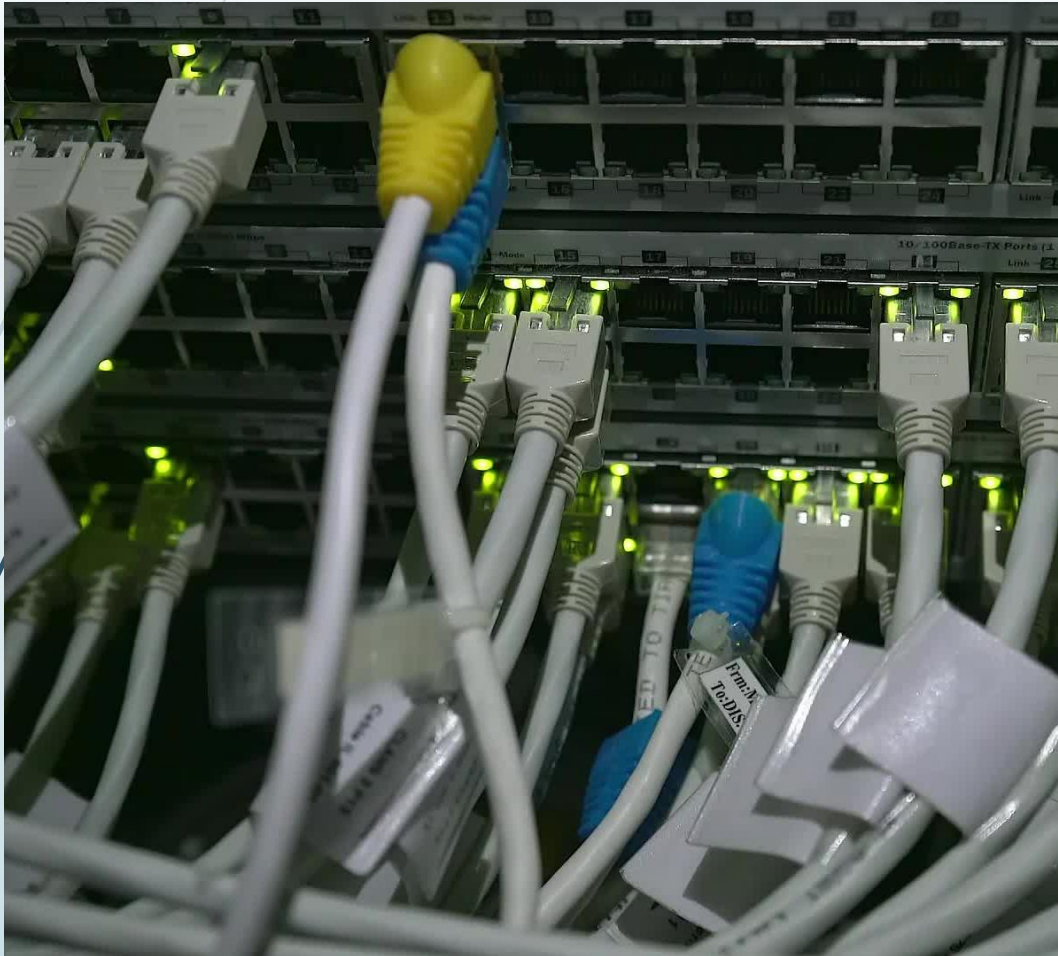




Avatar Representation and Animation (MPAI-ARA)

08 and 15 UTC, 07 September 2023

Contents of presentation



- [About MPAI](#)
- [About MPAI-ARA](#)
- [Avatar-Based Videoconference](#)
- [ARA-ABV Subsystems](#)
- [MPAI-ARA Data Types](#)
- [What's next](#)



About MPAI

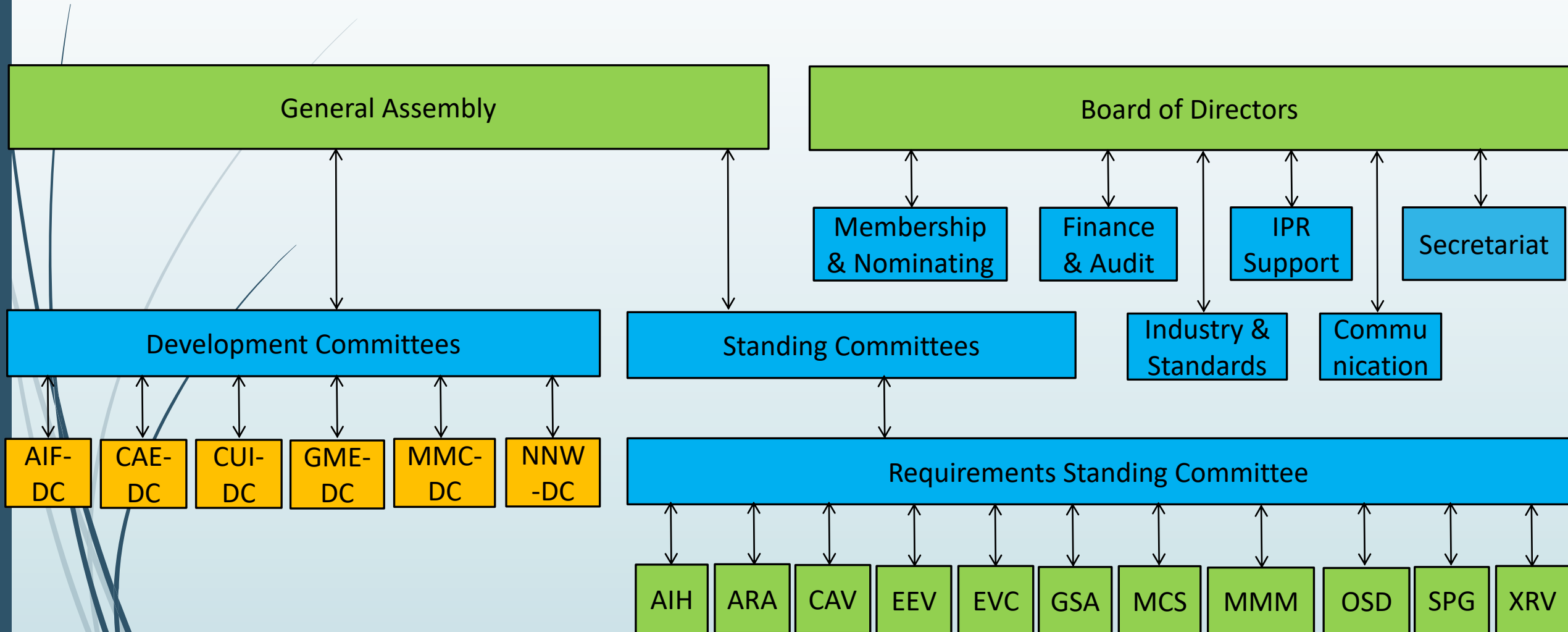
MPAI stands for Moving Picture, Audio, and Data Coding by Artificial Intelligence.

International, unaffiliated, non-profit SDO.

