MPAI talks to industry

Introduction to MPAI

Leonardo Chiariglione

Technology evolves through ages

- Technology has progressed through ages:
  - The mechanical age.
  - The electrical age.
  - The electronic age.
  - The digital processing age.
- Today we are in **The Artificial Intelligence (AI) age.**
- Data coding standards are important because they enable interoperability
  - MPEG used to be the standards organisation for data processing-based digital media with a business model of

    "**make good standards by rewarding patents worth inclusion**"
  - That business model is now **worn out**.

This is the first of a set of presentations about **the standards organisation** designed to cope with the needs for **data coding standards** in the **AI age.**

MPAI. community

Moving Picture, Audio and Data Coding by Artificial intelligence

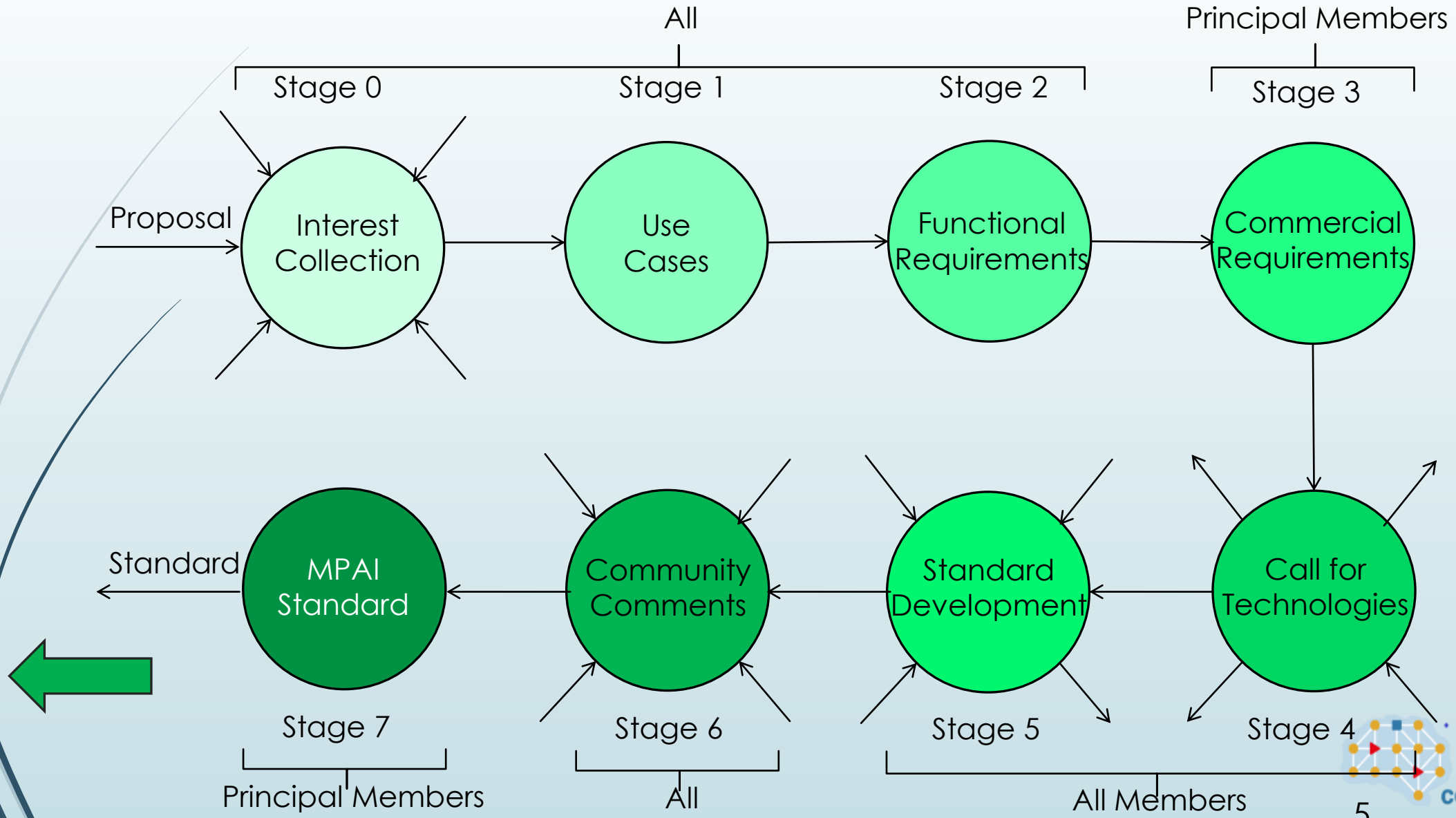
International, unaffiliated, not-for-profit organisation developing **AI-centred data coding standards**

