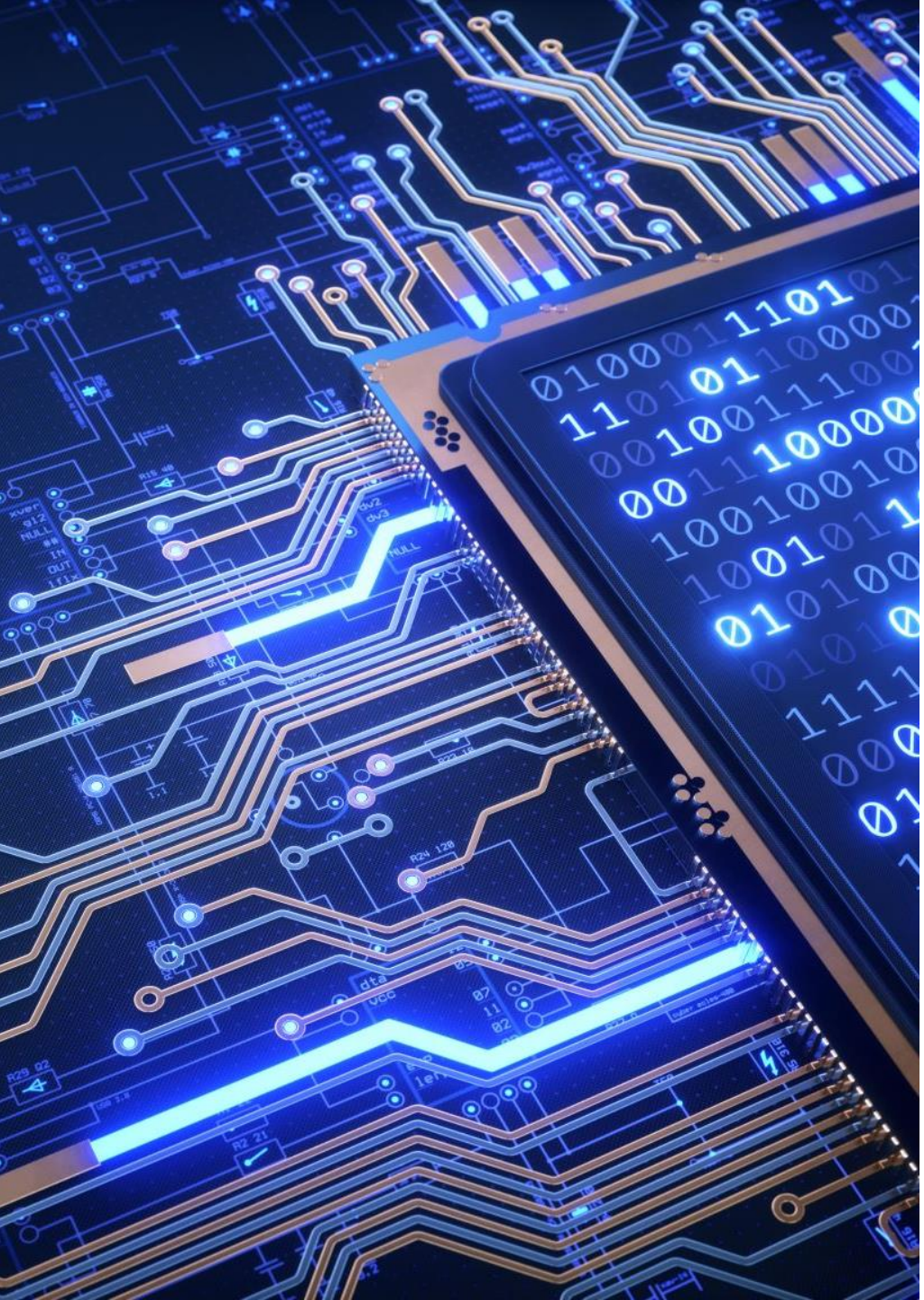
사실표준화기구 MPAI의 인공지능 표준화 현황

GISC 2021

한국전자통신연구원
최 미 란

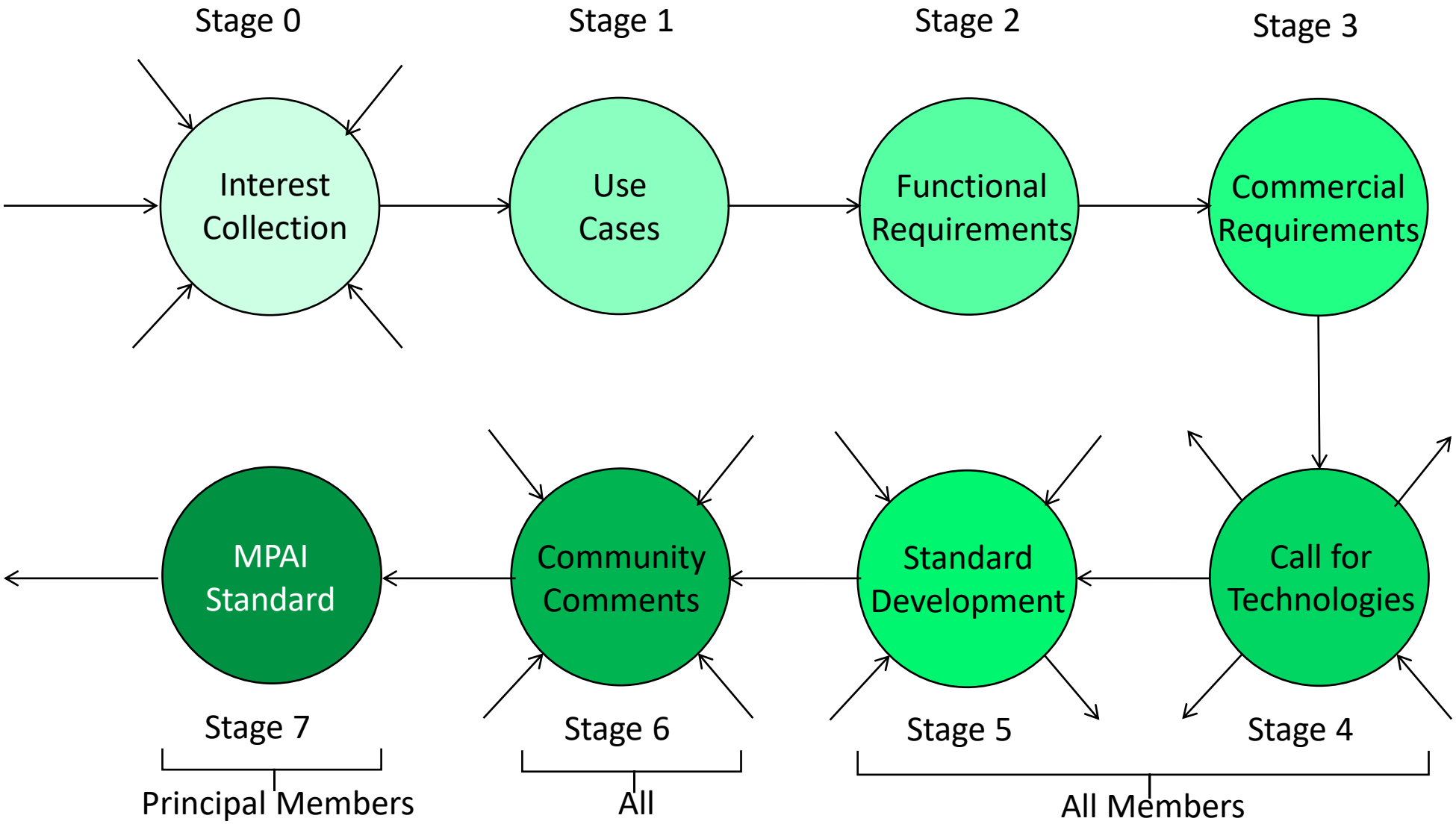


MPAI
Moving Picture, Audio and Data Coding by Artificial intelligence

- MPEG 창시자 Leonardo Chiariglione 가 설립한 비영리 국제표준단체
- AI 에 기반한 데이터 코딩 표준 개발 목표
- 2020년 9월 30일에 창립총회를 마치고 사실표준화 그룹 선언
- 현재 15개 국가의 40 개의 단체회원
  https://mpai.community

*Data coding: the **transformation of data from a format** into another format more suitable to an application.