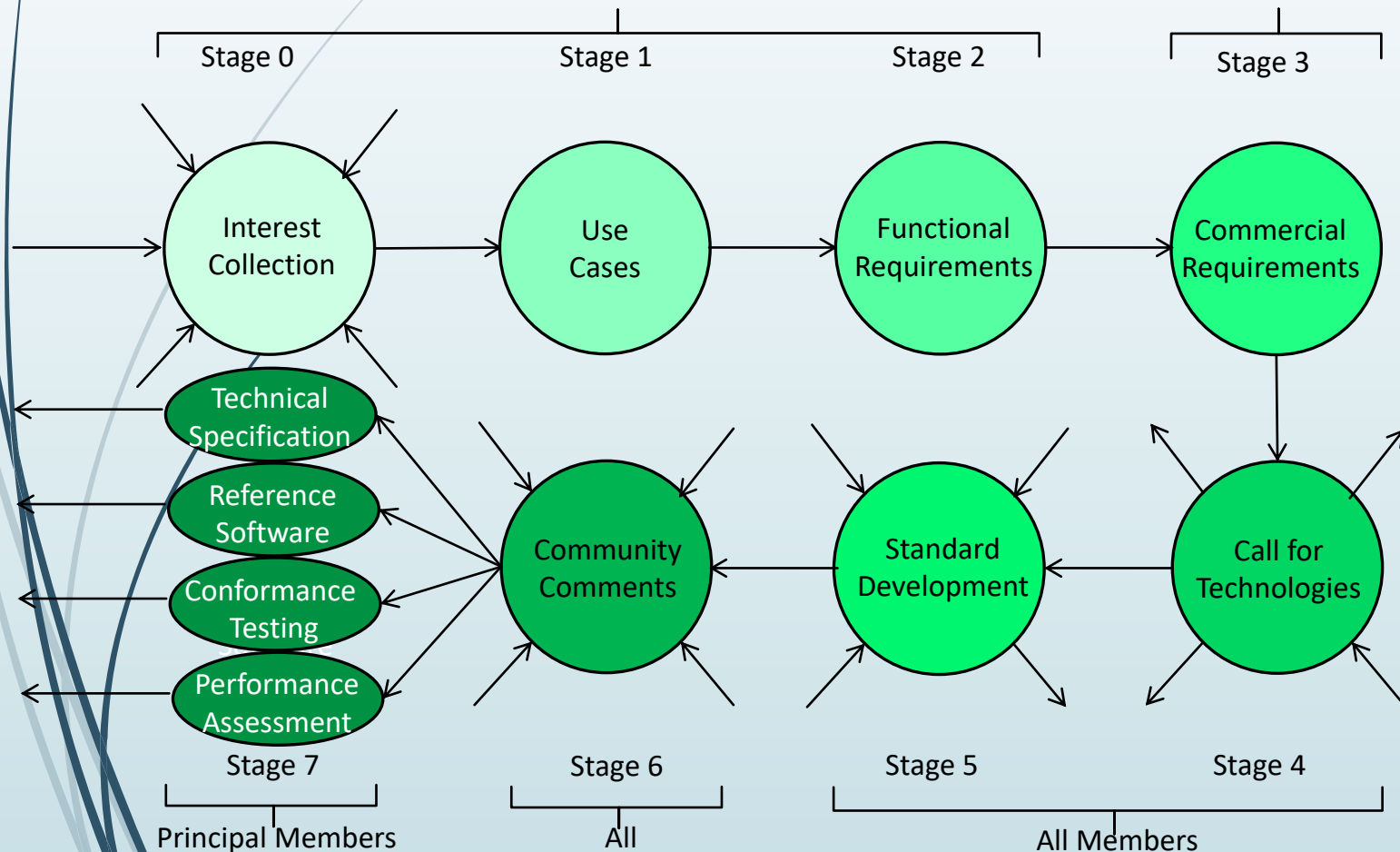
Developing AI-based data coding standards.

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**.

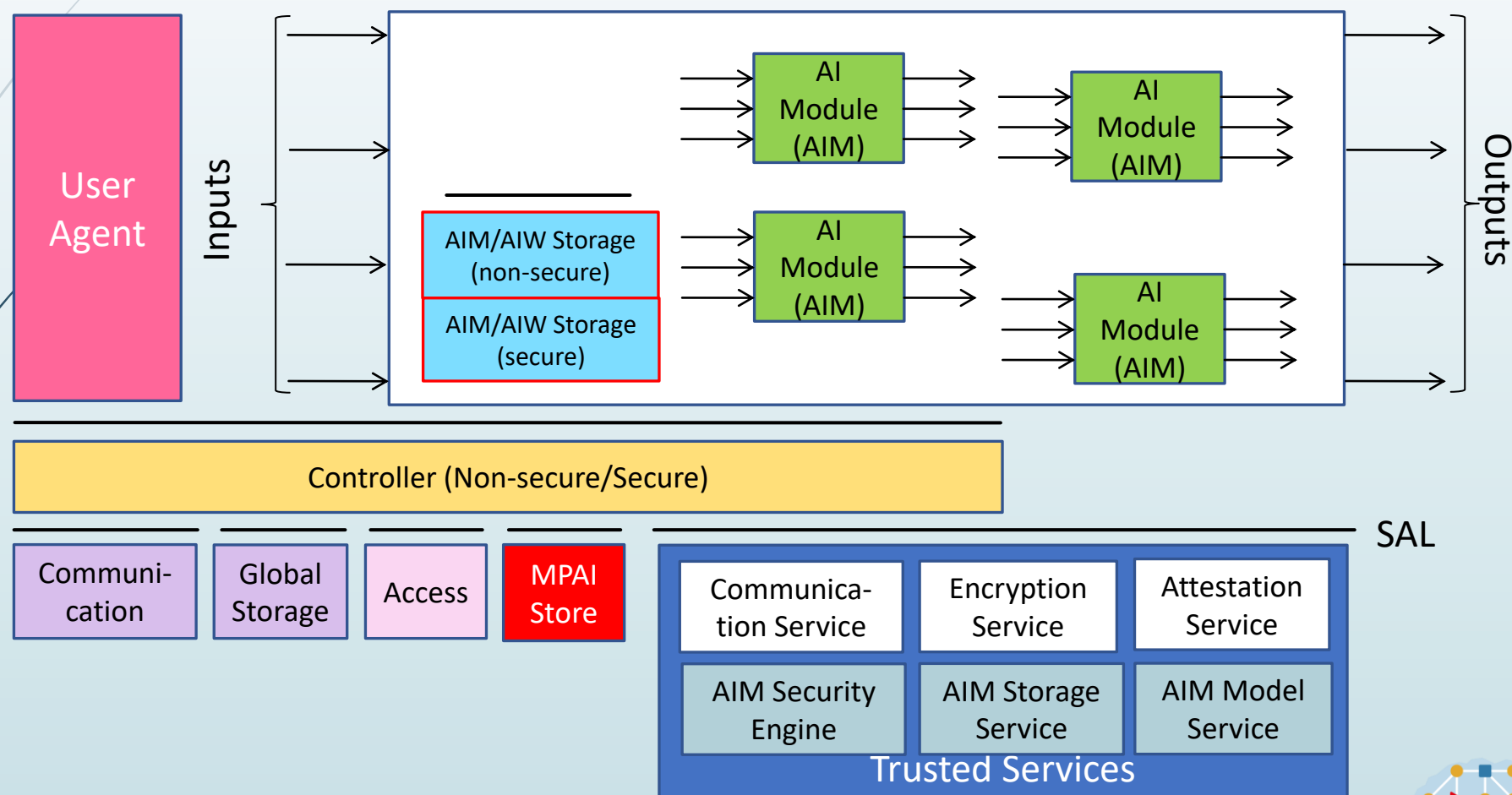
*MPAI's AI
standardisation is
“component-based”.*