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

MPAI is based on 4 pillars

#	Pillar
1	Develop standards based on a rigorous process
2	Set IPR Guidelines before developing a standard
3	AI modules (AIM) aggregated in AI Workflows (AIW) executed in a standard AI Framework (AIF)
4	Governance of the MPAI ecosystem

Rigorous standards development process



Goal: make standards accessible & timely available

Before initiating a standard, *Active Principal Members* **develop & adopt** the Framework Licence (FWL), a licence without *values: $, %, dates etc.* declaring that the eventual licence(s) will be issued

1. Not after products are on the market.
2. At a price comparable with similar standard technologies.

During the development, *any Member* making a contribution **declares** it will make its licence available according to the FWL.

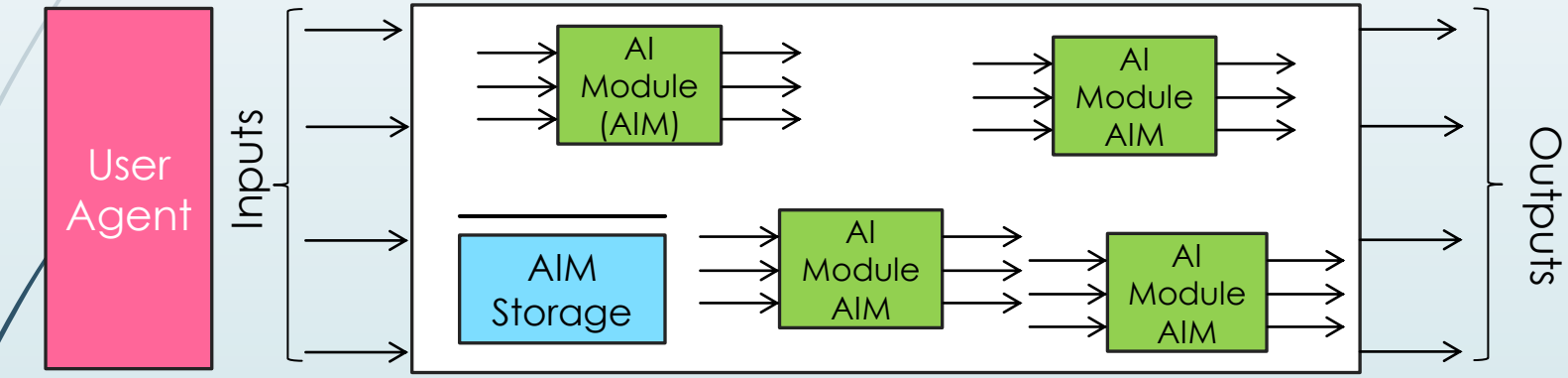
After the development, *Members holding IP* in the standard **select** the preferred patent pool administrator.

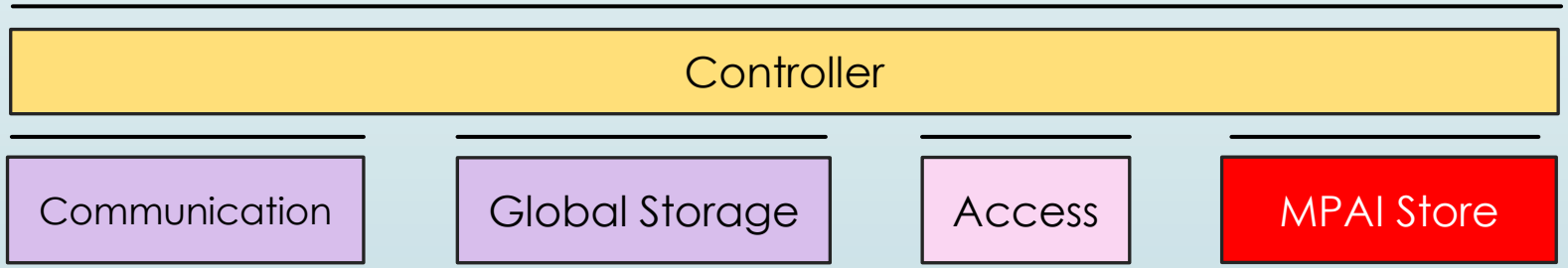
MPAI-AIF: AI Framework

MPAI develops component-based standards

AI module (AIM)

AI Workflow (AIW)