MPAI 표준개발 프로세스



MPAI Work Plan/1

Published

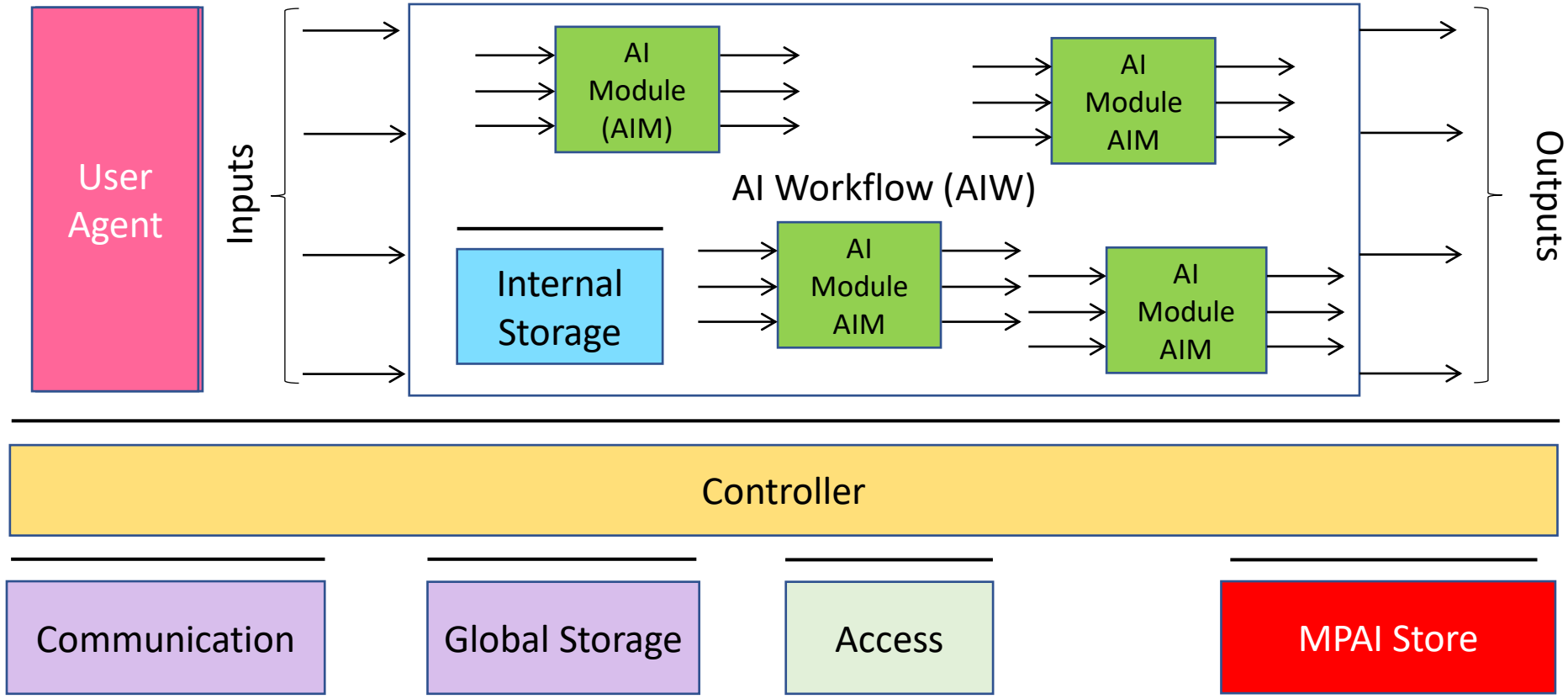
Planned: in 24 days – in 52 days

#	Work area	Project	Stage
1	AI Framework		6-SD
2	Context-based Audio Enhancement	Emotion Enhanced Speech	6-SD
		Audio Recording Preservation	
		Speech Restoration System	
		Enhanced Audioconference Experience	
3	Multimodal Conversation	Conversation with Emotion	7-MS
		Multimodal Question Answering	
		Automatic Speech Translation	
4	Compression and Understanding of Industrial Data	AI-based Performance Prediction	7-MS
5	Governance of the MPAI Ecosystem		7-MS

MPAI Work Plan/2

#	Work area	Project	Stage
6	Server-based Predictive Multiplayer Gaming	Data recovery and cheating detection	2-FR
7	Connected Autonomous Vehicles	Human-CAV interaction	2-FR
		Environment Sensing Subsystem	2-FR
		Autonomous Motion Subsystem	2-FR
		CAV-to-Everything	2-FR
		Motion Actuation Subsystem	2-FR
8	AI-Enhanced Video Coding	MPAI-EVC Evidence Project	2-FR
9	Mixed Reality Collaborative Spaces		2-FR
10	Integrative Genomic/Sensor Analysis		2-FR
11	Visual Object and Scene Description		1-UC
12	AI-based End-to-End Video Coding		0-IC

The MPAI-AIF (AI Framework) Model and Components

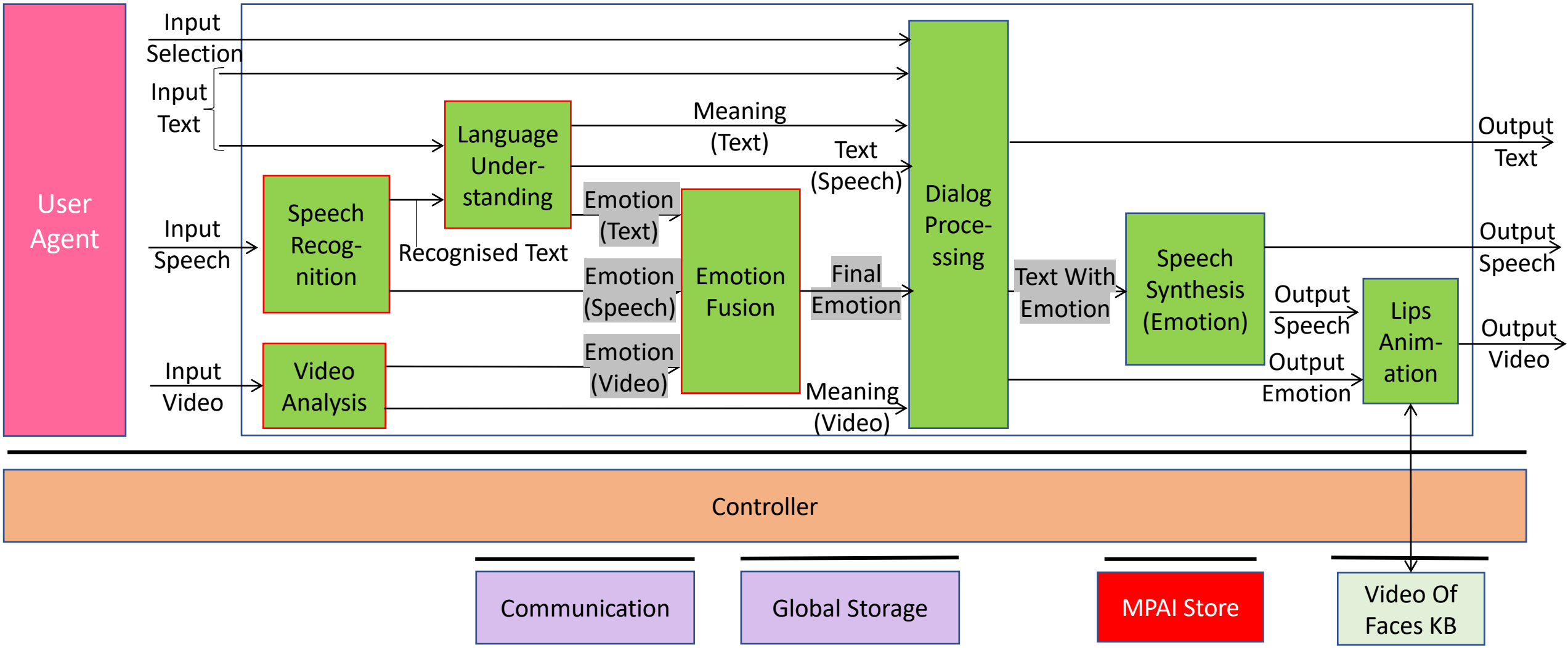


Multimodal Conversation (MPAI-MMC)