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

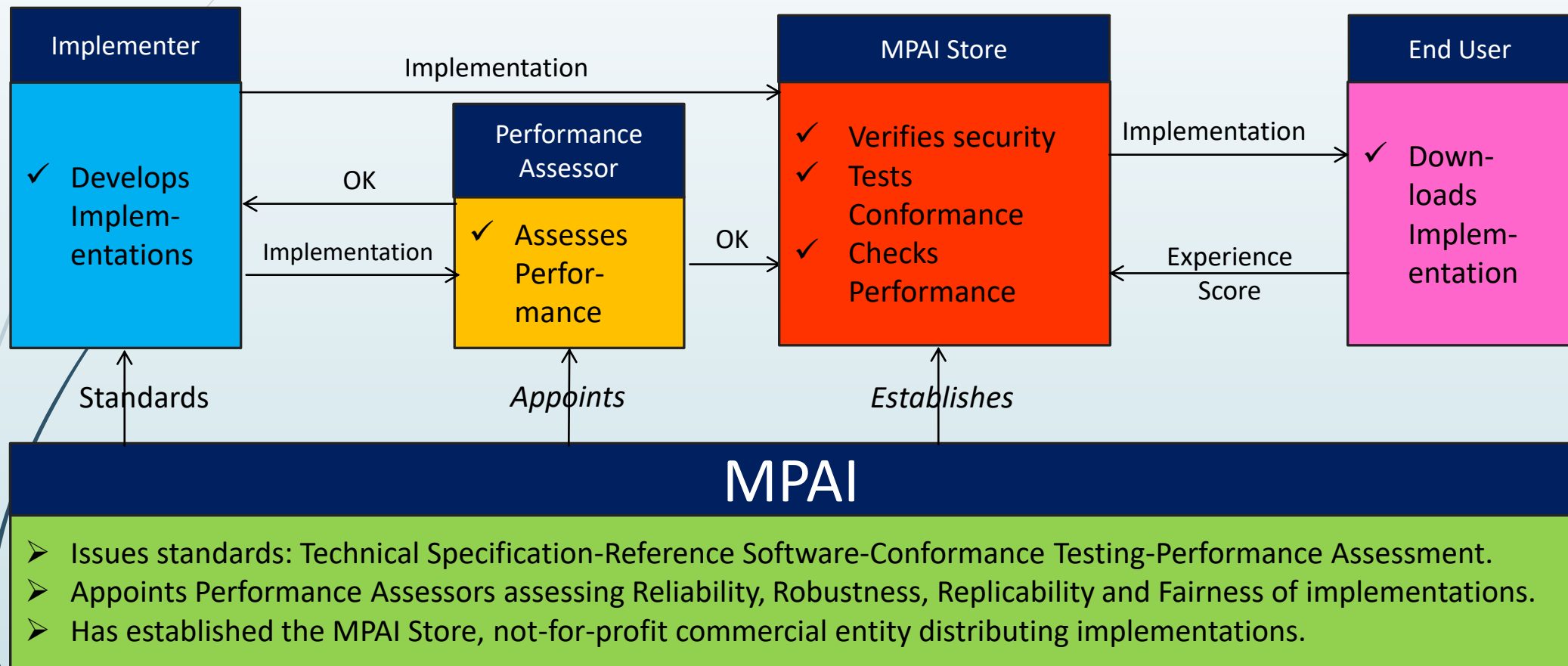


A sustainable MPAI Ecosystem

- **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).***

The MPAI ecosystem



More published MP AI standards

4 Technical Specifications

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

2 Technical Reports

- 1 - MP AI Metaverse Model (MP AI-MMM) – Functionalities*
- 2 - MP AI Metaverse Model (MP AI-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-ARA

Why a standard for digital humans

- Many computer-created objects called “**digital humans**” rendered with a human appearance.
- In most cases, creation, animation, and rendering **in a closed environment**. → No need for standards.
- In a communication context, e.g., in an interoperable metaverse, digital humans may **not** be **constrained to be in a closed environment**.
- If a sender requires that a remote receiving client **reproduce a digital human as intended by the sender**, standards are needed.
- Technical Specification: Avatar Representation and Animation is a **first response to this need**.



Avatar Representation and Animation

- Objective1: To enable a user to **reproduce a virtual environment as intended**.
- Objective2: to enable a user to **reproduce a sender's avatar and its animation as intended by the sender**.
- Objective3: to **estimate the personal status of a human or avatar**.
- Objective4: to **display an avatar with a selected personal status**.
- *Definition: **Personal Status** is the ensemble of information internal to a person, including Emotion, Cognitive State, and Attitude.*

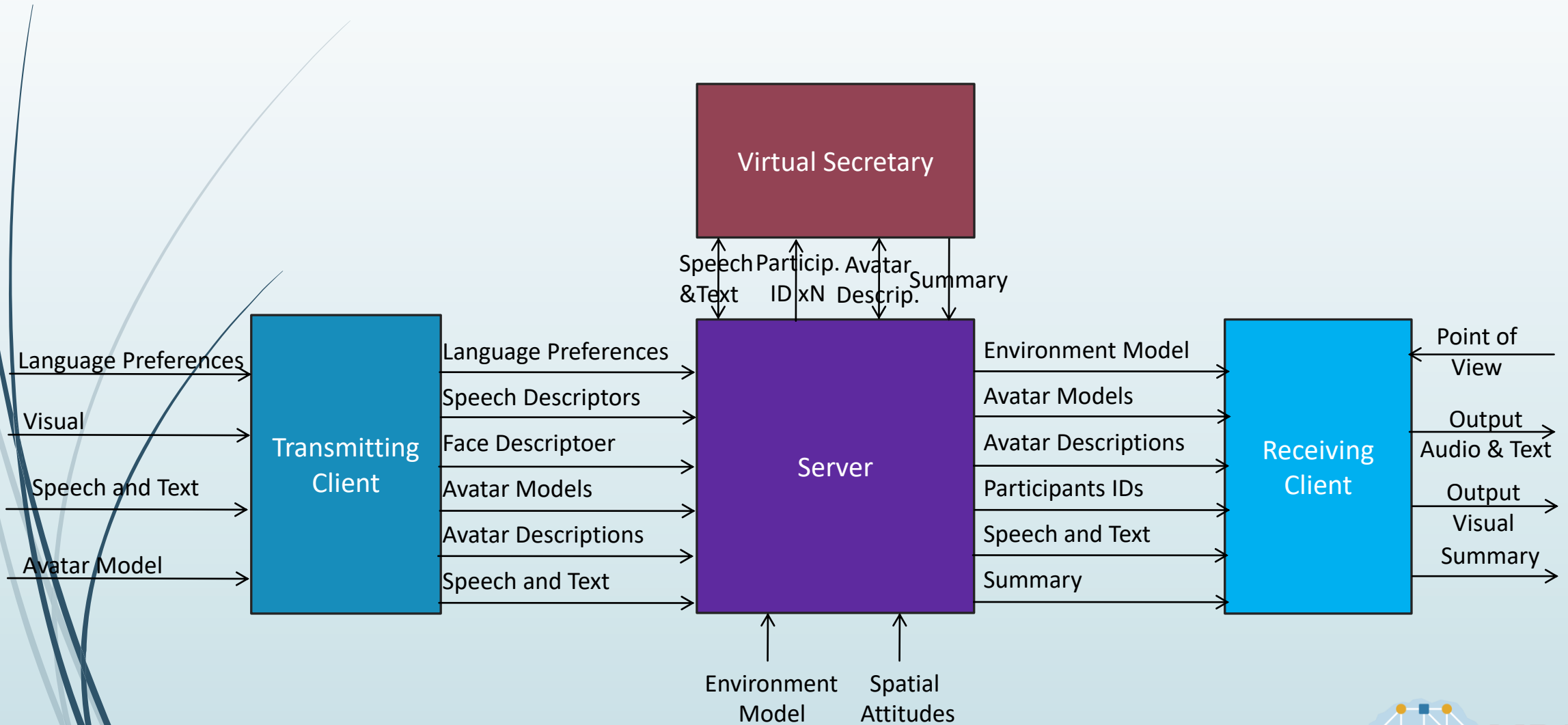


Avatar-Based Videoconference

Avatar-Based Videoconference (ABV) Use Case



Reference Model of Avatar-Based Videoconference



Functions of Avatar-Based Videoconference AIWs/1

Remotely located **Transmitting Clients** send:

1. At the beginning:

1. Receives and sends Avatar Model(s) and Language Preferences to server.
2. Sends Speech Object and Face Object to server for Authentication.

