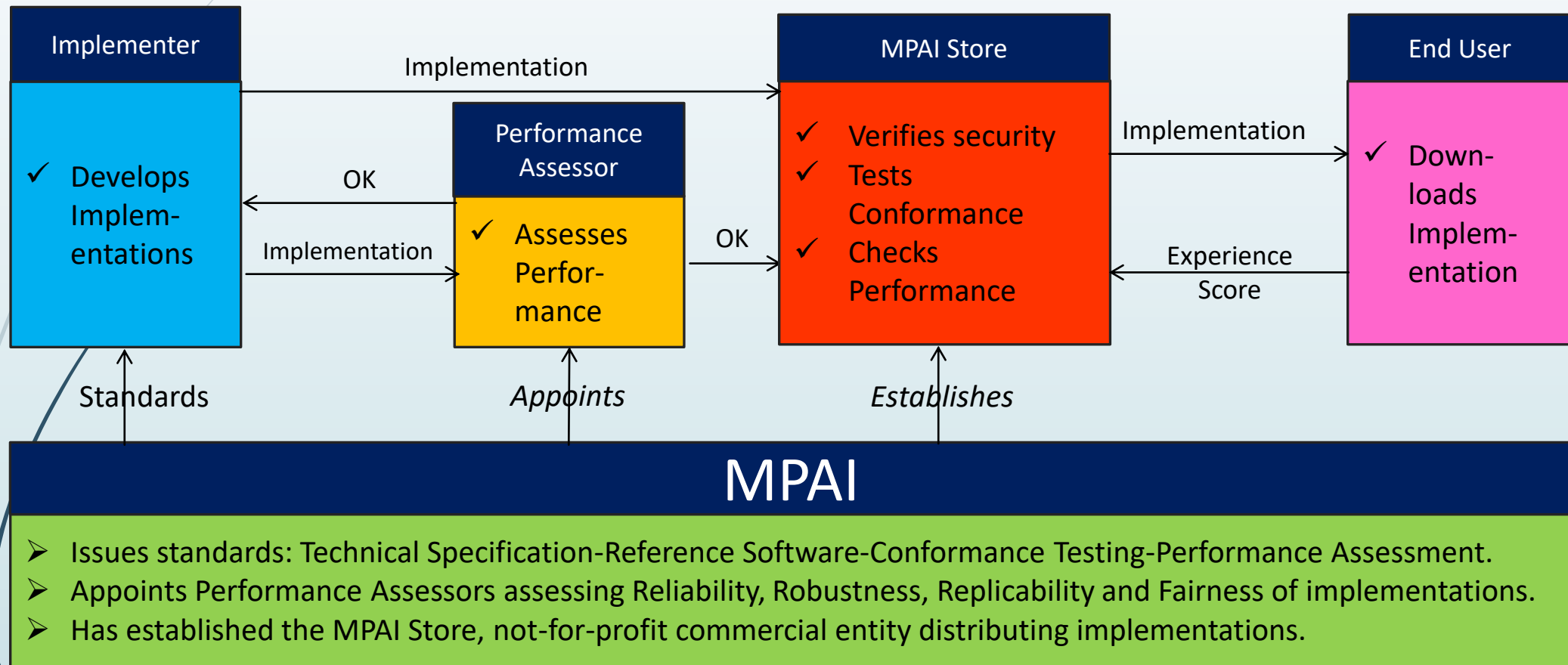


A sustainable MPAI Ecosystem

- **MPAI standards** create an ecosystem composed of:
 - **Developers:** develop components
→ require interoperability to bring their components to the market.
 - **Integrators:** assemble components
→ require ability to assemble third party components.
 - **Consumers:** use assembled components
→ require that the assembled components be trusted.
- The MPAI Store guarantees that AIMs/AIW's are:
 - Interoperable.
 - Trusted.
 - Available.

***1 system Technical Specification:
Governance of the MPAI Ecosystem (MPAI-GME).***

The MPAI ecosystem



More published MP AI standards

4 Technical Specifications

- 1 - Context-based Audio Enhancement (MPAI-CAE)*
- 2 - Compression and Understanding of Financial Data (MPAI-CUI)*
- 3 - Multimodal Conversation (MPAI-MMC)*
- 4 - Neural Network Watermarking (MPAI-NNW)*

2 Technical Reports

- 1 - MP AI Metaverse Model (MPAI-MMM) – Functionalities*
- 2 - MP AI Metaverse Model (MPAI-MMM) – Functionality Profiles*

Five standards published for Community Comments to become standards on 29 September

Existing MPAI standards extended

- 1 - AI Framework V2 (MPAI-AIF)**
- 2 - Multimodal Conversation V2 (MPAI-MMC)**

New MPAI standards being approved

- 3 - Avatar Representation and Animation V1 (MPAI-ARA)**
- 4 - Connected Autonomous Vehicles V1 (MPAI-CAV) – Architecture**
- 5 - MPAI Metaverse Model V1 (MPAI-MMM) – Architecture**

Brewing in the pot

Calls for Technologies issued

- 1 – Artificial Intelligence for Health (MPAI-AIH)*
- 2 – Object and Scene Description (MPAI-OSD)*
- 3 – Extended Reality Venues (MPAI-XRV) - Live Theatrical Stage Performance*

New opportunities being explored

- 1 - AI-based End-to-End Video Coding (MPAI-EEV)*
- 2 - AI-Enhanced Video Coding (MPAI-EVC)*
- 3 - Server-based Predictive Multiplayer Gaming (MPAI-SPG)*

MPAI and IEEE

MPAI Technical Specifications adopted as IEEE standards

1. ***MPAI-AIF – 3301-2022***
2. ***MPAI-CAE – 3302-2022***
3. ***MPAI-MMC – 3300-2022***
4. ***MPAI-CUI – 3303-2023***
5. ***MPAI-NNW (on its way)***

All this achieved in less than 3 years!