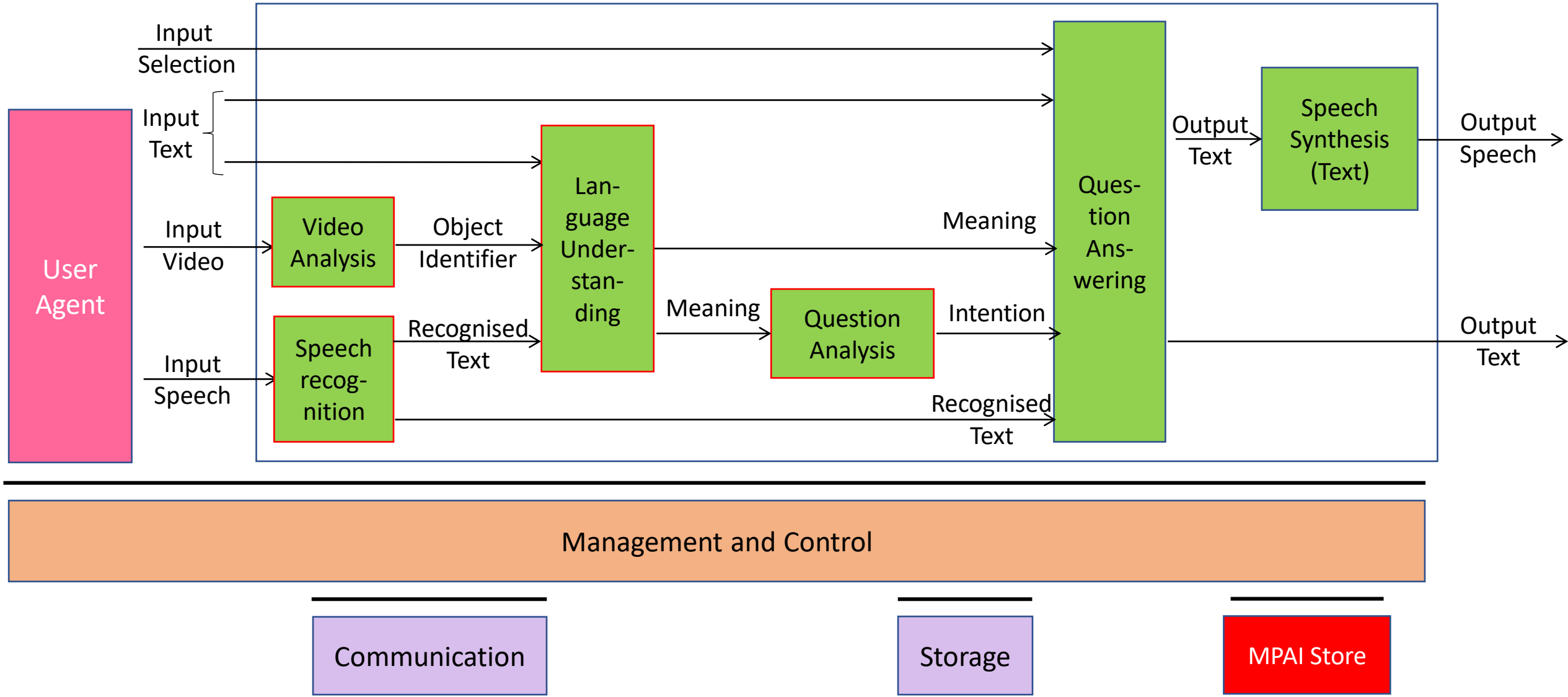
제안 기관: ETRI (CWE, MQA, PST), SpeechMorphing (PST), KLleon (CWE)

1. Common Characteristic
    1. Use of AI to enable a form of human-machine conversation that emulates human-human conversation in completeness and intensity
2. MPAI-MMC Use Cases:
    1. “Conversation with Emotion” (CWE):
       AV conversation with a machine impersonated by a synthetic voice and an animated face.
    2. “Multimodal Question Answering” (MQA): Request for information about a displayed object.
    3. Translate a spoken of text sentence preserving the speech features of the speaker in the translated speech:
        1. “Unidirectional Speech Translation” (UST).
        2. “Bidirectional Speech Translation” (BST).
        3. “One-to-Many Speech Translation” (MST).