2. Continuously sends:

1. Avatar Descriptors and Speech to Server.

Functions of Avatar-Based Videoconference AIWs/2

The **Server**:

1. At the beginning:

1. Selects an Environment, e.g., a meeting room.
2. Equips the room with objects, i.e., meeting table and chairs.
3. Places Avatar Models around the table.
4. Distributes Environment, Avatars, and their positions to all receiving Clients.
5. Authenticates Speech and Face Objects

2. Continuously:

1. Translates Speech from participants according to Language Preferences.
2. Sends Avatar Descriptors and Speech to receiving Clients.

Functions of Avatar-Based Videoconference AIWs/3

Virtual Secretary

1. Receives Text, Speech, and Avatar Descriptors of conference participants.
2. Recognises Speech streams.
3. Refines Recognised Text and extracts Meaning.
4. Extracts Avatars' Personal Status.
5. Produces a Summary.
6. Produces Edited Summary using comments.
7. Produces Text and Personal Status.
8. Creates Speech and Avatar Descriptors from Text and Personal Status.

Functions of Avatar-Based Videoconference AIWs/4

Receiving Clients:

1. At the beginning:

1. Environment Model
2. Avatar Models
3. Spatial Attitudes

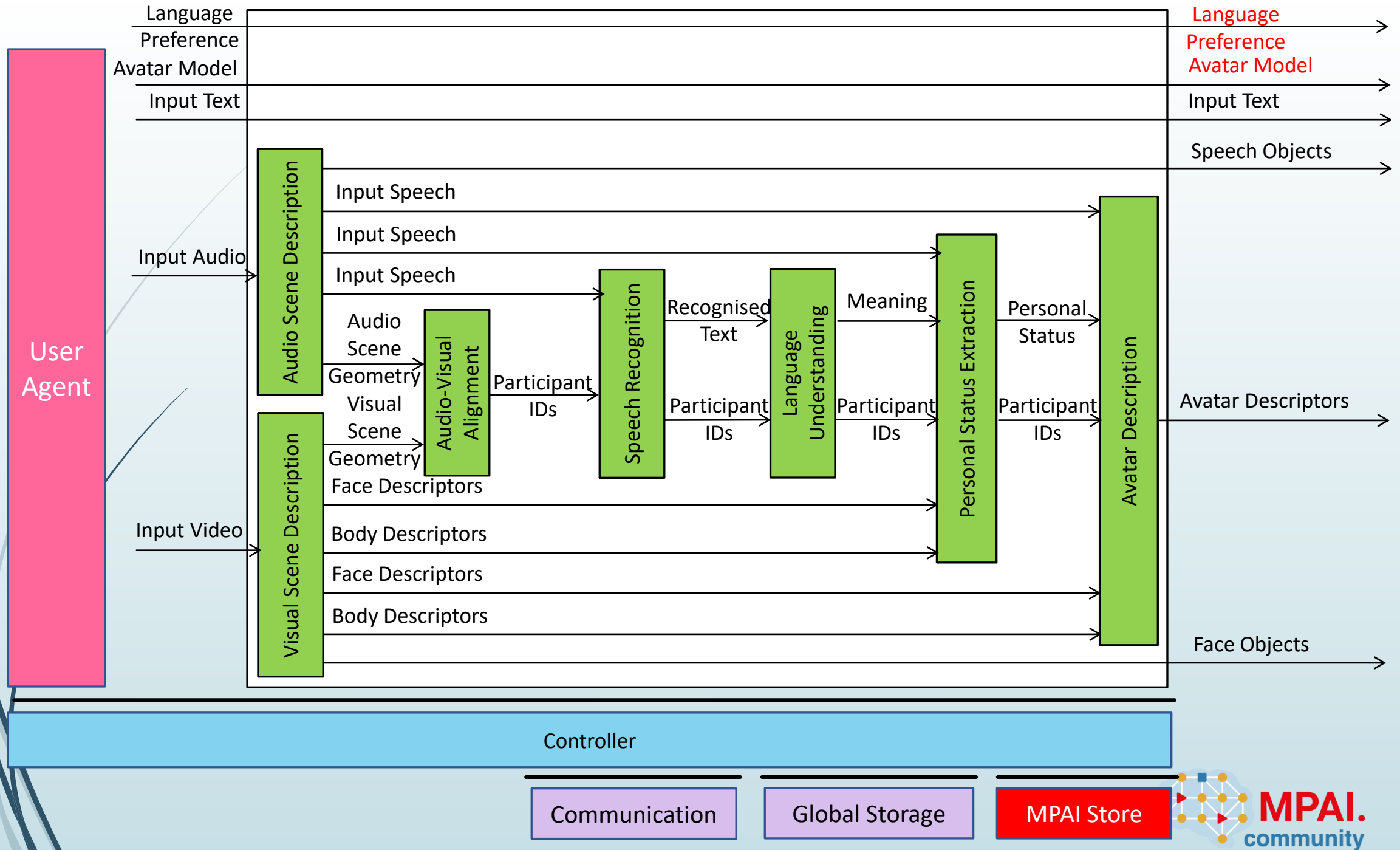
2. Continuously:

1. Creates Audio and Visual Scene Descriptors.
2. Renders the Audio-Visual Scene from the Point of View selected by Participant.



ARA-ABV Subsystems

ARA-ABV: Transmitting Client



Input/output Data of Transmitting Client

Input	Comments
Language Preference	The language participant wishes to speak and hear at the videoconference.
Input Text	Chat text used for Virtual Secretary/Participants communication.
Avatar Model	The avatar model selected by the participant.
Input Audio	Environment Audio with Participant Speech.
Input Video	Video of participants' body.
Output	Comments
Language Preference	As in input.
Input Text	As in Input.
Avatar Descriptors	As in input.
Speech Object	Speech for Authentication
Participant's Speech	Speech as separated from Environment Audio.
Avatar Descriptors	Descriptors produced by Transmitting Client.
Participant ID	ID of a Participant in the room
Face Object	Face of a Participant

Functions of Transmitting Client's AIMs

AIM	Function
Audio Scene Description	Provides audio objects and their audio scene geometry.
Visual Scene Description	Provides visual objects and their visual scene geometry.
Audio-Visual Alignment	Assigns identifiers to Audio, Visual and Audio-Visual Objects.
Speech Recognition	Recognises the speech of a human.
Language Understanding	Extracts Meaning from Recognised Text.
Personal Status Extraction	Extracts Personal Status from Speech, Meaning, and Face and Body Descriptors.
Avatar Description	Provides the Descriptors of the human represented by the Avatar.