About MPAI-MMC

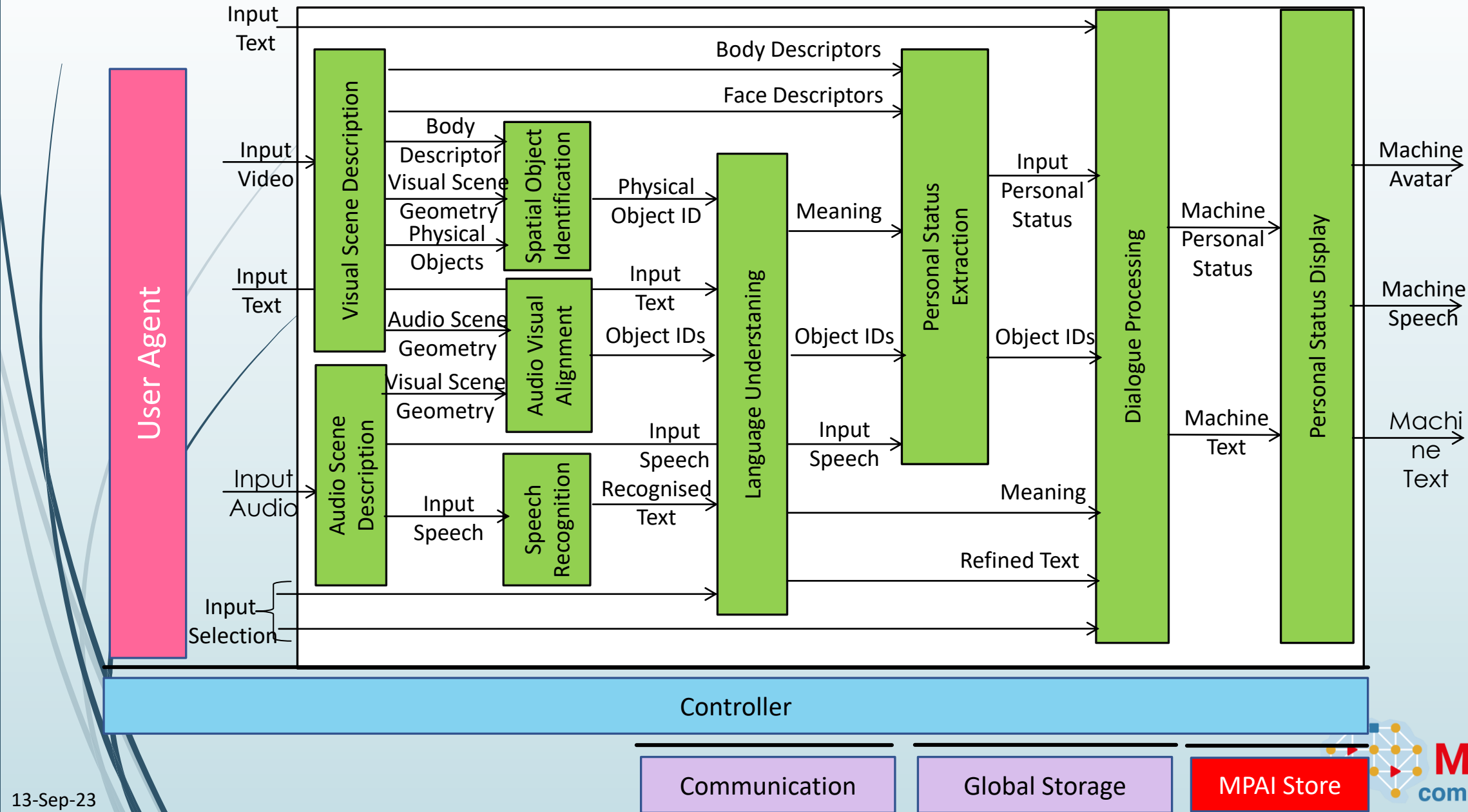
Technical Specification: Multimodal Conversation V2

- **Goal:** to provide technologies that enable a human-machine conversation that is more **human-like** and able to **emulate human-human conversation** in completeness and intensity.
- Improve machine's "conversational" capabilities in the two main phases of conversation: **understanding** the meaning and **generating** a pertinent response.
- MPAI-MMC V 2 achieves this goal by providing, among other technologies, a **new standard data type – Personal Status** – that represents the "internal status" of a conversing human expressed with text, speech, face, and gesture.
- Personal Status can be **used by the machine** to represent its own internal status as if it were a human.

Scope of Multimodal Conversation

- **Technologies for analysis** of text, speech and other non-verbal components used in human-machine and machine-machine conversation.
- **Functions and Formats of input and output Data** of AImS, Composite AImS, and AIWs (omplementing Use Cases).
- **Use Cases** that apply the technologies, currently:
 - Conversation with Personal Status.
 - Conversation with Emotion.
 - Multimodal Question Answering.
 - Conversation About a Scene.
 - Human-CAV Interaction.
 - Virtual Secretary for Videoconference
 - Speech Translation (one way, two ways, one to many)

A representative Use Case: Conversation with Personal Status



Input/output data of the AIW

Input	Comments
Input Text	Text typed by the human as additional information stream or as a replacement of the Speech.
Input Speech	Speech of the human having a conversation with the machine.
Input Video	Video of the Face of the human having a conversation with the machine.
Input Selection	Data determining the use of Speech vs Text.
Output	Comments
Machine Text	Text of the Speech produced by the machine.
Machine Speech	Synthetic Speech produced by the machine.
Machine Video	Avatar representing the machine.