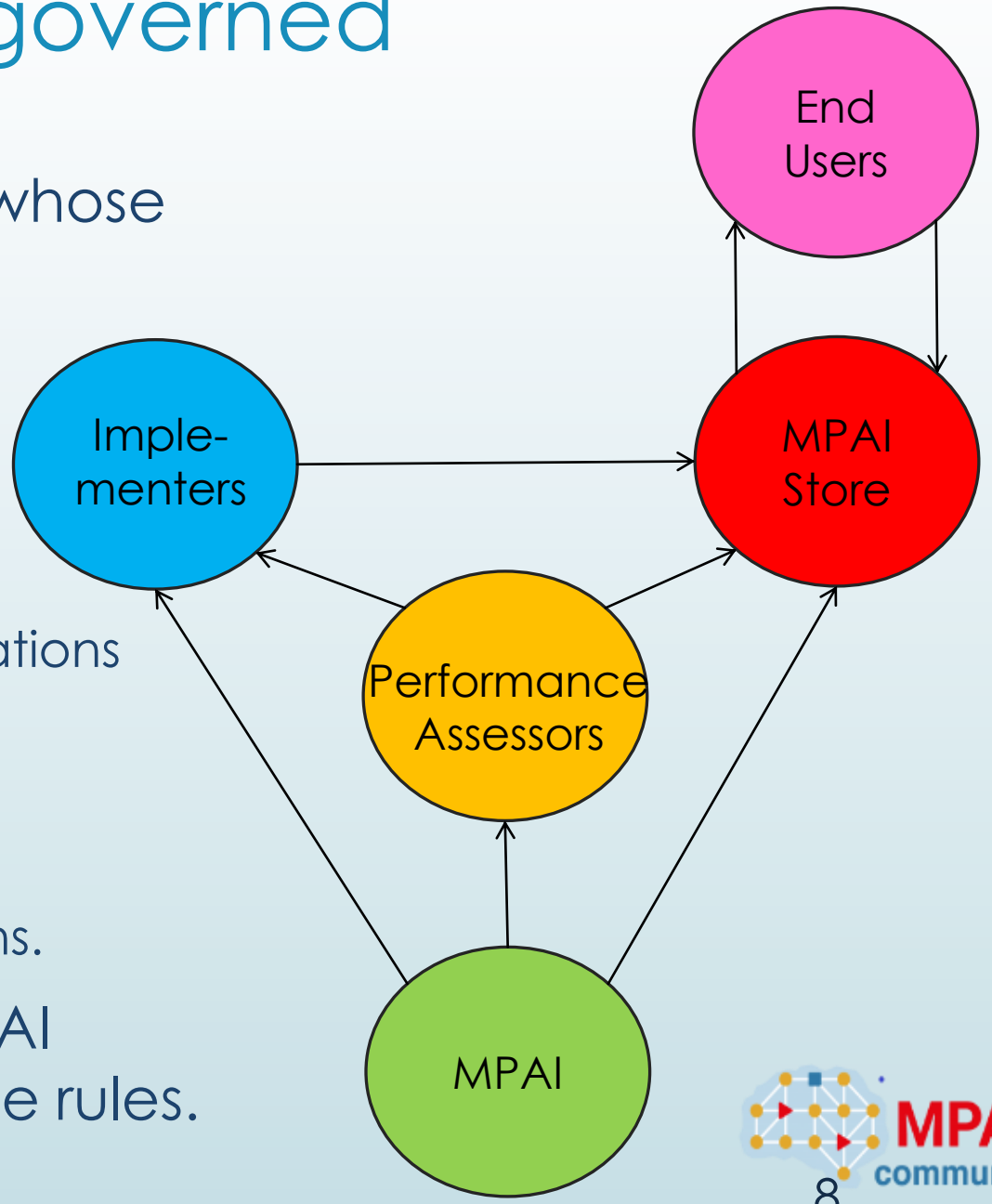




AI Framework (AIF)

The MPAI ecosystem is governed

- MPAI gives rise to an ecosystem whose components are:
  - **MPAI** produces MPAI **standards**.
  - **Implementers** make components and applications
  - **MPAI Store** verifies security and conformance and label implementations for interoperability and reliability.
  - **Performance Assessors** grade reliability of implementations.
  - **End Users** download implementations.
- MPAI is the root of trust of the MPAI Ecosystem issuing the governance rules.