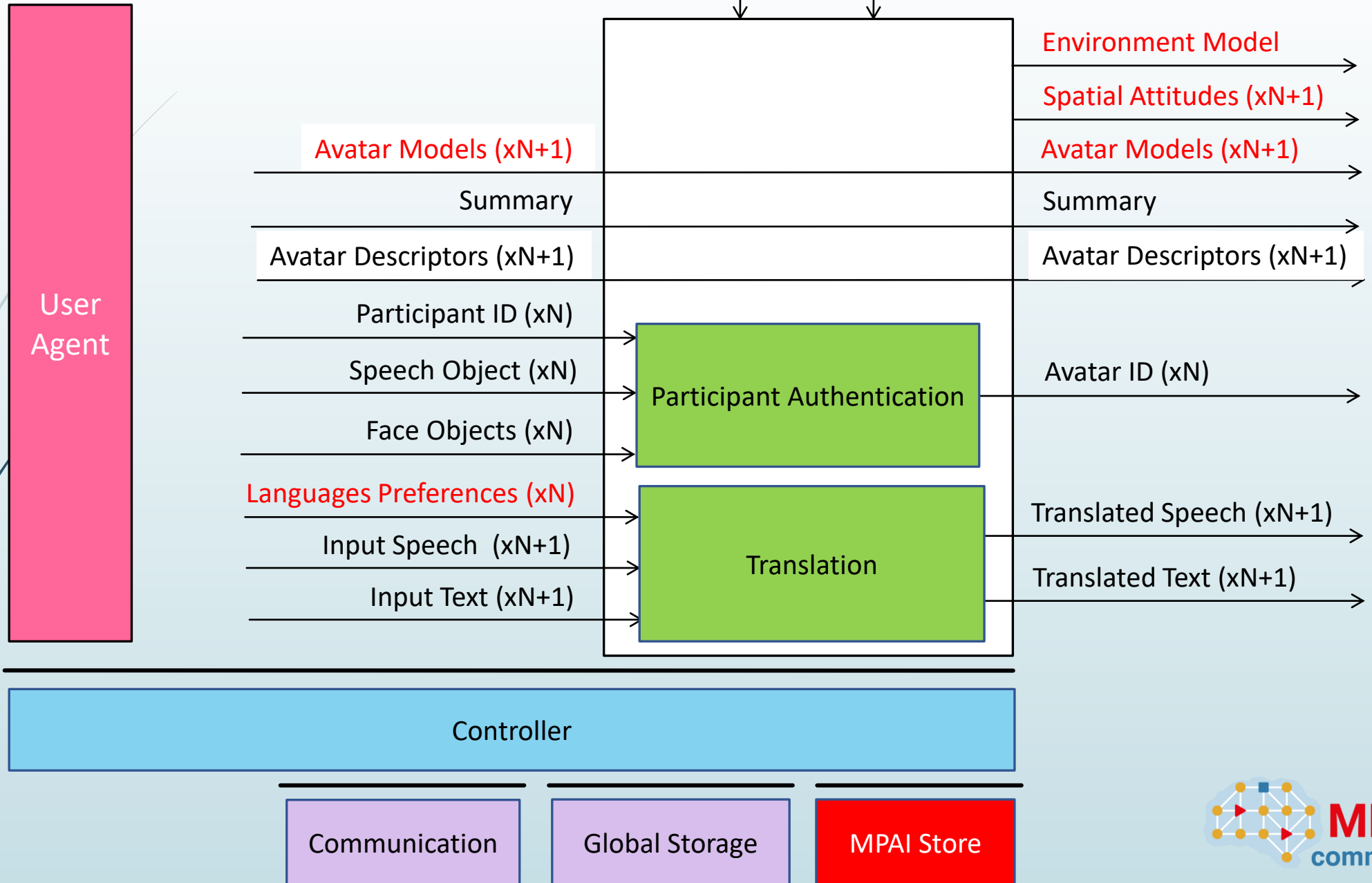
I/O Data of Transmitting Client AIMS

AIMAudio	Input	Output
Audio Scene Description	Input Audio	Audio Scene Descriptors
Visual Scene Description	Input Video	Face Descriptors, Body Descriptors Face Object
Audio-Visual Alignment	Audio Scene Geometry Visual Scene Geometry	Participant IDs
Speech Recognition	Speech Objects, Participant IDs	Recognised Text
Language Understanding	Recognised Text Participant IDs	Refined Text, Meaning
Personal Status Extraction	Meaning, Speech, Face Descriptors Body Descriptors, Participant IDs	Personal Status
Avatar Description	Personal Status, Face Descriptors Gesture Descriptors, Participants IDs	Avatar Descriptors.

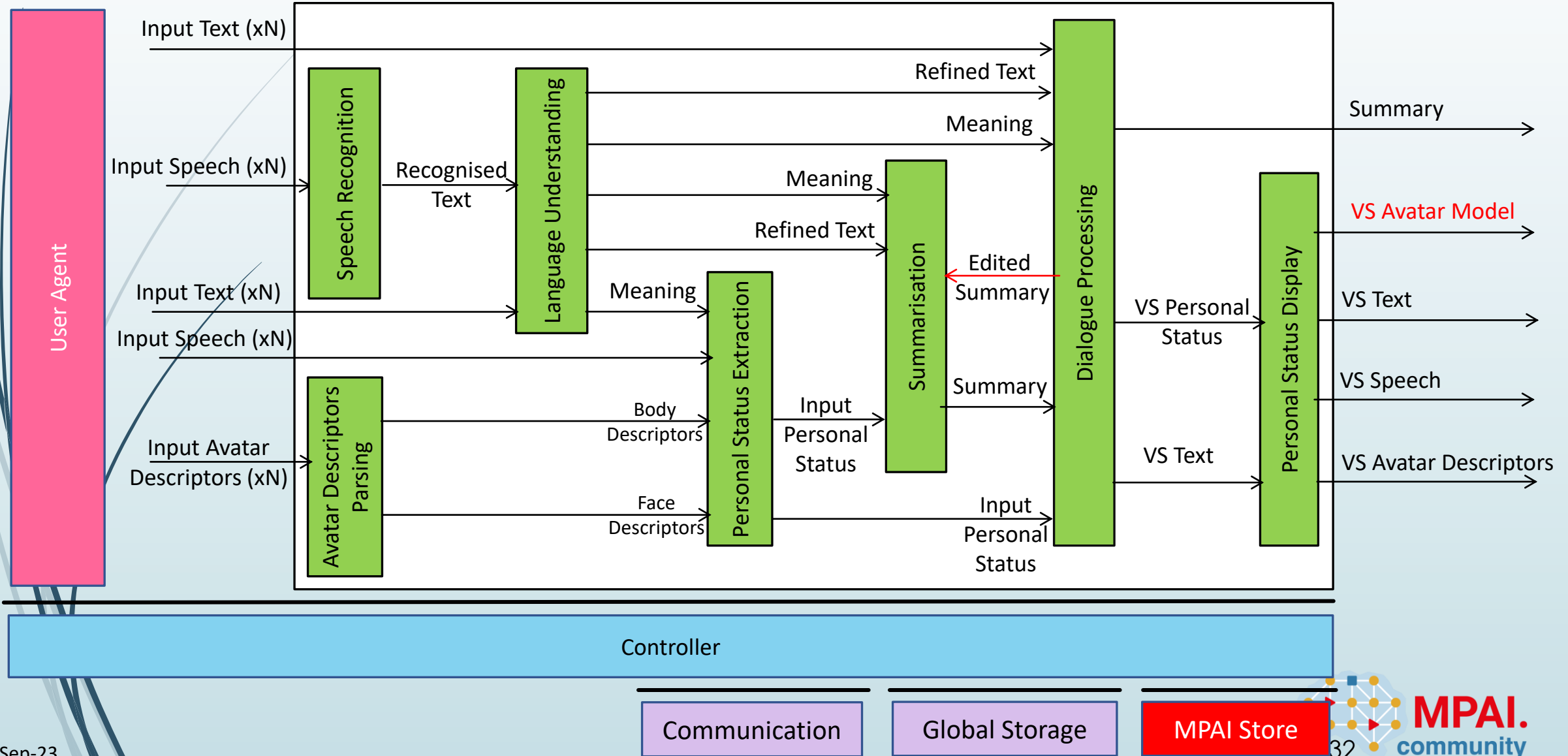
JSON metadata

- The Transmitting Client's AIW executed in an AIF.
- The capabilities of the AIF described by standard JSON metadata.
- The capabilities of the AIW described by standard JSON metadata.
- The capabilities of (Composite) AIMs described by a standard JSON metadata (like AIW metadata).
- An AIF downloads suitable AIW and AIMs from the MPAI Store using the JSON metadata.