Functions of the AIMs

AIM	Function
Visual Scene Description	Provides <ol style="list-style-type: none">1. Face and Body Descriptors2. Physical Objects3. Visual Scene Descriptors.
Audio Scene Description	Provides Speech Objects and their Spatial Attitudes.
Speech Recognition	Recognises Speech and produces Recognised Text.
Language Understanding	Refines Recognised Text and extracts Meaning.
Personal Status Extraction	Extracts Input Personal Status:
Dialogue Processing	<ol style="list-style-type: none">1. Processes Refined Text and Input Personal Status2. Produces machine's Text and Personal Status.
Personal Status Displays	<ol style="list-style-type: none">1. Synthesises Machine Speech from Machine Text and Personal Status.2. Synthesises Machine Avatar.

Input and output data of the AIMS

AIM	Receives	Produces
Visual Scene Description	Input Video	<ol style="list-style-type: none"> 1. Face Descriptors 2. Body Descriptors 3. Visual Scene Geometry 4. Physical Objects
Audio Scene Description	Input Audio	<ol style="list-style-type: none"> 1. Speech 2. Audio Scene Geometry
Spatial Object Identification	<ol style="list-style-type: none"> 1. Body Descriptors 2. Visual Scene Geometry 3. Physical Objects 	Physical Object ID
Speech Recognition	Input Speech	Recognised Text
Language Understanding	<ol style="list-style-type: none"> 1. Physical Object ID 2. Input Text 3. Recognised Text 4. Input Selection 	<ol style="list-style-type: none"> 1. Meaning 2. Refined Text
Personal Status Extraction	<ol style="list-style-type: none"> 1. Body Descriptors 2. Face Descriptors 3. Meaning 4. Speech 	Input Personal Status
Dialogue Processing	<ol style="list-style-type: none"> 1. Input Text 2. Refined Text 3. Input Personal Status 4. Input Selection 	<ol style="list-style-type: none"> 1. Machine Personal Status 2. Machine Text
Personal Status Displays	<ol style="list-style-type: none"> 1. Machine Text 2. Machine Personal Status 	<ol style="list-style-type: none"> 1. Machine Avatar 2. Machine Speech 3. Machine Text

JSON metadata



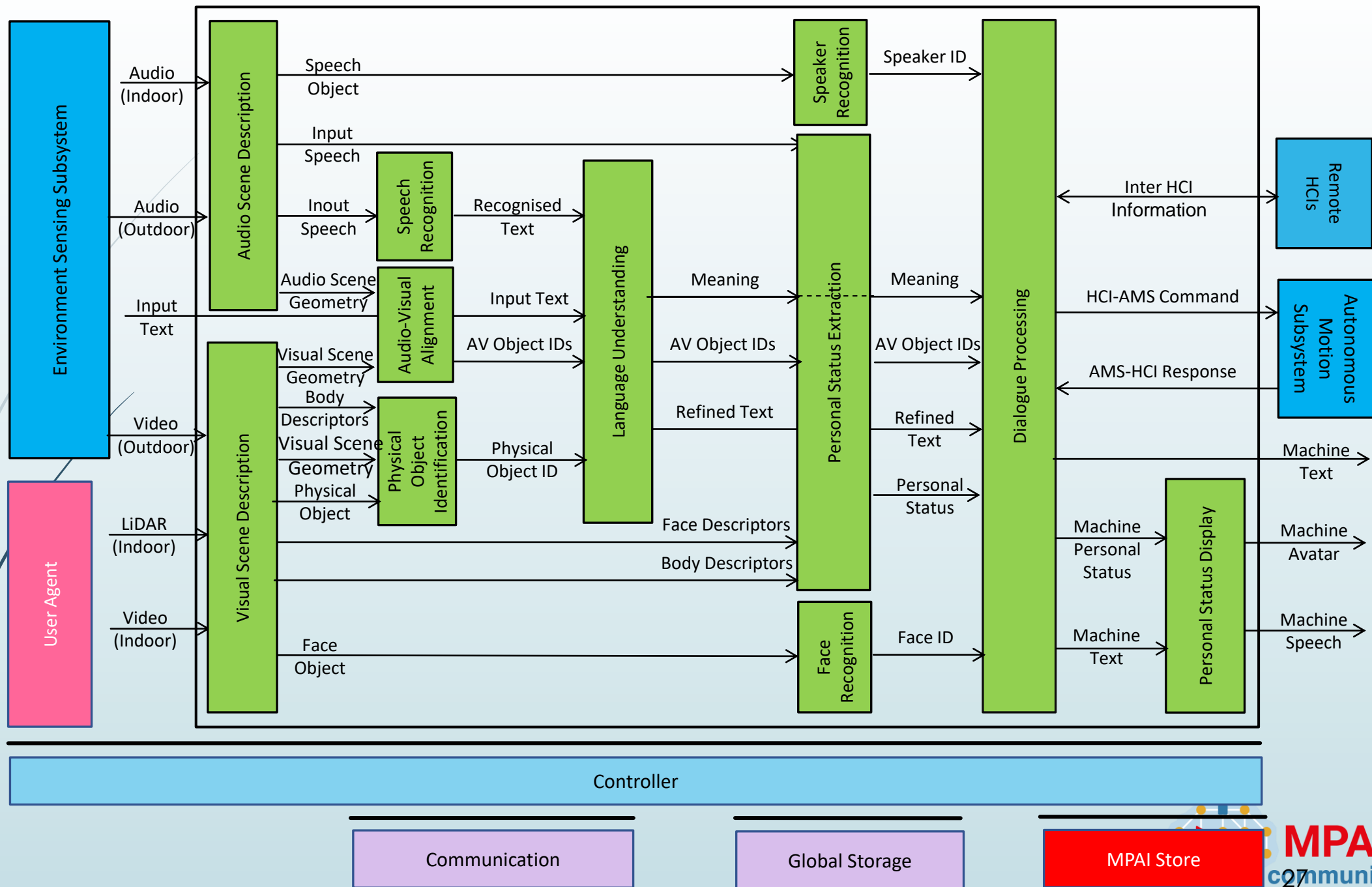
- The Conversation About a Scene AIW should be **executed in an AIF**.
- The **capabilities of the AIF** are **described** by a standard JSON metadata format.
- The **capabilities of the AIW** are **described** by a standard JSON metadata format.
- The **capabilities of (Composite) AIMs** are **described** by a standard JSON metadata format (similar to the AIW metadata format).
- **An AIF downloads a suitable AIW and suitable AIMs** from the MPAI Store using the JSON metadata.