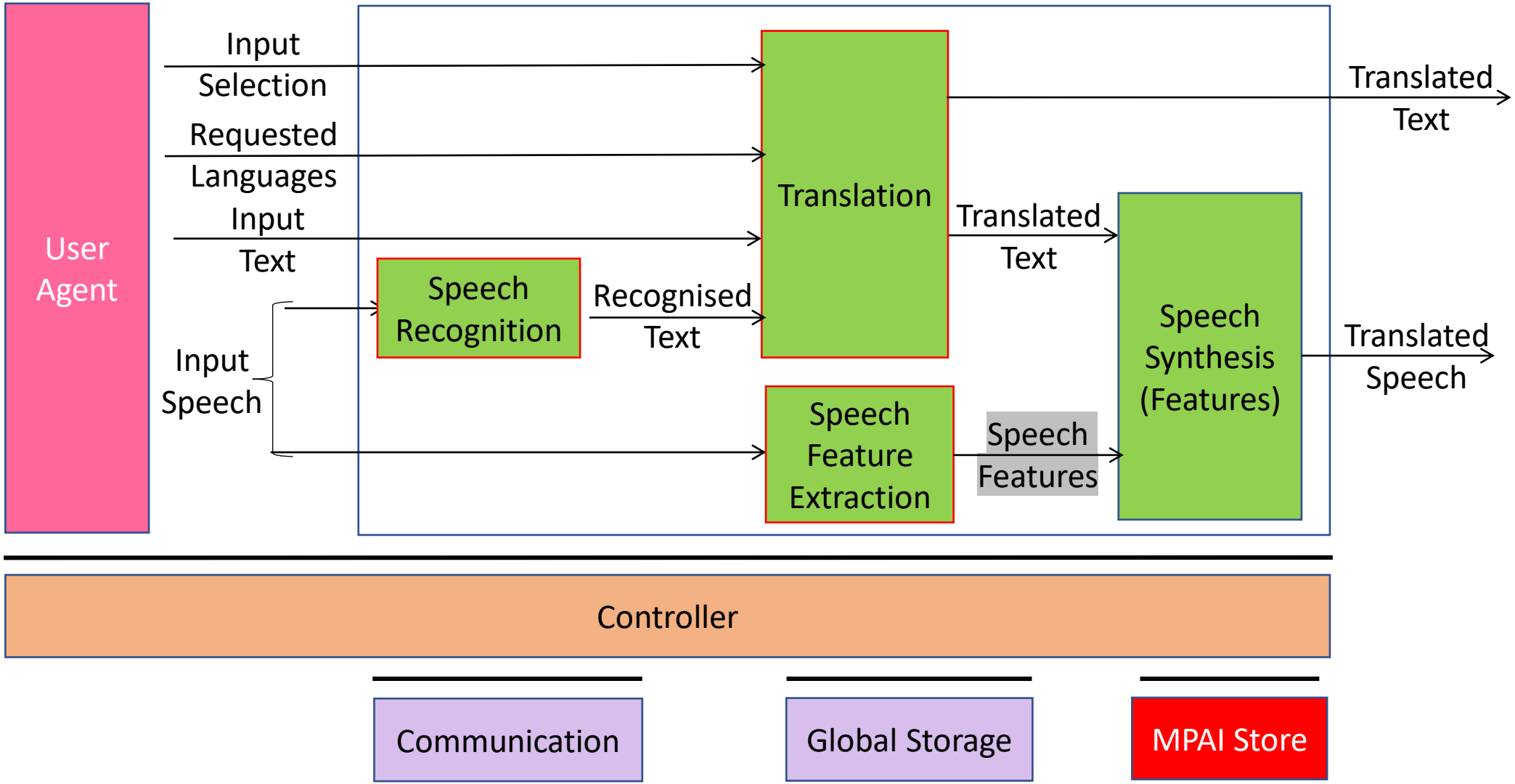
Conversation with Emotion



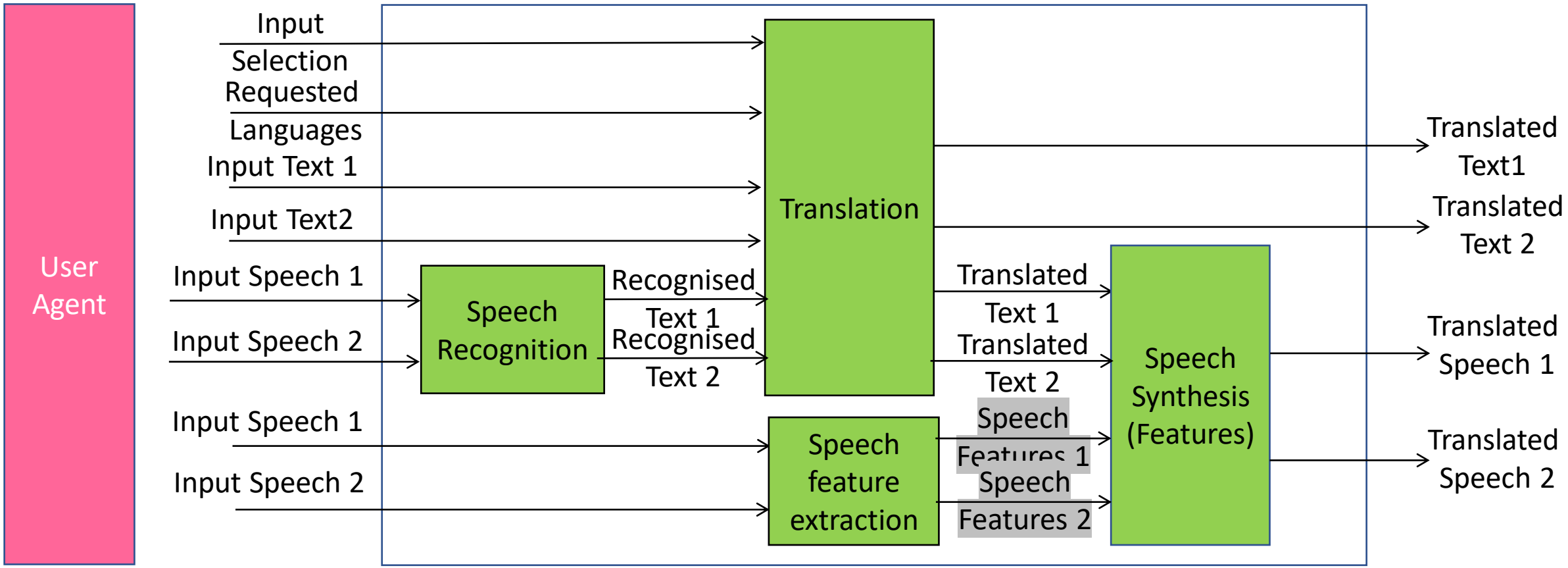
Multimodal Question Answering



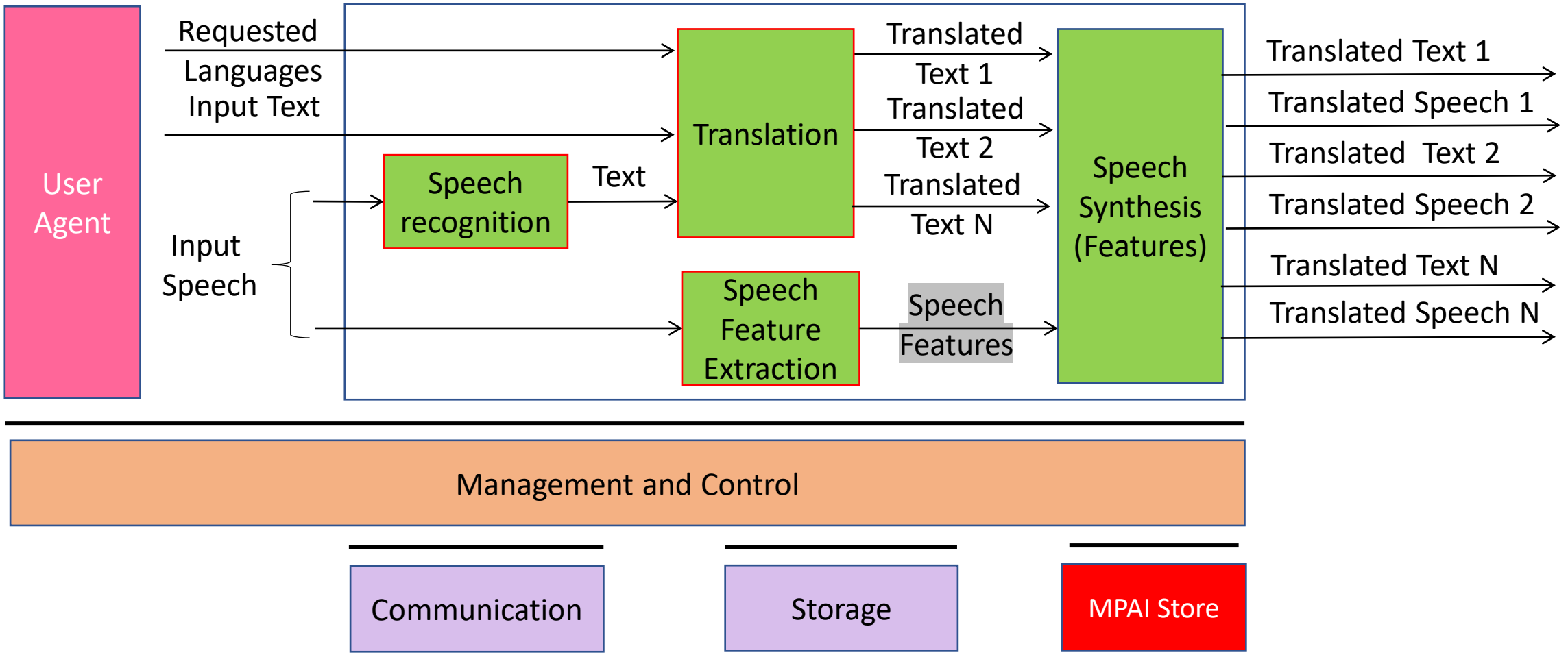
Unidirectional Speech Translation



Bidirectional Speech Translation



One-to-Many Speech Translation



MPAI-MMC data formats

1. Emotion
2. Intention
3. Language identifier
4. Meaning
5. Object Identifier
6. Speech
7. Speech Features
8. Text
9. Text with Emotion
10. Video
11. Video File
12. Video of Faces KB Query Format

Emotion – Syntax

```
{
  "$schema":"http://json-schema.org/draft-07/schema",
  "definitions":{
    "EmotionType":{
      "type":"object",
      "properties":{
        "emotionDegree":{
          "type":"{Enum high | Enum medium | Enum low}"
        },
        "emotionName":{
          "type":"string"
        },
        "emotionSetName":{
          "type":"string"
        }
      }
    },
    "type":"object",
    "properties":{
      "primary":{
        "$ref":"#/definitions/EmotionType"
      },
      "secondary":{
        "$ref":"#/definitions/EmotionType"
      }
    }
  }
```

Emotion – Semantics

Name	Definition
emotionType	Specifies the Emotion that the input carries.
emotionDegree	Specifies the Degree of Emotion as one of “Low,” “Medium,” and “High.”
emotionName	Specifies the name of an Emotion.
emotionSetName	Specifies the name of the Emotion set which contains the Emotion. The Basic Emotion Set is used as a baseline, but other sets are possible.