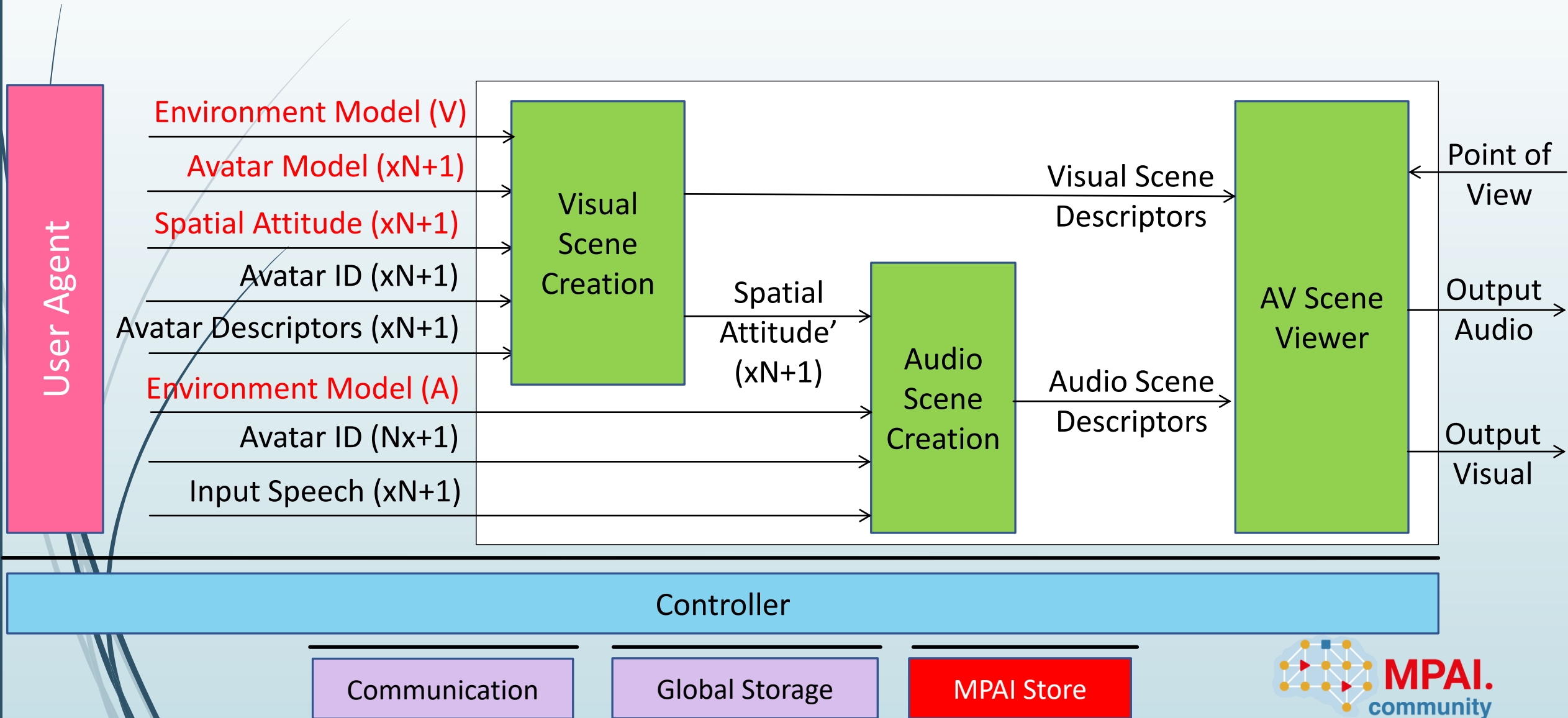
ARA-ABV: Server



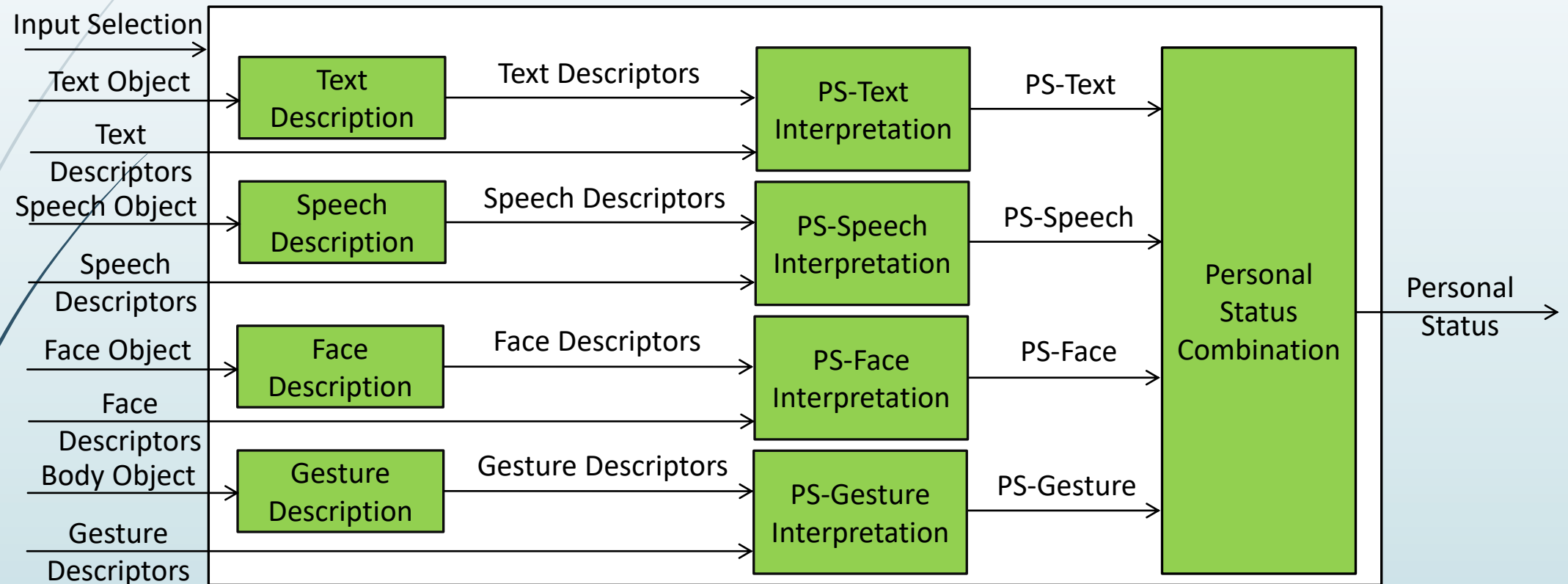
ARA-ABV: Virtual Secretary for Videoconference