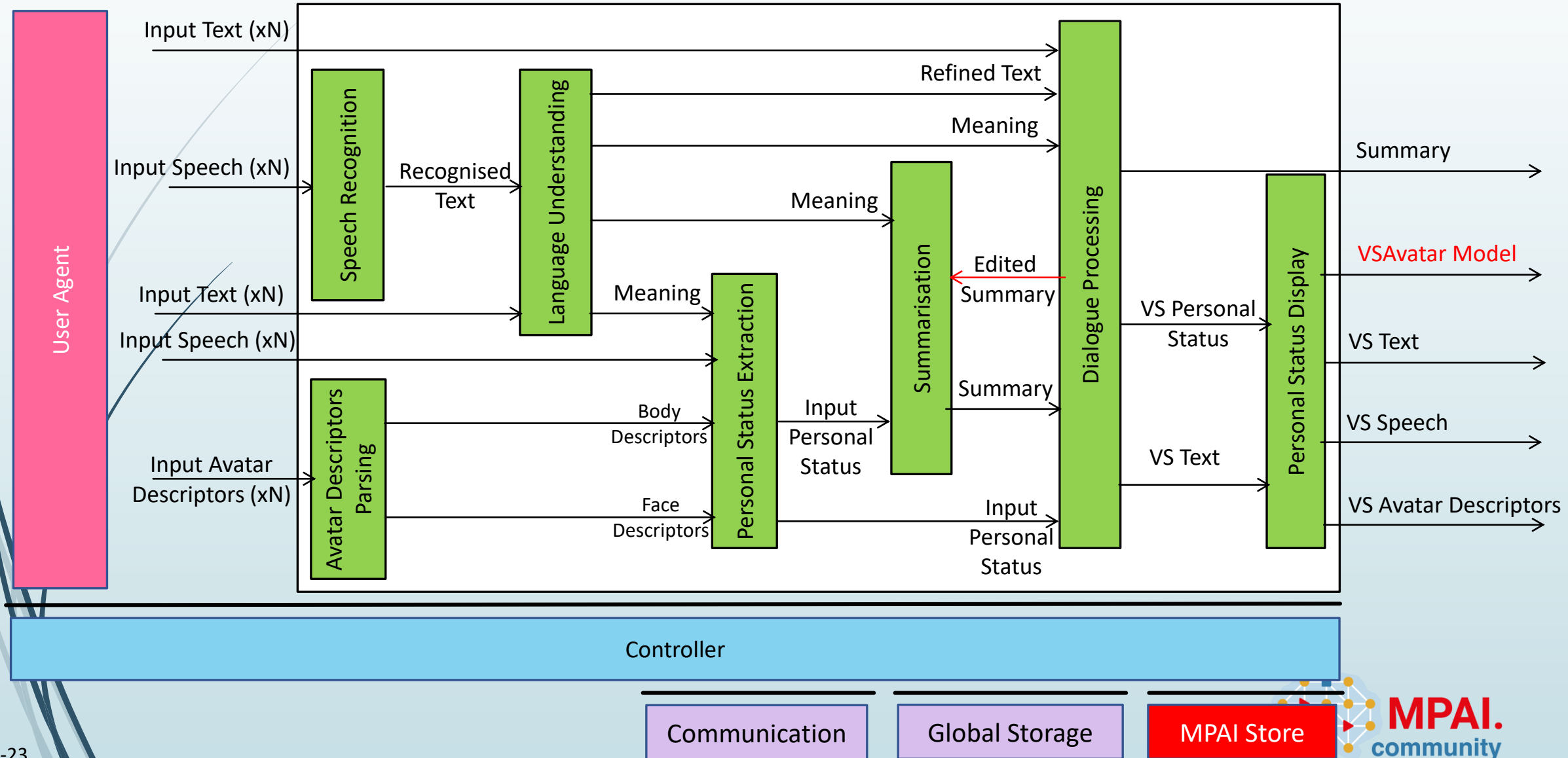
Use Cases and AI Modules

Another Use Case: Human-CAV

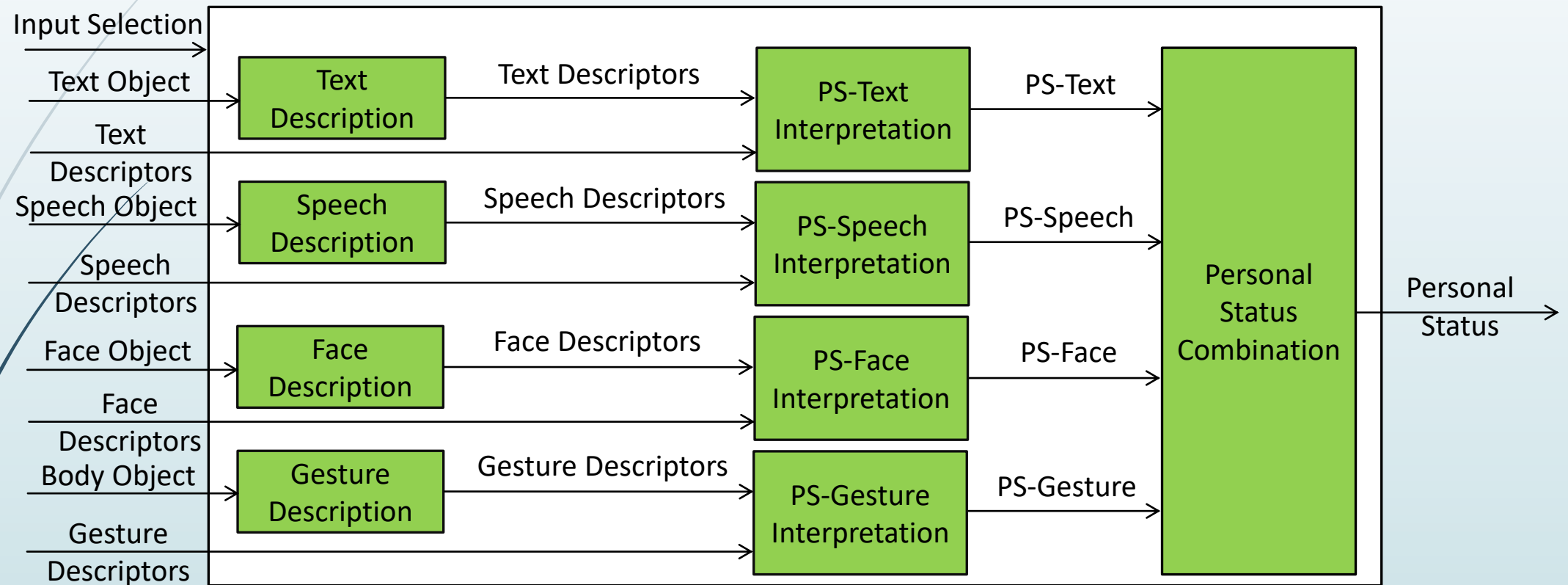
Interaction



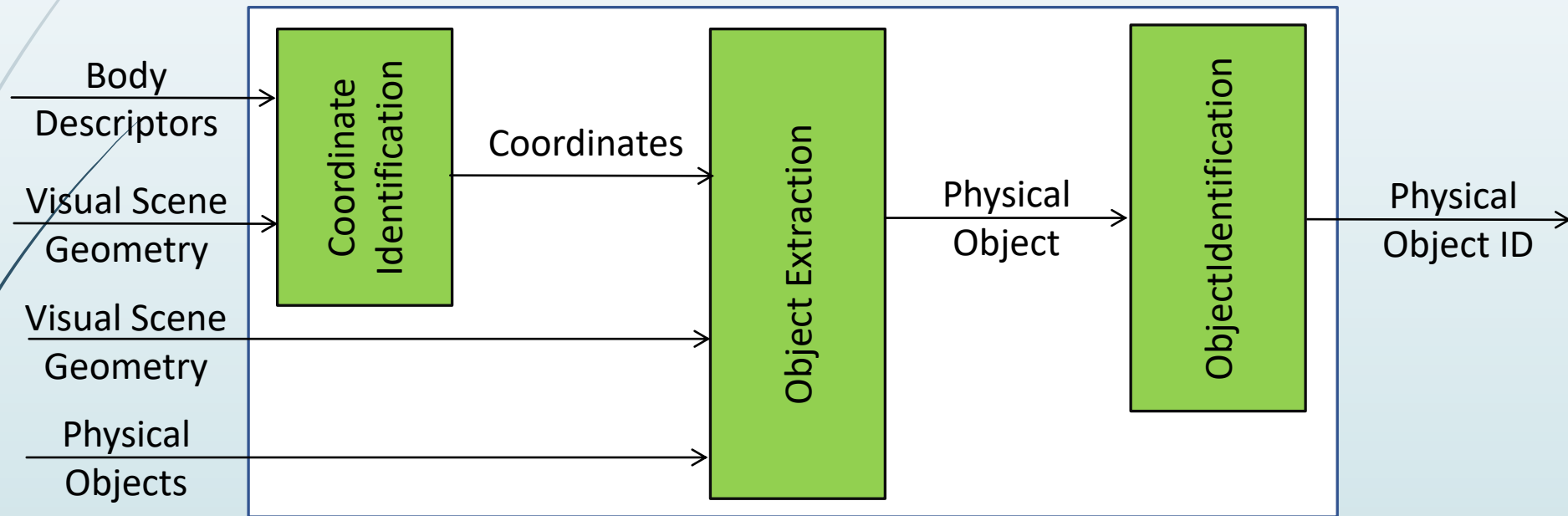
Virtual secretary for Videoconference



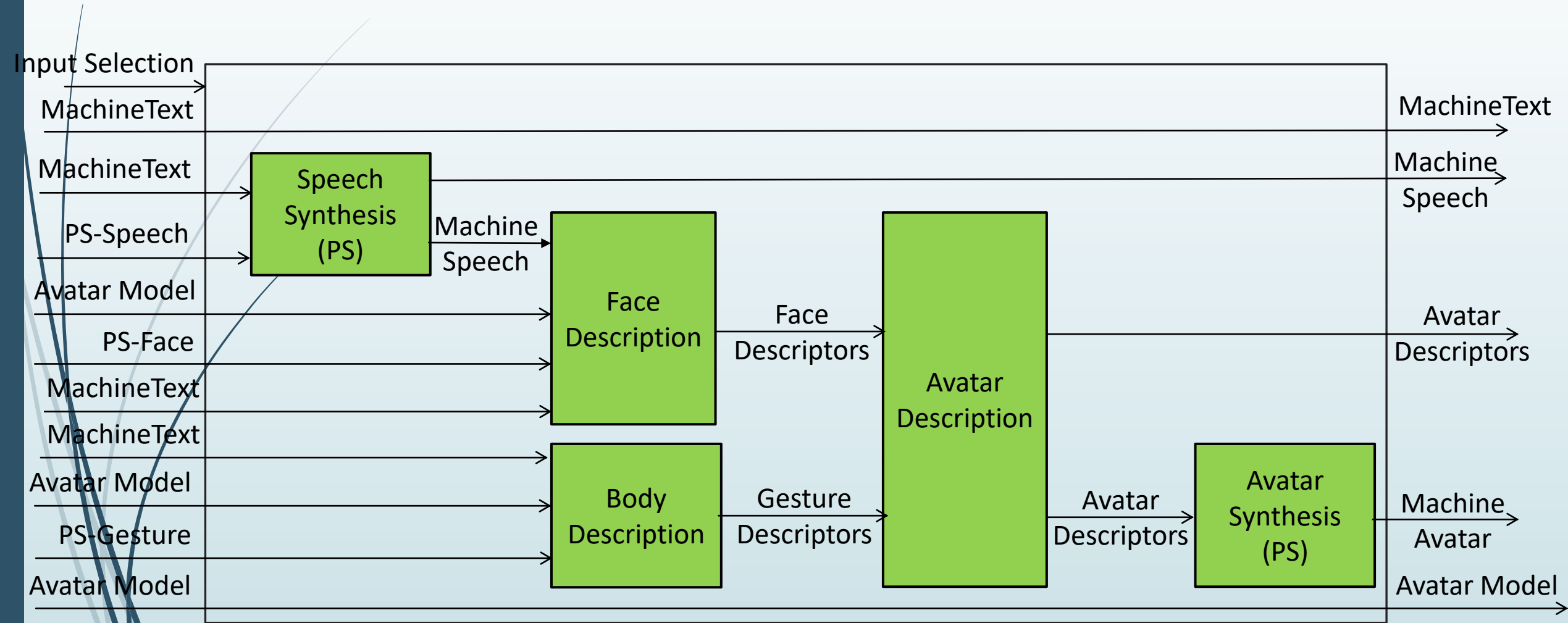
Personal Status Extraction



Spatial Object Identification



Personal Status Display





MPAI-MMC Data Types

Name of Data Format	Specified by
Audio File	MMC
Audio Scene Descriptors	CAE
Cognitive State	MMC
Emotion	MMC
Face Descriptors	ARA
Gesture Descriptors	ARA
Instance Identifier	MMC
Language Identifier	MMC
Meaning/Text Descriptors	MMC
Personal Status	MMC
Social Attitude	MMC
Spatial Attribute	OSD
Speech Descriptors/Speech Features	MMC
Text	MMC
Video	MMC
Video File	MMC
Video Of Faces KB Query Format	MMC
Visual Scene Descriptors	ARA