5 MPAI standards developed in 15 months

1. AI Framework
2. Company Performance Prediction
3. Context-based Audio Enhancement
4. Multimodal Conversation
5. Governance of the MPAI Ecosystem

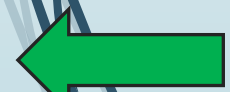
AI Framework

Component-based standard

Enabling execution of **AI Workflows**

Composed of **AI Modules**

Both defined in terms of their **functions** and **interfaces**



Company Performance Prediction

A standard to

- Predict a company's
  - Default probability, and
  - Business discontinuity probability

  in a given time frame measured in years, and
- Assess governance adequacy.

Context-based Audio Enhancement (MPAI-CAE)

- A standard enabling the improvement of the user experience for audio-related applications in
  - 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
- Using context information

Multimodal Conversation (MPAI-MMC)

To enable forms of human-machine conversation that emulate human-human conversation in completeness and intensity.

- Conversation with Emotion supporting audio-visual conversation with a machine impersonated by a synthetic voice and an animated face;
- Multimodal Question Answering supports request for information about a displayed object;
- Automatic Speech Translation – Unidirectional, Bidirectional and One-to-Many – support conversational translation using a synthetic voice preserving the speech features of the human.

Governance of the MPAI Ecosystem

Guarantees end users that implementations are

- Secure.
- Conform to an MPAI application standard (e.g., MPAI-CAE).
- Are reliable, e.g., unbiased.

MPAI is now developing 9 standards

1. AI-Enhanced Video Coding
2. End-to-End Video Coding
3. Online Gaming
4. Connected Autonomous Vehicles
5. Mixed-reality Collaborative Spaces
6. Genomics
7. Visual Scene and Objects
8. Neural Network Watermarking
9. Avatar Representation and Animation

AI-based Video Coding

- AI-Enhanced Video Coding (MPAI-EVC)
  - A video compression standard that substantially enhances the performance of a traditional video codec (MPEG-5 EVC in this case) by improving or replacing traditional tools with AI-based tools.
- AI-Based End-to-End Video Coding (MPAI-EEV)
  - A video compression standard that seeks to compress video by exploiting AI-based data coding technologies. No legacy assumed – as in MPAI-EVC – on traditional use of data processing technologies.

Online Gaming (MPAI-SPG)

- Aims to minimise the audio-visual and gameplay discontinuities caused by high latency or packet losses during an online real-time game.
- If information from a client is missing, client data are fed to an AI-based system that predicts the missing data.
- Data from a cheating player can also be detected.

Connected Autonomous Vehicles (MPAI-CAV)

- A Connected Autonomous Vehicle (CAV) is a system
  - capable of moving autonomously
  - based on the analysis of the data produced by sensors exploring the environment and the information transmitted by other sources in range, e.g., other CAVs.
- MPAI-CAV aim at standardising the IT components of a CAV.

Mixed-reality Collaborative Spaces (MPAI-MCS)

- A project for scenarios where humans represented by avatars collaborate in virtual-reality spaces.
- Virtual twins of humans – avatars – are represented with as realistic as possible features, especially: face, head, arms, hands and speech.

Neural Network Watermarking (MPAI-NNW)

A project developing requirements for a standard enabling the measure, for a given size of the watermarking payload, of

1. The impact, e.g., the degradation of the user experience caused by the watermark applied to a neural network.
2. The resistance to attacks, e.g., transfer learning, pruning.
3. The processing cost of watermarking injection, e.g., time, processing cost.

Program

#	Title	Speaker	Country
1	Introduction to MPAI	Leonardo	CH
2	MPAI-AIF – AI Framework	Andrea	IT
3	MPAI-CAE – Context-based Audio Enhancement	Marina	US
4	MPAI-MMC – Multimodal Conversation	Miran	KR
5	MPAI-CUI – Compression and Understanding if Industrial Data	Guido	IT
6	Reference Software, Conformance and Performance	Panos	UK
7	MPAI-GME – Governance of the MPAI Ecosystem	Paolo	UK
8	MPAI-SPG – Server-based Predictive Multiplayer Gaming	Marco	IT
9	MPAI-EVC – AI-Enhanced Video Coding	Roberto	IT
10	MPAI-EEV – AI-based End-to-End Video Coding	Chuanmin	CN
11	MPAI-CAV – Connected Autonomous Vehicles	Gianluca	IT
12	MPAI-MCS – Mixed-reality Collaborative Spaces	David	US
13	Conclusions	Leonardo	CH

The MPAI AI Framework Standard (MPAI-AIF)

Andrea Basso

Why an AI Framework standard?