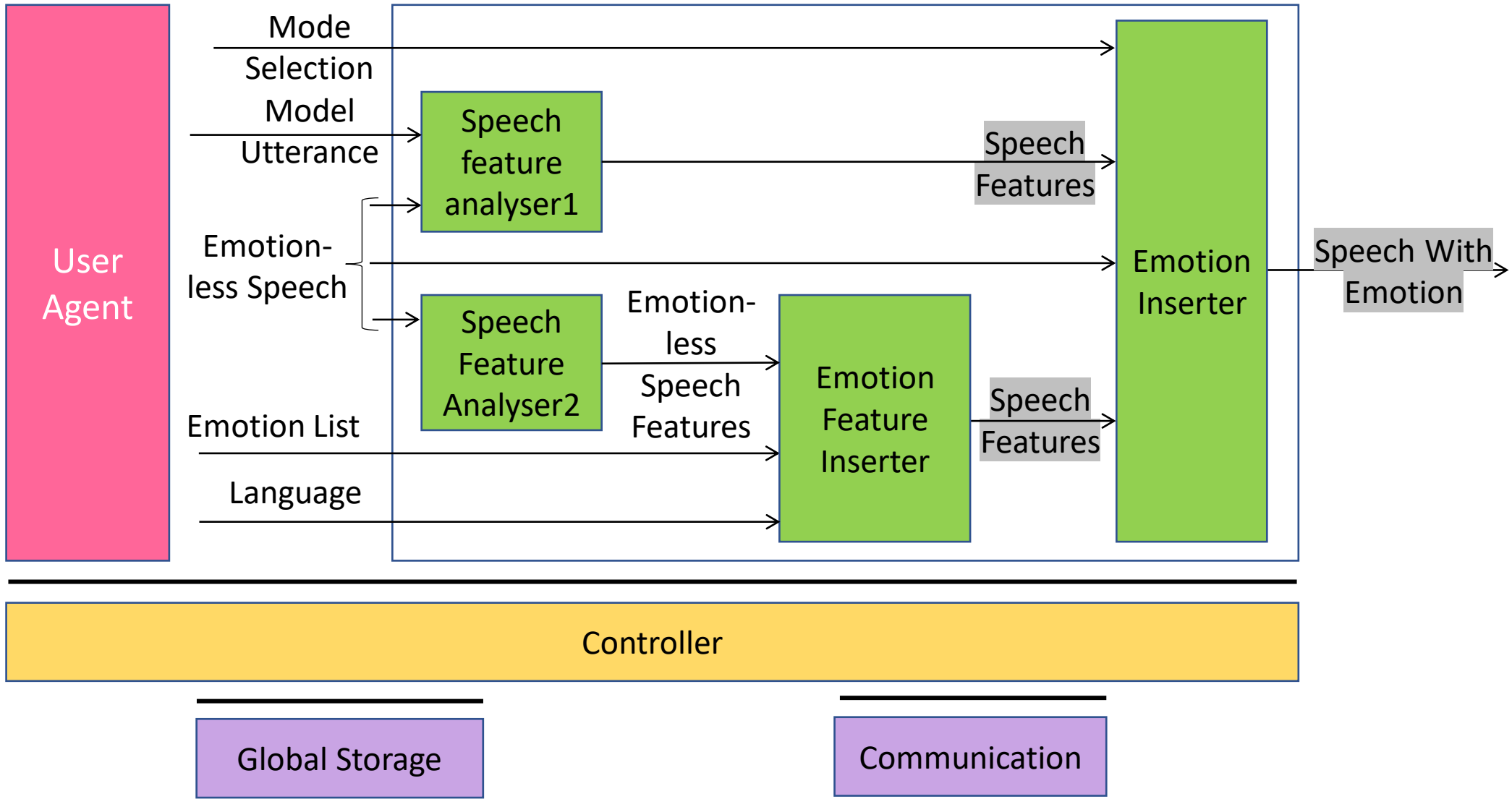
Emotion – Basic Set (the first entries)

EMOTION CATEGORIES	GENERAL ADJECTIVAL	SPECIFIC ADJECTIVAL
HAPPINESS	happy	joyful content delighted amused
SADNESS	sad	lonely grief-stricken discouraged depressed disappointed
CALMNESS	calm	peaceful/serene resigned
FEAR	fearful/scared	terrified anxious/uneasy
ANGER	anger	furious irritated frustrated
DISGUST	disgust	loathing
SOCIAL DOMINANCE, CONFIDENCE	arrogant confident submissive	
PRIDE/SHAME	proud ashamed	arrogant guilty/remorseful/sorry embarrassed

Context-based Audio Enhancement (MPAI-CAE)

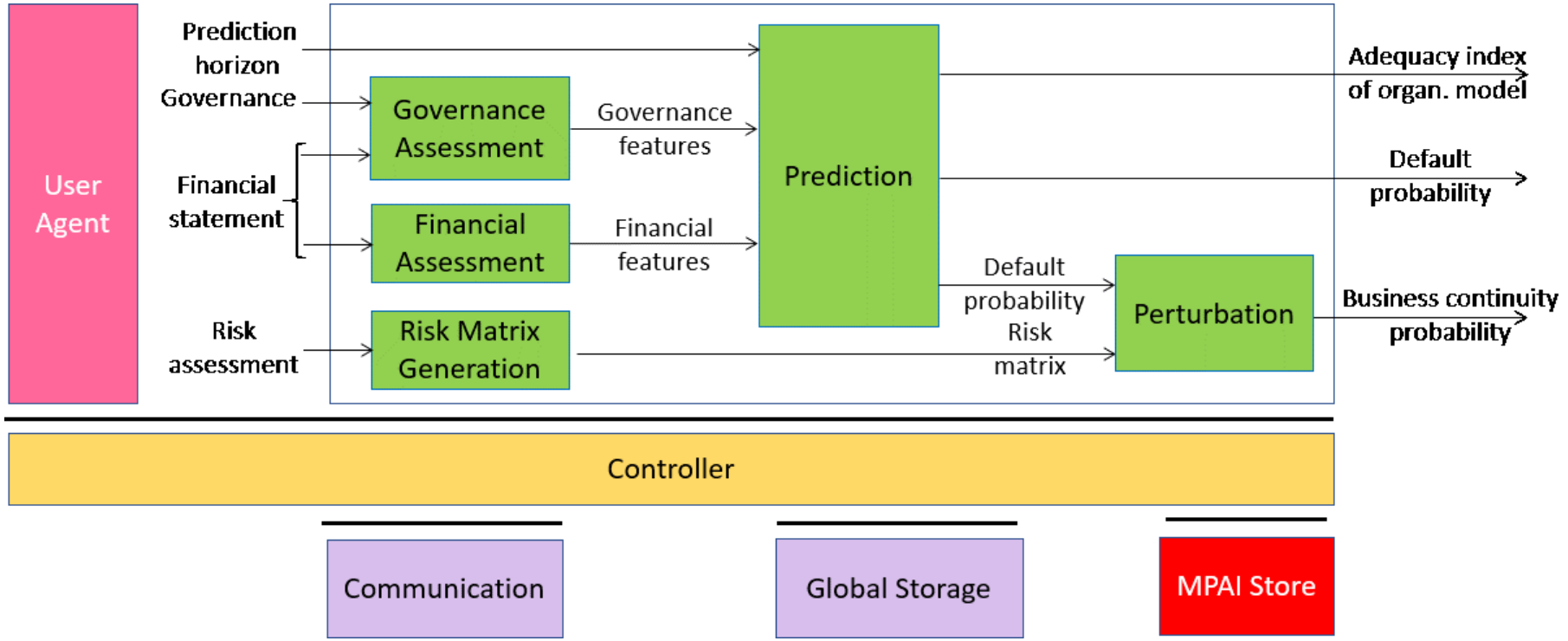
- Common characteristics
  - Improving the user experience for audio-related applications including: entertainment, communication, teleconferencing, gaming, post-production, restoration etc.
  - in a variety of contexts such as in the home, in the car, on-the-go, in the studio etc.
  - using context information to act on the input audio content, and potentially deliver the processed output via an appropriate protocol.
- MPAI-CAE Use Cases
  - Emotion Enhanced Speech (EES): adding emotion to emotion-less speech
  - Audio Recording Preservation (ARP): restoration of old audio tapes
  - Speech Restoration System (SSR): restoration of damaged speech recordings
  - Enhanced Audioconference Experience (EAE): improve audioconference.