ARA: Avatar Representation and Animation

CAE: Context-based Audio Enhancement

OSD: Object and Scene Description

Personal Status/1

- Personal Status is a data structure composed of three Personal Status *Factors*:
 1. Emotion (such as “angry” or “sad”).
 2. Cognitive State (such as “surprised” or “interested”).
 3. Social Attitude (such as “polite” or “arrogant”).
- Each *Factor* can be conveyed by *Modalities*: Text, Speech, Face, and Gestures.
- In the future, other *Modalities* may be added, e.g., body posture.
- Within a given *Modality*, *Factors* are analysed/interpreted via *Descriptors*.
- E.g., when using Speech, the elements may be expressed by combinations of:
 - Prosody (pitch, rhythm, and volume variations);
 - Separable speech effects (such as degrees of voice tension, breathiness, etc.);
 - Vocal gestures (laughs, sobs, etc.).

Personal Status/2

Two tables for each of Factor (Emotion, Cognitive State, and Social Attitude):

- A *Label Set Table* containing descriptive labels relevant to the element type in a 3-level format:
 - The CATEGORIES column: specifies the relevant categories using nouns (e.g., “ANGER”).
 - The GENERAL ADJECTIVAL column: adjectival labels for general or basic labels within a category (e.g., “angry”).
 - The SPECIFIC ADJECTIVAL column: sub-category labels in the category (e.g., “furious”).
- A *Label Semantics Table* providing the semantics for each label in the GENERAL ADJECTIVAL and SPECIFIC ADJECTIVAL columns of the Label Set Table.
 - E.g., for “angry” the semantic gloss is “emotion due to perception of physical or emotional damage or threat.”