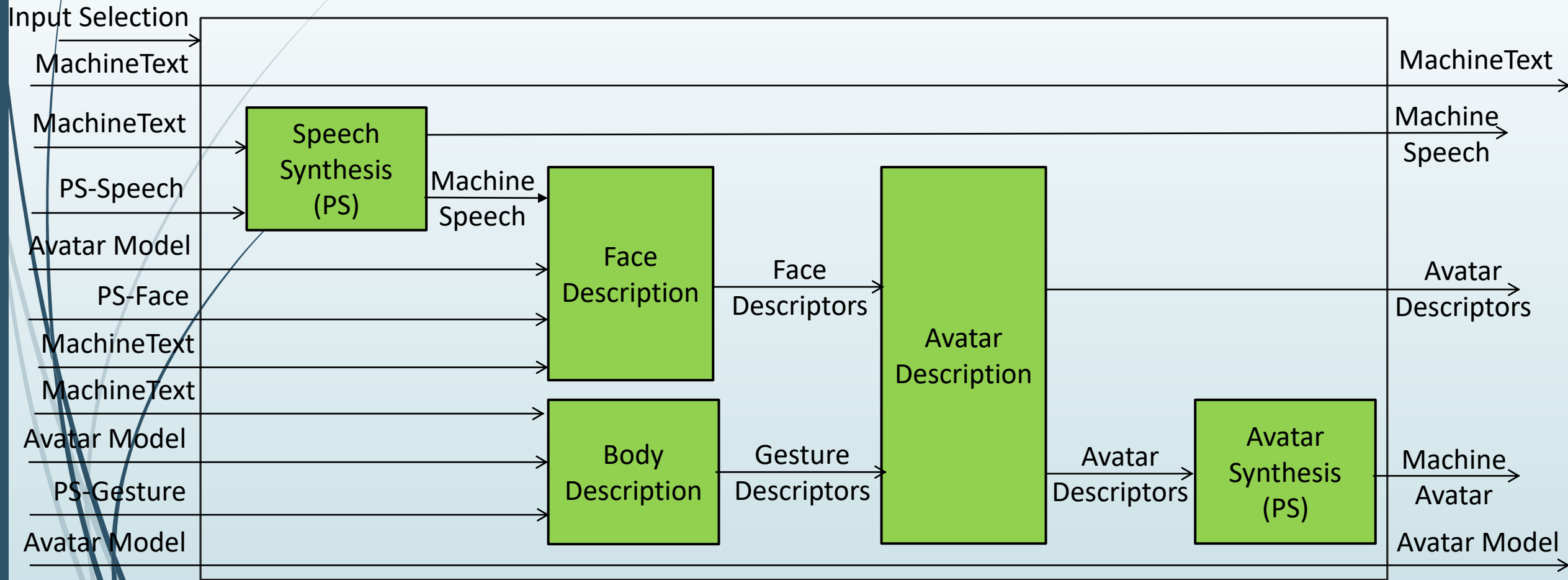
ARA-ABV: Receiving Client



Personal Status Extraction



Personal Status Display



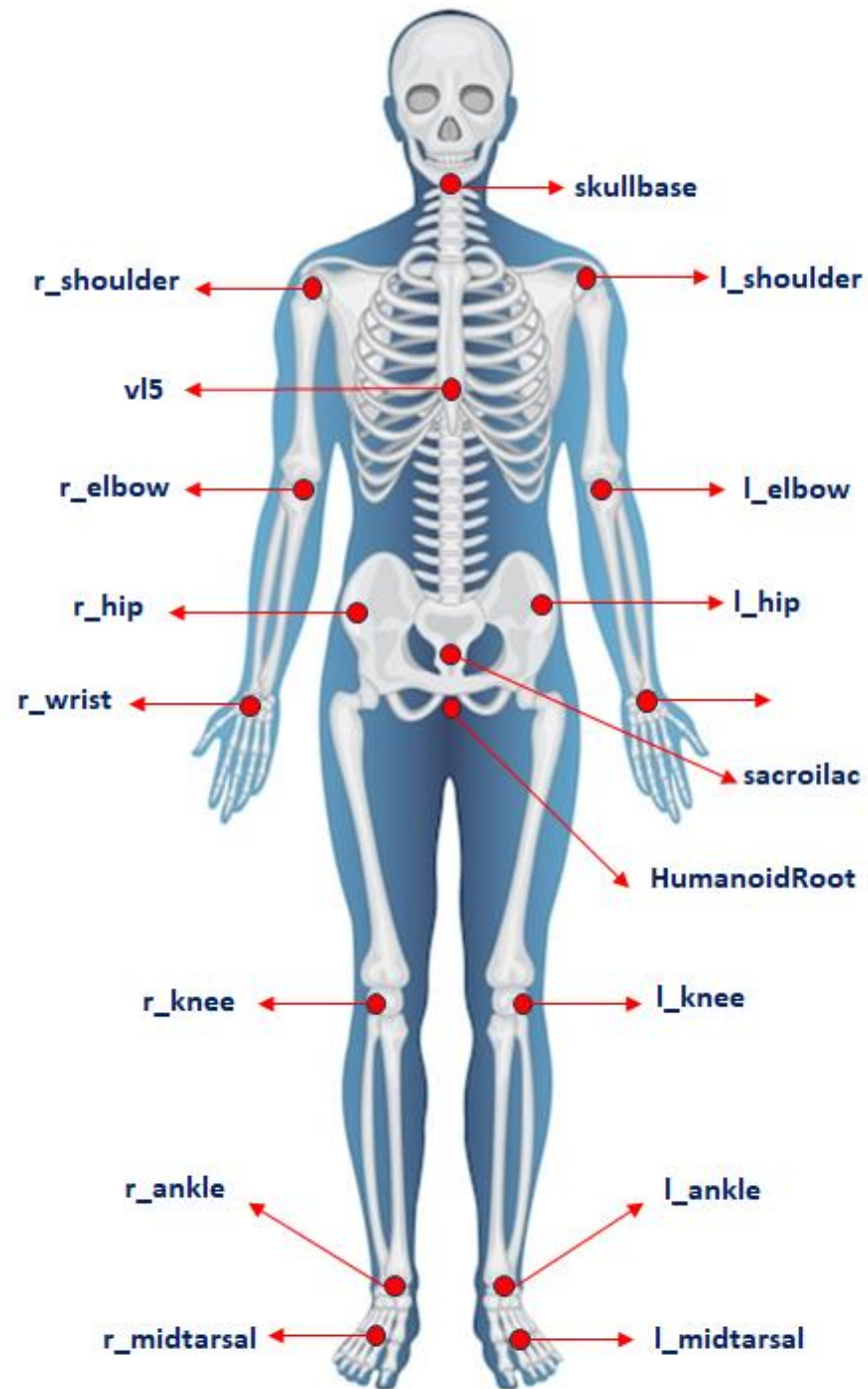


MPAI-ARA Data Types

Data Types

Name of Data Format	Specified by
Environment	OSD
Body Model	ARA
Body Descriptors	ARA
Face Model	ARA
Face Descriptors	ARA
Avatar Model	ARA
Avatar Descriptors	ARA
Spatial Attitude	OSD
Audio Scene Descriptors	CAE
Visual Scene Descriptors	OSD
Text	MMC
Language identifier	MMC
Meaning	MMC
Personal Status	MMC