Emotion-Enhanced Speech



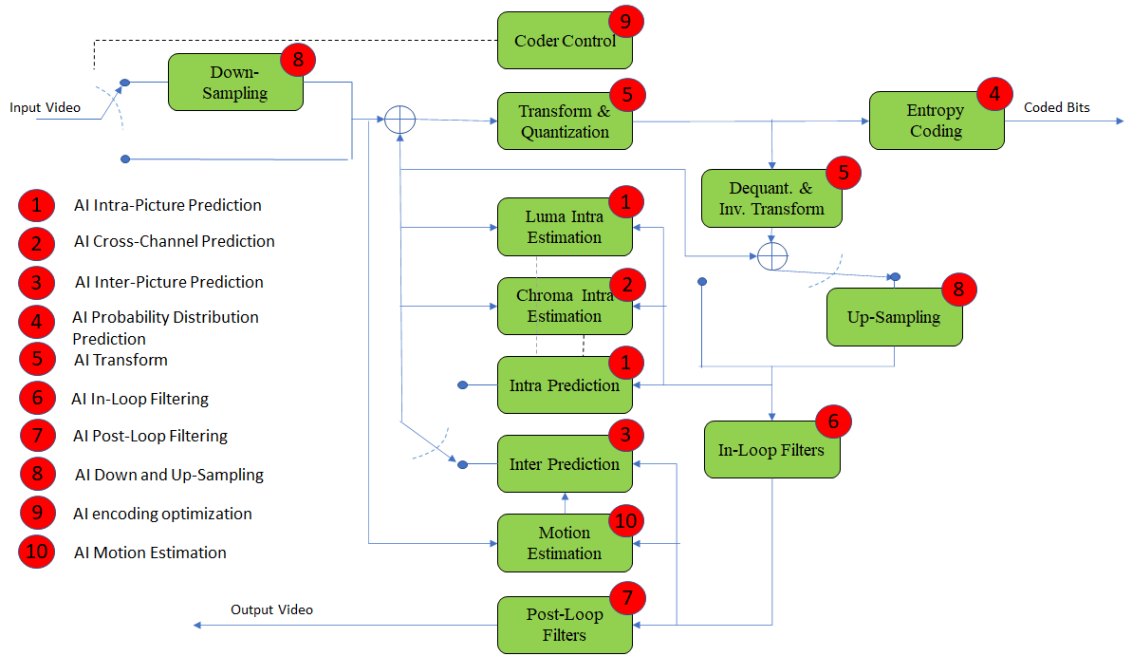
MPAI-CUI: Compression and understanding of industrial data

- MPAI-CUI aims to enable AI-based filtering and extraction of key information to predict company performance by applying Artificial Intelligence to governance, financial and risk data.
- MPAI-CUI requires standardisation of all data formats to be fed into an AI machine to extract information that is relevant to the intended use.
- Converted data undergo a further conversion and are then fed to specific neural networks.



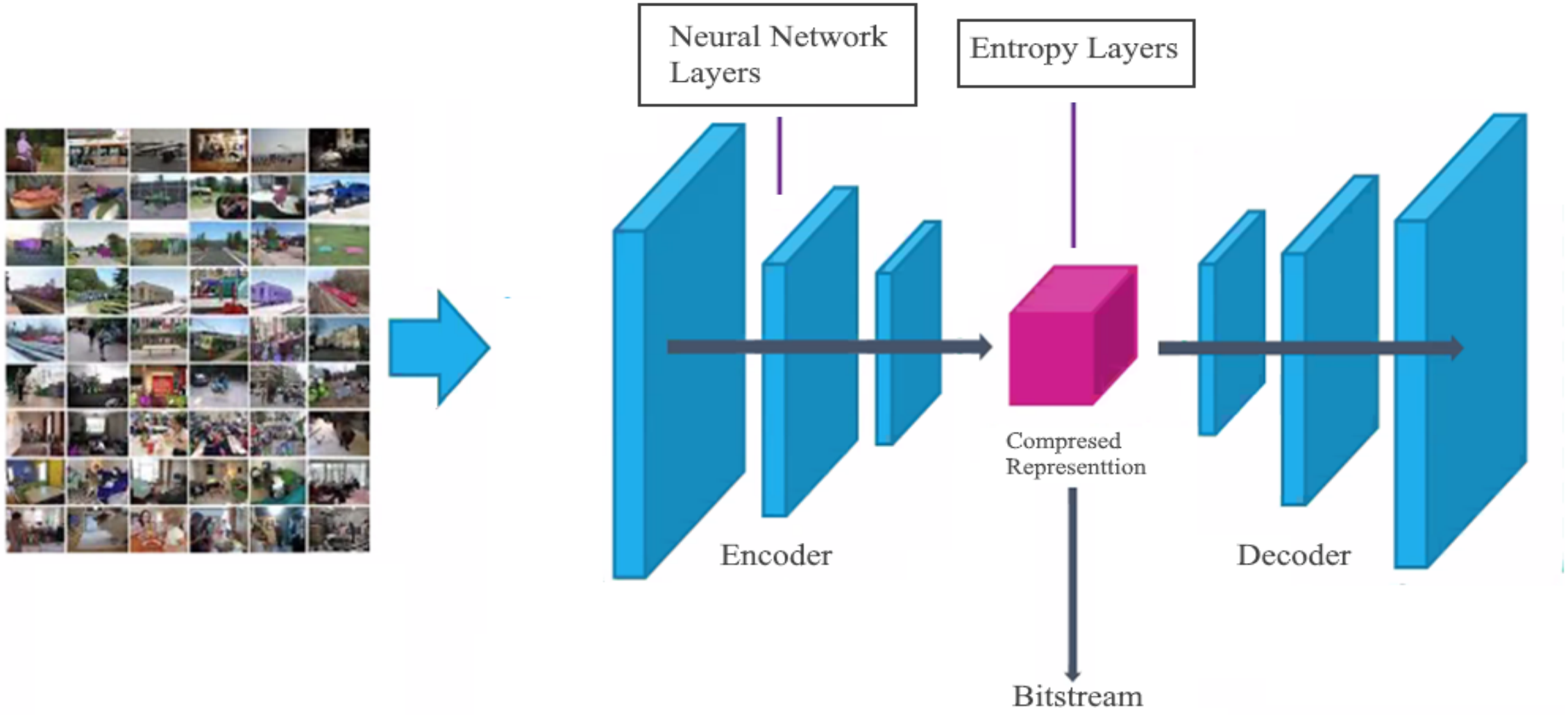
MPAI-EVC: AI-Enhanced Video Coding

- Uses AI to substantially enhance the performance of a traditional video codec by improving or replacing traditional tools with AI-based tools.