Personal Status/3

Note: MPAI-MMC provides the JSON representation of Personal Status.

Variable name	Code
Timestamp	Timestamp type
	Timestamp value
Emotion	Emotion version
	Fused Emotion value
	<i>Text</i> Emotion value
	<i>Speech</i> Emotion value
	<i>Face</i> Emotion value
Cognitive State	<i>Gesture</i> Emotion value
	Cognitive State version
	Fused Cognitive State value
	<i>Text</i> Cognitive State value
	<i>Speech</i> Cognitive State value
Social Attitude	<i>Face</i> Cognitive State value
	<i>Gesture</i> Cognitive State value
	Social Attitude version
	Fused Social Attitude value
	<i>Text</i> Social Attitude value
	<i>Speech</i> Social Attitude value
	<i>Face</i> Social Attitude value
	<i>Gesture</i> Social Attitude value

Basic Cognitive State Label Set (samples)

COGNITIVE CATEGORIES	GENERAL ADJECTIVAL	SPECIFIC ADJECTIVAL
AROUSAL	aroused/excited/energetic	cheerful playful lethargic sleepy
ATTENTION	Attentive	expectant/anticipating thoughtful distracted/absent-minded vigilant hopeful/optimistic
BELIEF	Credulous	sceptical
INTEREST	Interested	fascinated curious bored
SURPRISE	Surprised	astounded startled
UNDERSTANDING	Comprehending	uncomprehending bewildered/puzzled

Basic Cognitive State Semantics Set (samples)

ID	Cognitive State	Meaning
1	aroused/excited/energetic	cognitive state of alertness and energy
2	astounded	high degree of surprised
3	attentive	cognitive state of paying attention
4	bewildered/puzzled	high degree of incomprehension
5	bored	not interested
6	cheerful	energetic combined with and communicating happiness
7	comprehending	cognitive state of successful application of mental models to a situation
8	credulous	cognitive state of conformance to mental models of a situation
9	curious	interest due to drive to know or understand
10	distracted/absent-minded	not attentive to present situation due to competing thoughts
11	expectant/anticipating	attentive to (expecting) future event or events
12	fascinated	high degree of interest
13	interested	cognitive state of attentiveness due to salience or appeal to emotions or drives
14	lethargic	not aroused
15	playful	energetic and communicating willingness to play
16	sceptical	not credulous
17	sleepy	not aroused due to need for sleep
18	surprised	cognitive state due to violation of expectation
19	startled	surprised by a sudden event or perception
20	surprised	cognitive state due to violation of expectation
21	thoughtful	attentive to thoughts
22	uncomprehending	not comprehending

Basic Emotion Label Set (samples)

EMOTION CATEGORIES	GENERAL ADJECTIVAL	SPECIFIC ADJECTIVAL
ANGER	angry	furious irritated frustrated
CALMNESS	calm	peaceful/serene resigned
DISGUST	disgusted	repulsed
FEAR	fearful/scared	terrified anxious/uneasy
HAPPINESS	happy	joyful content delighted amused
HURT	hurt jealous	insulted/offended resentful/disgruntled bitter
PRIDE/SHAME	proud ashamed	guilty/remorseful/sorry embarrassed
RETROSPECTION	nostalgic	homesick
SADNESS	sad	lonely grief-stricken depressed/gloomy disappointed

Basic Emotion Semantics Set (samples)

ID	Emotion	Meaning
1	amused	positive emotion combined with interest (cognitive state)
2	angry	emotion due to perception of physical or emotional damage or threat
3	anxious/uneasy	low or medium degree of fear, often continuing rather than instant
4	ashamed	emotion due to awareness of violating social or moral norms
5	bitter	persistently angry due to disappointment or perception of hurt or injury
6	calm	relatively lacking emotion
7	content	medium or low degree of happiness, continuing rather than instant
8	delighted	high degree of happiness, often combined with surprise
9	depressed/gloomy	high degree of sadness, continuing rather than instant, combined with lethargy (see AROUSAL)
10	disappointed	sadness due to failure of desired outcome
11	disgusted	emotion due to urge to avoid, often due to unpleasant perception or disapproval
12	embarrassed	shame due to consciousness of violation of social conventions
13	fearful/scared	emotion due to anticipation of physical or emotional pain or other undesired event or events
14	frustrated	angry due to failure of desired outcome
15	furious	high degree of angry
16	grief-stricken	sadness due to loss of an important social contact
17	happy	positive emotion, often continuing rather than instant

Basic Social Attitude Label Set (samples)

SOCIAL ATTITUDE CATEGORIES	GENERAL ADJECTIVAL	SPECIFIC ADJECTIVAL
ACCEPTANCE	accepting exclusive/cliqish	welcoming/inviting friendly unfriendly/hostile
AGREEMENT, DISAGREEMENT	like-minded argumentative/disputatious	sarcastic
AGGRESSION	aggressive peaceful submissive	combative/belligerent passive-aggressive mocking
APPROVAL, DISAPPROVAL	admiring/approving disapproving indifferent	awed contemptuous
ACTIVITY, PASSIVITY	assertive passive	controlling permissive/lenient
COOPERATION	cooperative/agreeable uncooperative	flexible subversive/undermining uncommunicative stubborn disagreeable
RESPONSIVENESS	responsive/demonstrative emotional/passionate unresponsive/undemonstrative unemotional/detached	enthusiastic unenthusiastic passionate dispassionate