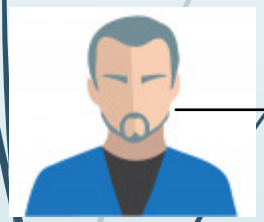
- **A standard AI Framework** can
  - ✓ Create, compose, execute, and update component-based workflows.
  - ✓ Interconnect components trained to specific tasks from multiple vendors.
  - ✓ Implement possibly high-complexity AI Workflows
  - ✓ Execute AI Workflows exchanging standard format data between AIMs.
- **Benefits various actors**
  - ✓ **Technology providers** can offer conforming AI components to the market
  - ✓ **Application developers** find components for their applications on the market
  - ✓ **Innovation** fueled by demand for novel/more performing AI components
  - ✓ **Consumers** have a wider choice of AI applications in a competitive market
  - ✓ **Society** can lift opaque veils from large, monolithic AI-based applications.

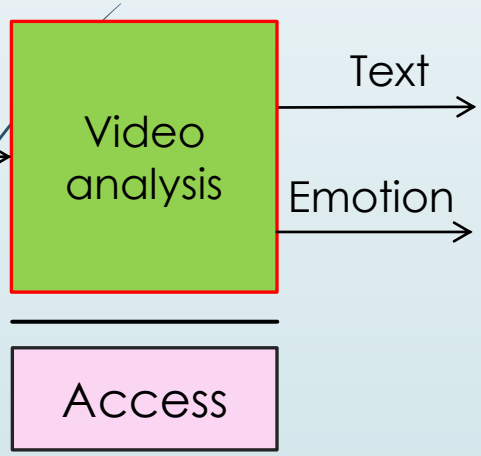
What the MPAI-AIF standard does/1

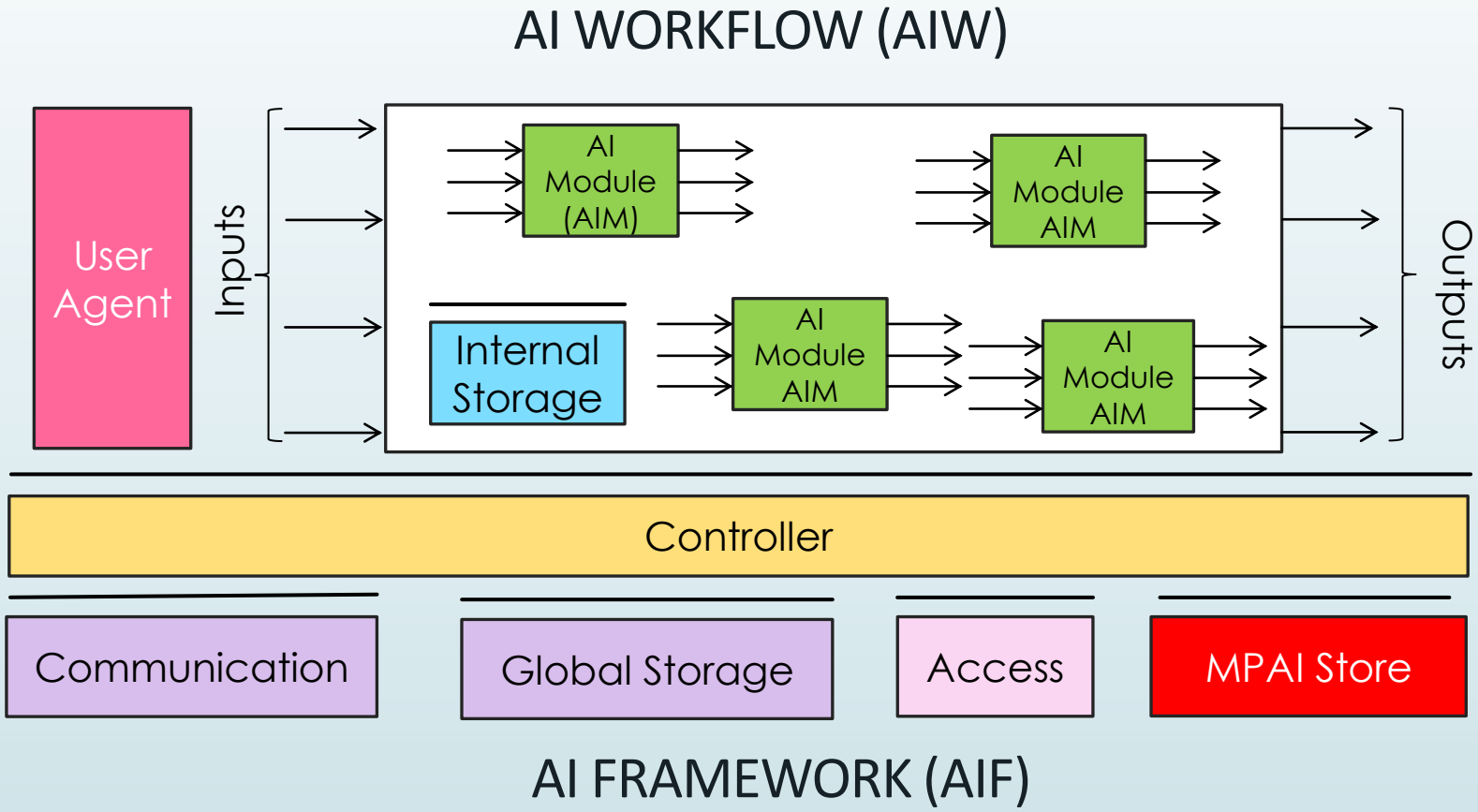
- Specifies
  - Architecture
  - Interfaces
  - Protocols
  - APIs