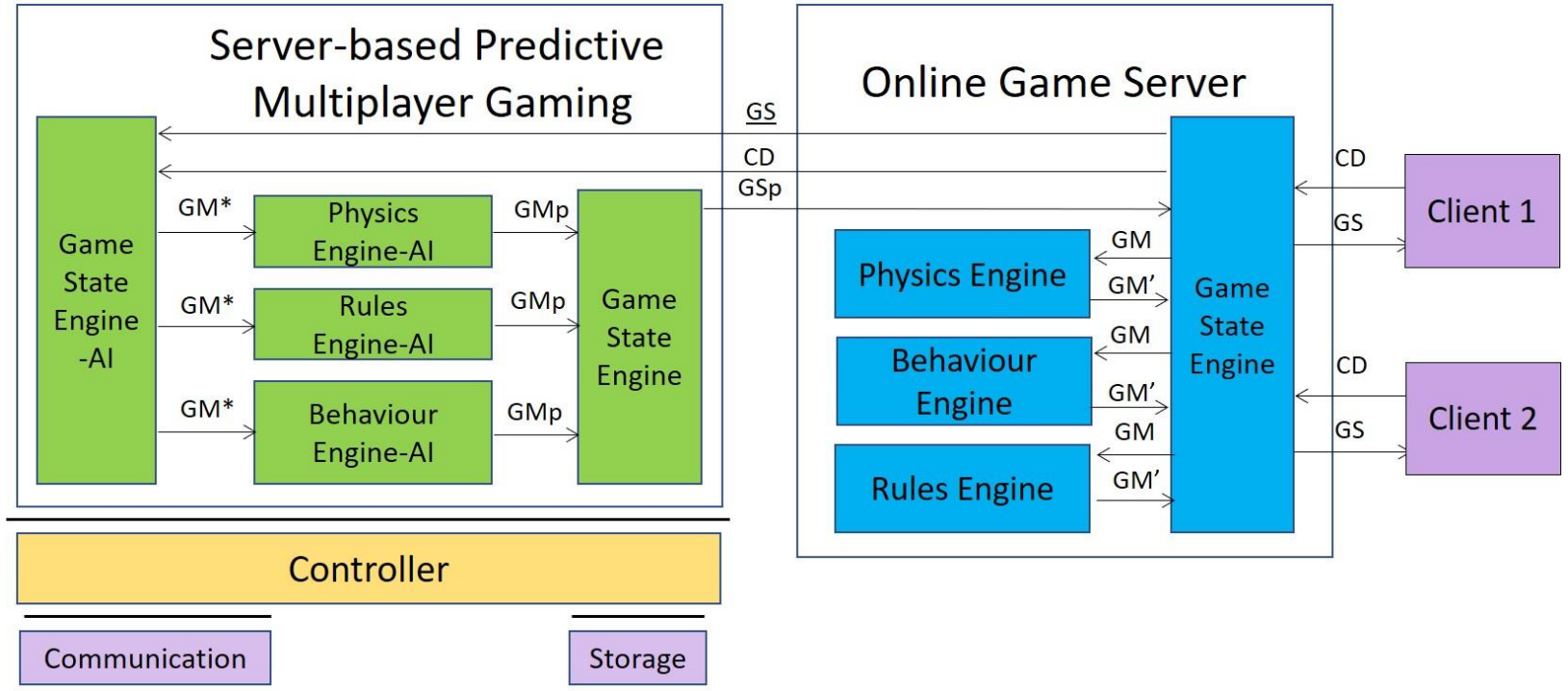
A reference diagram for the Horizontal Hybrid approach

A conceptual MPAI-EEV Reference Model



MPAI-SPG: Server-based Predictive Multiplayer Gaming

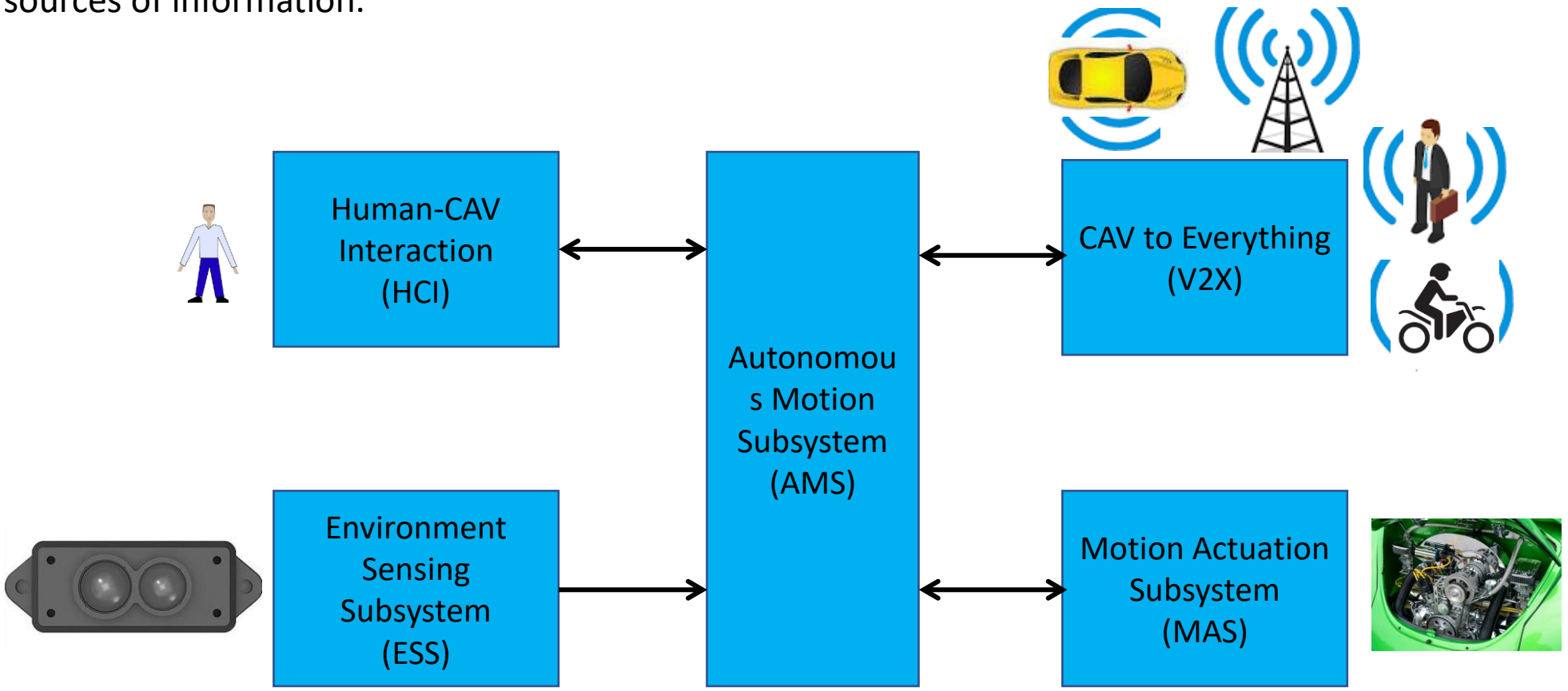
- aims to minimise the audio-visual and gameplay discontinuities caused by high latency or packet losses during an online real-time game.
- In case information from a client is missing, the data collected from the clients are fed to an AI-based system that predicts the moves of the client whose data are missing.
- provides a response to the need to detect who amongst the players is cheating.