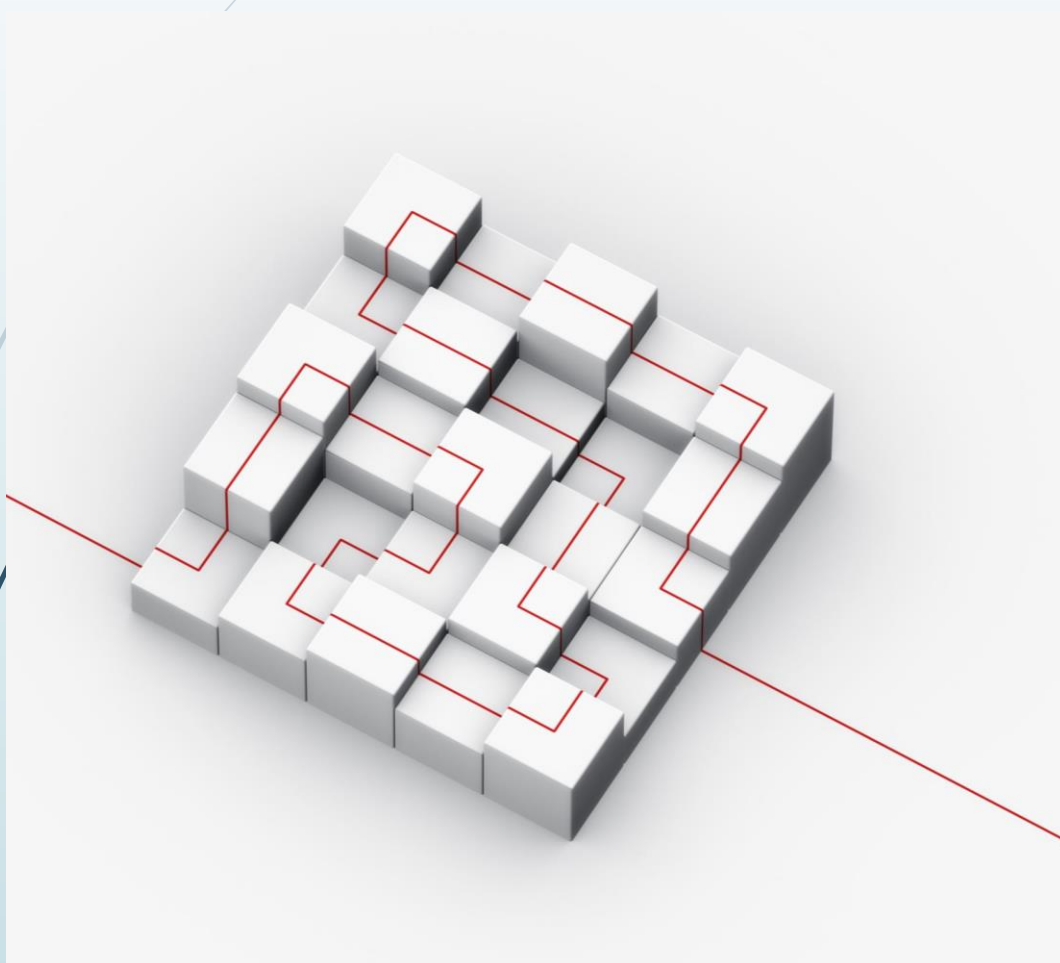
Basic Social Attitude Semantics Set (samples)

ID	Social Attitude	Meaning
1	accepting	attitude communicating willingness to accept into relationship or group
2	admiring/approving	attitude due to perception that others' actions or results are valuable
3	aggressive	tending to physically or metaphorically attack
4	apathetic/indifferent	showing lack of interest
5	approachable	sociable and not inspiring inhibition
6	argumentative	tending to argue or dispute
7	arrogant	emotion communicating social dominance
8	assertive	taking active role in social situations
9	awed	approval combined with incomprehension or fear
10	belittling	criticising by understating victim's achievements, personal attributes, etc.
11	boastful	tending to praise or promote self
12	brazen	high degree of forwardness/presumption
13	candid/frank	open in linguistic communication
14	closed/distant	not open
15	commanding/domineering	tending to assert right to command
16	combative/belligerent	high degree of aggression, often physical
17	communicative	evinced willingness to communicate as needed



What's next

And now?



- Technical Specification: Multimodal Conversation WD for Community Comments is **open to comments** by anybody.
- **Comments should reach** secretariat@mpai.community by 2023/09/25 T 23:59 UTC.
- **No specific format** is required to make comments.
- MP AI intends to **publish the standard** on 29 September 2023.

What is next

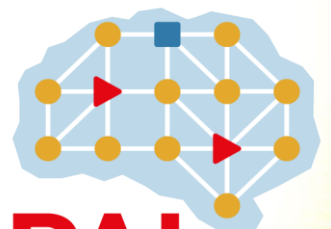


- MMC-DC intends to develop:
 - **Reference software** for all use cases.
 - **Conformance testing** for all technologies.
 - **Performance Assessment.**
- MMC-DC also intends to explore the relationship of **Large Language Models** with the **MPAI-AIM component-based approach.**



We look forward to working
with you
on this exciting MPAI project!

Join MPAI
Share the fun
Build the future



MPAI.
community