  of a Framework executing AI-based applications.
- Has the following main features:
  - ✓ Is component-based.
  - ✓ Defines the interfaces among its components.
  - ✓ Is secure as the components operate in a trusted zone.
  - ✓ Supports mixed hardware-software implementations.
  - ✓ Supports distributed and local execution environments.
  - ✓ Supports Machine Learning.
  - ✓ Supports operation of AIFs in proximity

The MPAI-AIF standard/2







AIF Components: Controller

- Provides functionalities such as scheduling, inter AIMs communication, access to all the AIF components.
- Activates/suspends/resumes/deactivates AIWs or AIMs according to the user's or other inputs.
- Load balancing.
- Exposes three APIs:
  - *AIM API* through which modules can communicate (register themselves, communicate and access the rest of the AIF environment)
  - *User API* through which the user or other Controllers can perform high-level tasks (e.g., switch the Controller on and off, give inputs to the AIW through the Controller).
  - *MPAI Store API* to enable communication between the AIF and the Store.
- May run one or more AIWs.

AIF: Other Components

AI Module (AIM): data processing element receiving AIM-specific Inputs and producing AIM-specific Outputs according to its Function. An AIM may aggregate AIMs.

AI Workflow (AIW): organised aggregation of AIMs implementing a Use Case receiving AIM-specific Inputs and producing AIM-specific Outputs according to its Function.

Access: provides access to static or slowly changing data that are required by an application such as domain knowledge data, data models, etc.

Communication: connects the Components of an AIF.

Global Storage: stores data shared by AIMs.

Internal Storage: stores data of the individual AIMs.

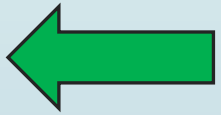
MPAI Store: makes available Implementations for users to download.

User Agent: The Component interfacing the user with an AIF

AI Framework: Features

- Event-based, and port and channel-based (unicast)
- Two types of Messages: High-Priority and Normal-Priority Messages
- Messages may be communicated through Channels or Events.
- Controller may run on a different computing platform than the AIW.
- AIMs may run on different computing platforms, e.g., on the cloud or in coordinated drone swarms.
- API Profiles allow Implementation on different computing platforms and using different programming languages.
- The Controller is always present even if Implementation is lightweight.
- AIMs may be hot-pluggable and register themselves on the fly.
- Swarms of devices with multiple controllers supported.

http://aif.mpai.community/



The Compression and Understanding of Industrial Data standard (MPAI-CUI)

Guido Perboli – Politecnico di Torino

MPAI-CUI: Contents

- The MPAI-CUI standard
- Applications of the MPAI-CUI standard
- MPAI-CUI Demo (anonymous companies)