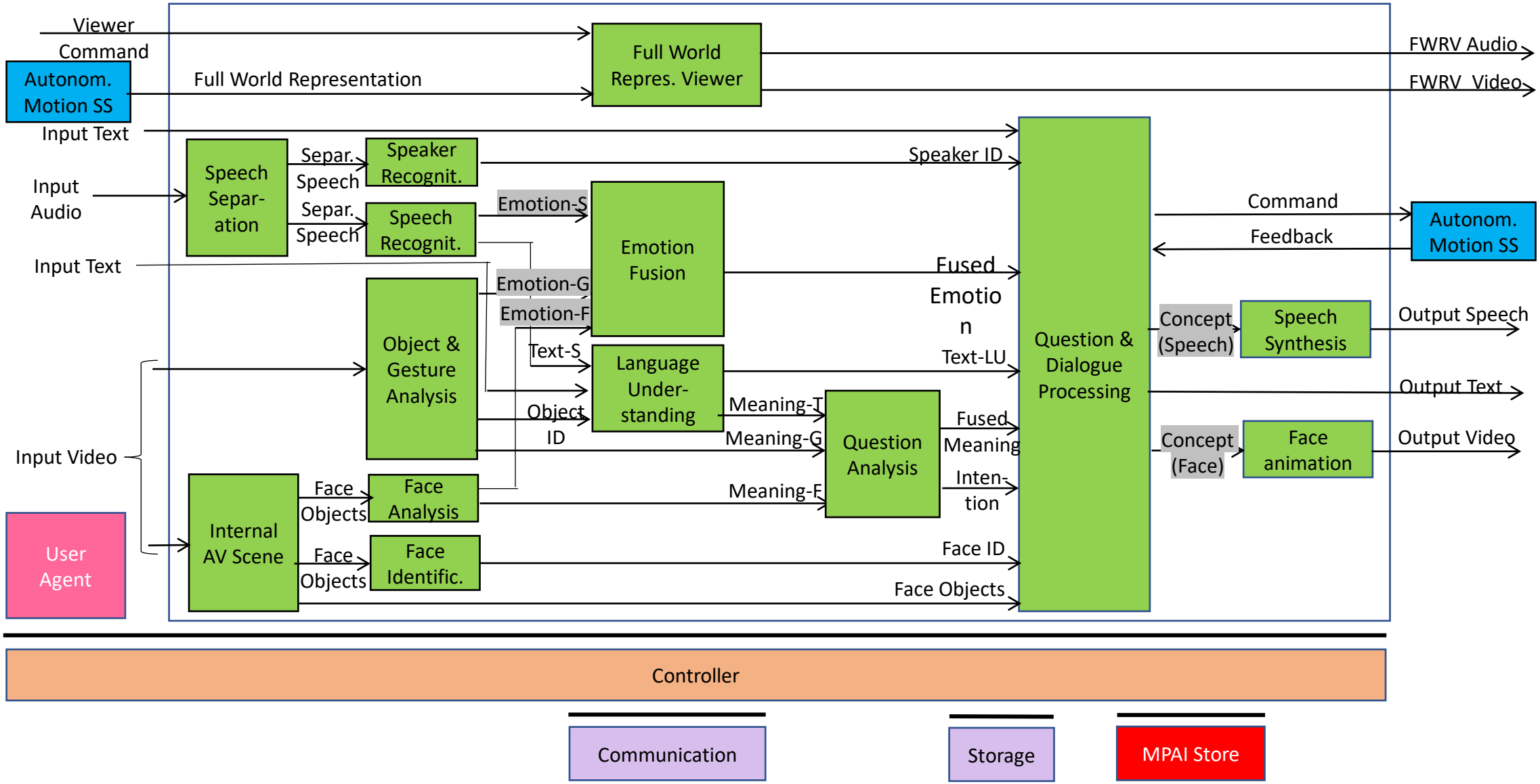
MPAI-CAV: Connected Autonomous Vehicles

- addressing the Connected Autonomous Vehicle (CAV) domain and the 3 main operating instances of a CAV:

1. Autonomous Motion, i.e., the operation of the portion of a CAV that enables its autonomous motion
2. Human-CAV interaction, i.e., the operation of the portion of a CAV that responds to humans' commands and queries and senses humans' activities
3. The CAV-environment interaction, i.e., the operation of the portion of a CAV that communicates with other CAVs and sources of information.



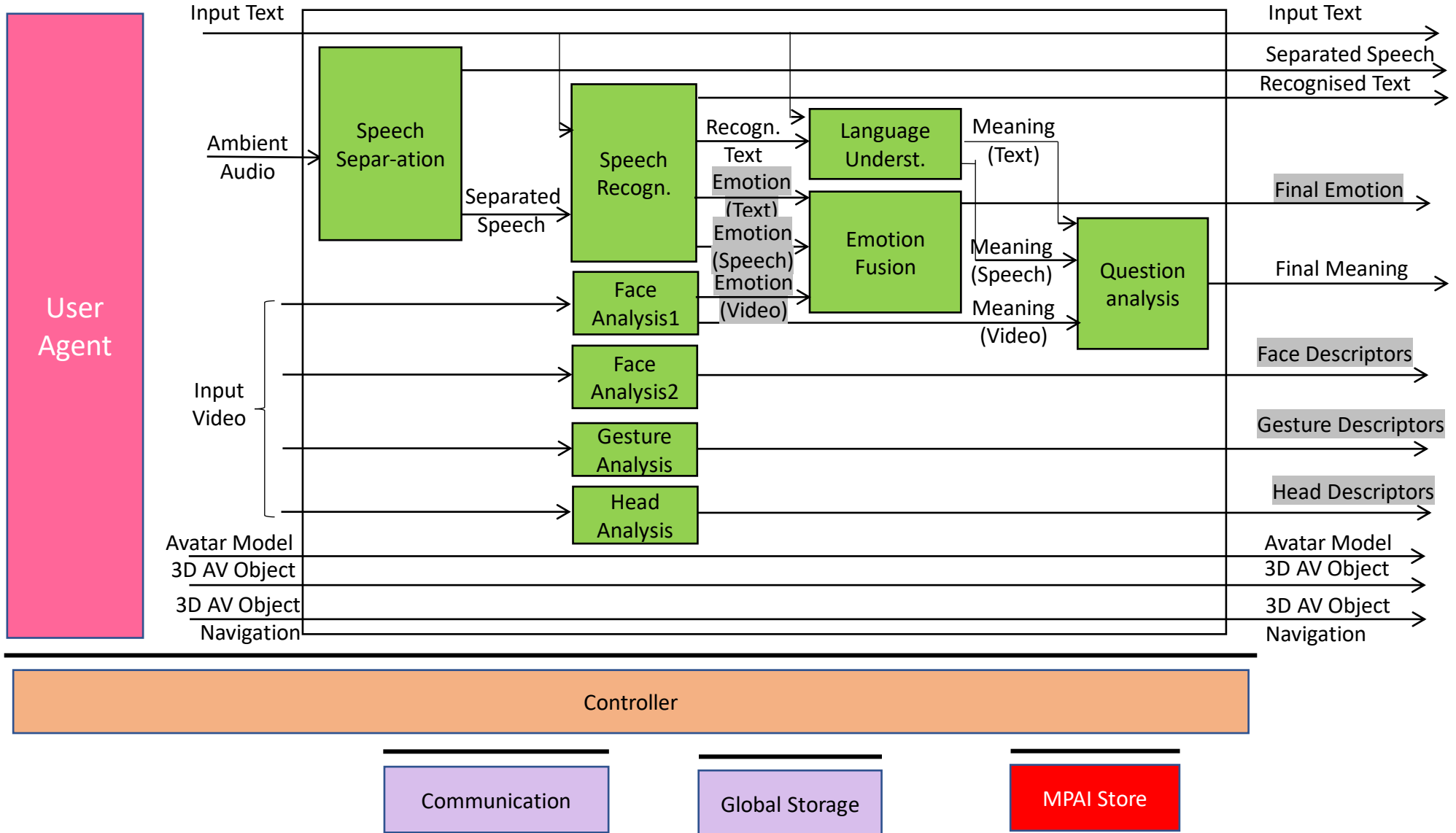
Human-Connected Autonomous Vehicle Interaction



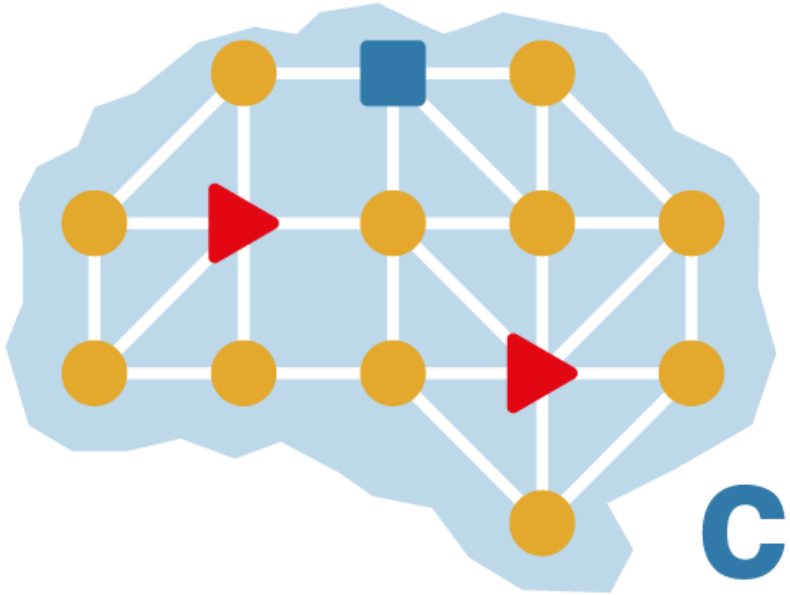
MPAI-MCS Use Cases and Functional Requirements

1. Four Use Cases
    1. Use Case #1 – Virtual e-learning
    2. Use Case #2 – Local Avatar Videoconference (LAV)
    3. Use Case #3 – Teleconsulting
    4. Use Case #4 – Multipoint videoconference
2. Functional requirements of UC #1 and UC#2 fully developed
3. Needs review by CAE-DC and MMC-DC
    1. Several technologies can be inherited from MPAI-CAE/MMC

MPAI-MCS: Virtual e-Learning (Metaverse)



Join the fun, build the future



MPAI.
community

www.mpai.community