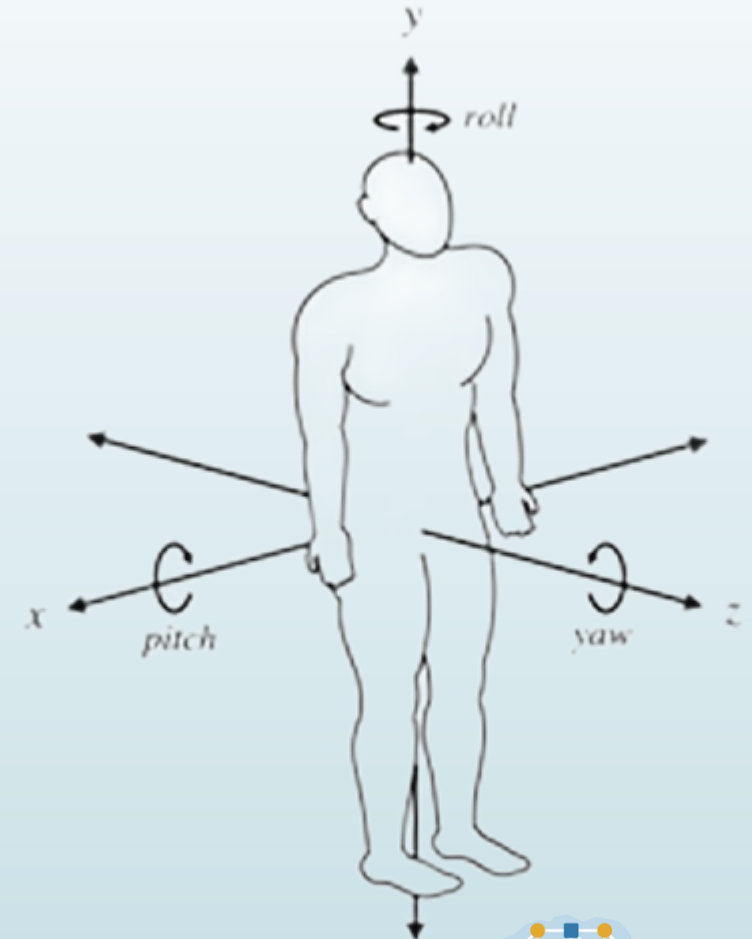
Body Model



Body Descriptors

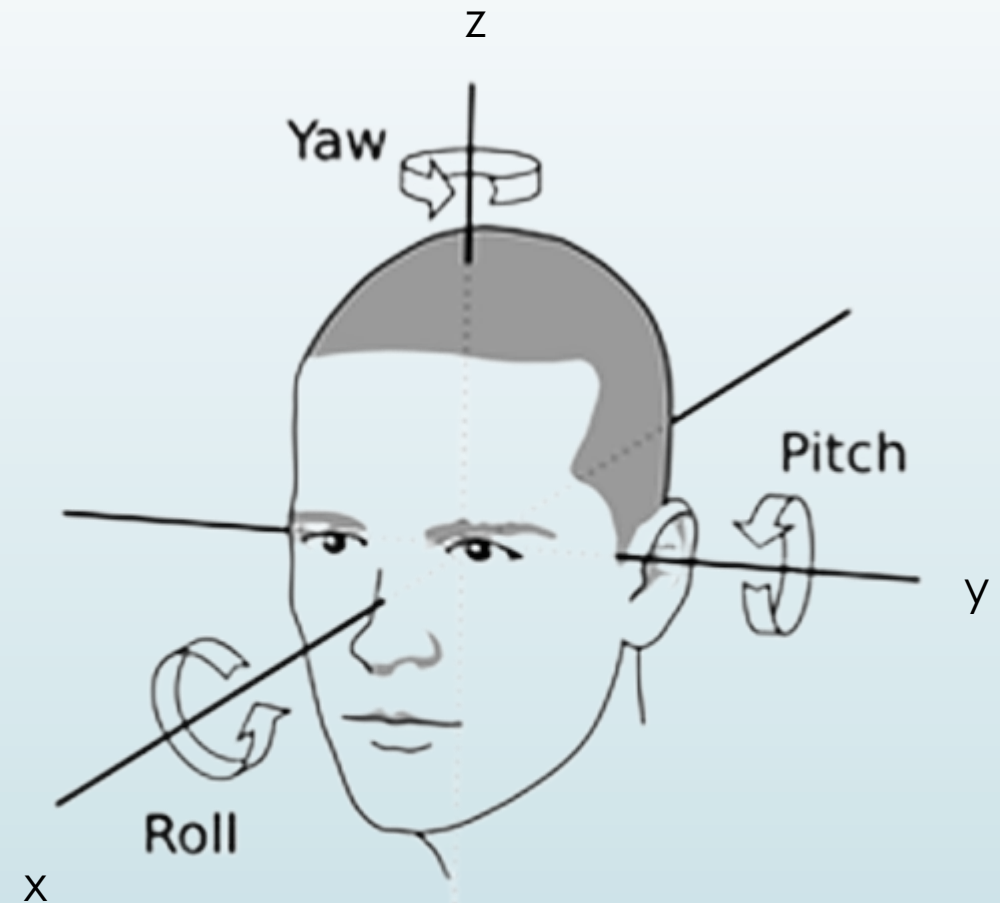
Body Descriptors describe the movement of the root and joints:

1. Position and Orientation of the root with respect to the Position.
2. Rotation angle of the y axis in the Figure.
3. Rotation angles of the joints.
4. The rotation of the head is treated as any other joint.



Head Descriptors

- The Head is described by:
 - Roll: head moves toward one of the shoulders.
 - Pitch: head moves up and down.
 - Yaw: head rotates left to right (around the vertical axis of the head).
- Roll, Pitch, and Yaw of a Head in Figure.



Face Descriptors/1

AU	Description	Facial muscle
<u>1</u>	Inner Brow Raiser	Frontalis, pars medialis
<u>2</u>	Outer Brow Raiser	Frontalis, pars lateralis
<u>4</u>	Brow Lowerer	Corrugator supercilii, Depressor supercilii
<u>5</u>	Upper Lid Raiser	Levator palpebrae superioris
<u>6</u>	Cheek Raiser	Orbicularis oculi, pars orbitalis
<u>7</u>	Lid Tightener	Orbicularis oculi, pars palpebralis
<u>9</u>	Nose Wrinkler	Levator labii superioris alaeque nasi
<u>10</u>	Upper Lip Raiser	Levator labii superioris
<u>11</u>	Nasolabial Deepener	Zygomaticus minor
<u>12</u>	Lip Corner Puller	Zygomaticus major
<u>13</u>	Cheek Puffer	Levator anguli oris (a.k.a. Caninus)
<u>14</u>	Dimpler	Buccinator
<u>15</u>	Lip Corner Depressor	Depressor anguli oris (a.k.a. Triangularis)
<u>16</u>	Lower Lip Depressor	Depressor labii inferioris
<u>17</u>	Chin Raiser	Mentalis
<u>18</u>	Lip Pucker	Incisivii labii superioris and Incisivii labii inferioris
<u>20</u>	Lip stretcher	Risorius with platysma

Face Descriptors/2

AU	Description	Facial muscle
22	Lip Funneler	Orbicularis oris
23	Lip Tightener	Orbicularis oris
24	Lip Pressor	Orbicularis oris
25	Lips part**	Depressor labii inferioris or relaxation of Mentalis, or Orbicularis oris
26	Jaw Drop	Masseter, relaxed Temporalis and internal Pterygoid
27	Mouth Stretch	Pterygoids, Digastric
28	Lip Suck	Orbicularis oris
41	Lid droop**	Relaxation of Levator palpebrae superioris
42	Slit	Orbicularis oculi
43	Eyes Closed	Relaxation of Levator palpebrae superioris; Orbicularis oculi, pars palpebralis
44	Squint	Orbicularis oculi, pars palpebralis
45	Blink	Relaxation of Levator palpebrae superioris; Orbicularis oculi, pars palpebralis
46	Wink	Relaxation of Levator palpebrae superioris; Orbicularis oculi, pars palpebralis
61	Eyes turn left	
62	Eyes turn right	
63	Eyes up	
64	Eyes down	

Avatar Descriptors

Variable name	Code
Avatar ID	ID of Avatar
Timestamp type	Absolute/relative
Timestamp value	Time from the start of duration (In seconds)
Space type	Global/local
Space value	Metres
Spatial Attitude	Metres/Degrees
Body Descriptors	From ARA
Face Descriptors	From ARA
Speech Segment	Compression identifier
Time duration	In second
Text snippet	Text



What's next

Anybody is entitled to comment on MPAI-ARA



- The MPAI-MMC Version 2 Working Draft is available ([html](#), [pdf](#)).
- Anybody is entitled to comment on the WD
- Responses should reach secretariat@mpai.community by 2023/09/27T23:59 UTC
- MPAI plans: **publish MPAI-ARA – Architecture** as an MPAI Technical Specification at the 36th General Assembly (29 September 2023).

What's next for MPAI-MMC?

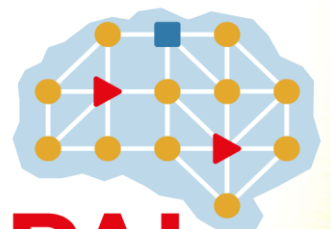
- Adoption of MPAI-ARA without modifications as an IEEE standard.
- Continue the development of Reference Software.
- Start the development of Conformance Testing.
- Study extensions of MPAI-ARA (e.g., compression of Avatar Description).





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

Join MPAI
Share the fun
Build the future



MPAI.
community