1. The standard

The «AI-based Compression and Understanding of Industrial Data» standard

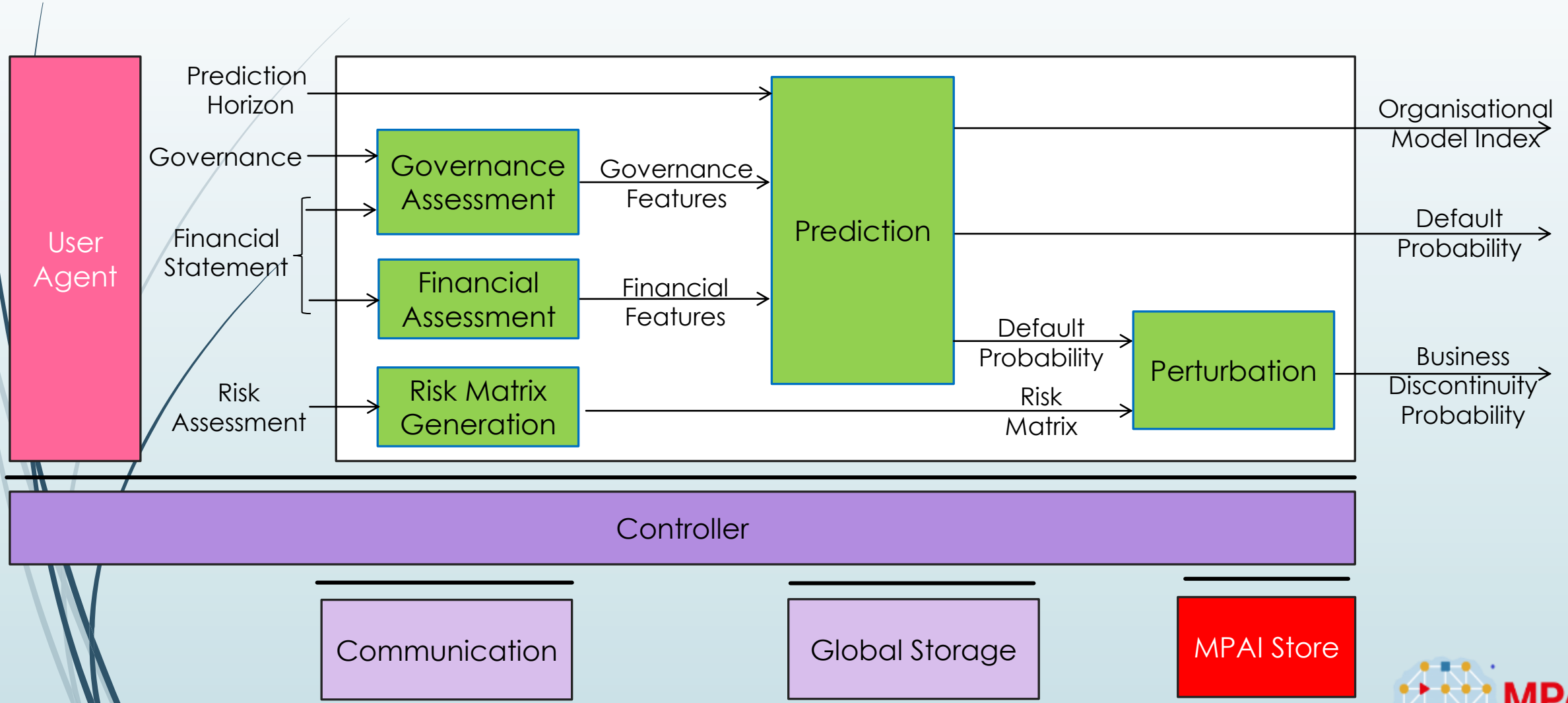
Company Performance Prediction use case

- The Company Performance Prediction of MPAI-CUI is a powerful and extensible way to predict the performance of a company
  - Financial risks
  - Vertical risks (seismic and cyber)
- Predicts the performance of a company from its governance, financial and risk data in a given time horizon of prediction.
- First full standard released by MPAI

What does "performance" mean?

- **Default probability**: the probability the company will default (e.g., crisis, bankruptcy) in a specified number of future months dependent on financial features
- **Organisational Model Index**: the adequacy of the organisational model (e.g., board of directors, shareholders, familiarity, conflicts of interest)
- **Business continuity Index**: the probability of an interruption of the operations of the company for a period of time less than 2% of the prediction horizon.

MPAI-CUI Workflow



MPAI-CUI Workflow

1. Input:
    1. Prediction Horizon
    2. Governance Data
    3. Financial Data
    4. Risk Assessment Data.
2. Processing

AI Module	Input data	Output data
Governance Assessment	Governance and Financial	Governance Features
Financial Assessment	Financial Statement	Financial Features
Risk Matrix Generation	Risk Assessment	Risk Matrix
Prediction	Governance Features, Financial Features	Organizational Model Index, Default Probability
Perturbation	Default Probability, Risk Matrix	Business Discontinuity Probability

3. Output
    1. Organizational Model Index
    2. Default Probability
    3. Business Discontinuity Probability

MPAI-CUI – An AI-based standard

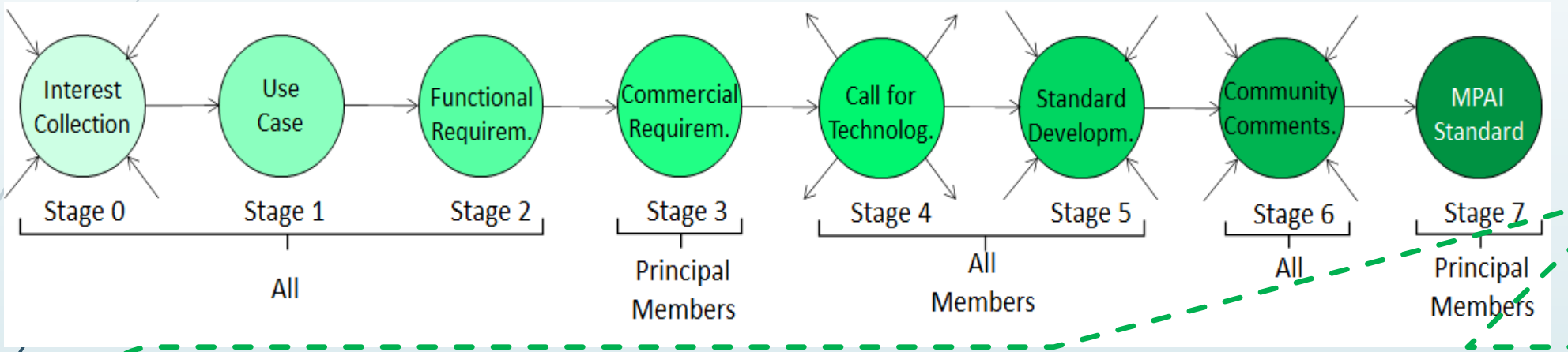
- Prediction AIM is a neural network.
  - Back testing on a sample of 160.000 companies both active and bankrupted
  - Prediction Accuracy: 85% vs 37% of traditional techniques
  - Reviewed by the scientific community

  - *See further details in G. Perboli and E. Arabnezhad. A Machine Learning-based DSS for mid and long-term company crisis prediction. Expert Systems with Applications, 174, 114758, 2021*

AI-based standard

- Novelties of MPAI-CUI
  - Extracts the most relevant data with **controlled information loss by** analysing the **large amount** of data required by regulation.
  - Extends prediction horizon **up to 60 months**, using AI.

Standard development process



What next?

Future versions with more use cases comprising other vertical risks (e.g., Environmental, Social) not included in the present version of the standard.

2. MPAI-CUI standard: Applications

How is MPAI-CUI going to be used?

COMPANY BOARDS

+ Develop efficient strategies.

+ Identify clues to crisis or bankruptcy years in advance.

+ Help to:
- Decide path to recovery,
- Conduct what-if analysis,
- Devise efficient strategies.

BANKS/FINANCIAL INSTITUTIONS

+ Assess the financial health of companies

+ Aids the financial institution to make the right decision in:
- funding or not funding that company,
- having a broad vision of its situation.

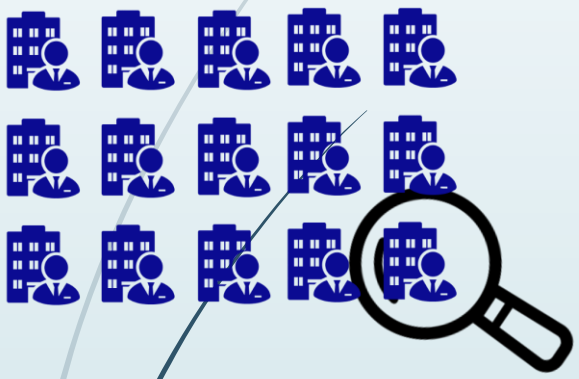
PUBLIC AUTHORITIES

+ Assess public policies in advance

+ Evaluate scenarios of public interventions

+ Identify proactive actions to increase resiliency of industrial sectors.

A real example: Evaluate the effects of a public policy



- Evaluate the impact of funds made available to SMEs in the Piedmont region by analysing the **probability of default** before and after public intervention.
- The study **confirmed the effectiveness of the public intervention** put in place by showing that public intervention improved the default probability.

See further details in G. Perboli et al. Using machine learning to assess public policies: a real case study for supporting SMEs development in Italy. TEMSCON 2021

3. MPAI-CUI – Demo

Prof. Guido Perboli
Politecnico di Torino – Arisk srl

What is the demo

- Illustrate a real application by applying an AI-based Company Performance Prediction (CPP) **decision support system**.

- **Compute instantaneously** the Default Probability and the Organizational Model Index after entering the company VAT number.

Demo - Home

Actions possible in the Action section:

View report, view the complete report of the selected company.

Fill out survey, view the parts to be filled in, (e.g., the input of risk data).

Delete, delete the selected company.

Change Company registry, change or update company data (e.g., Ateco Code).

Demo- Dashboard Menu

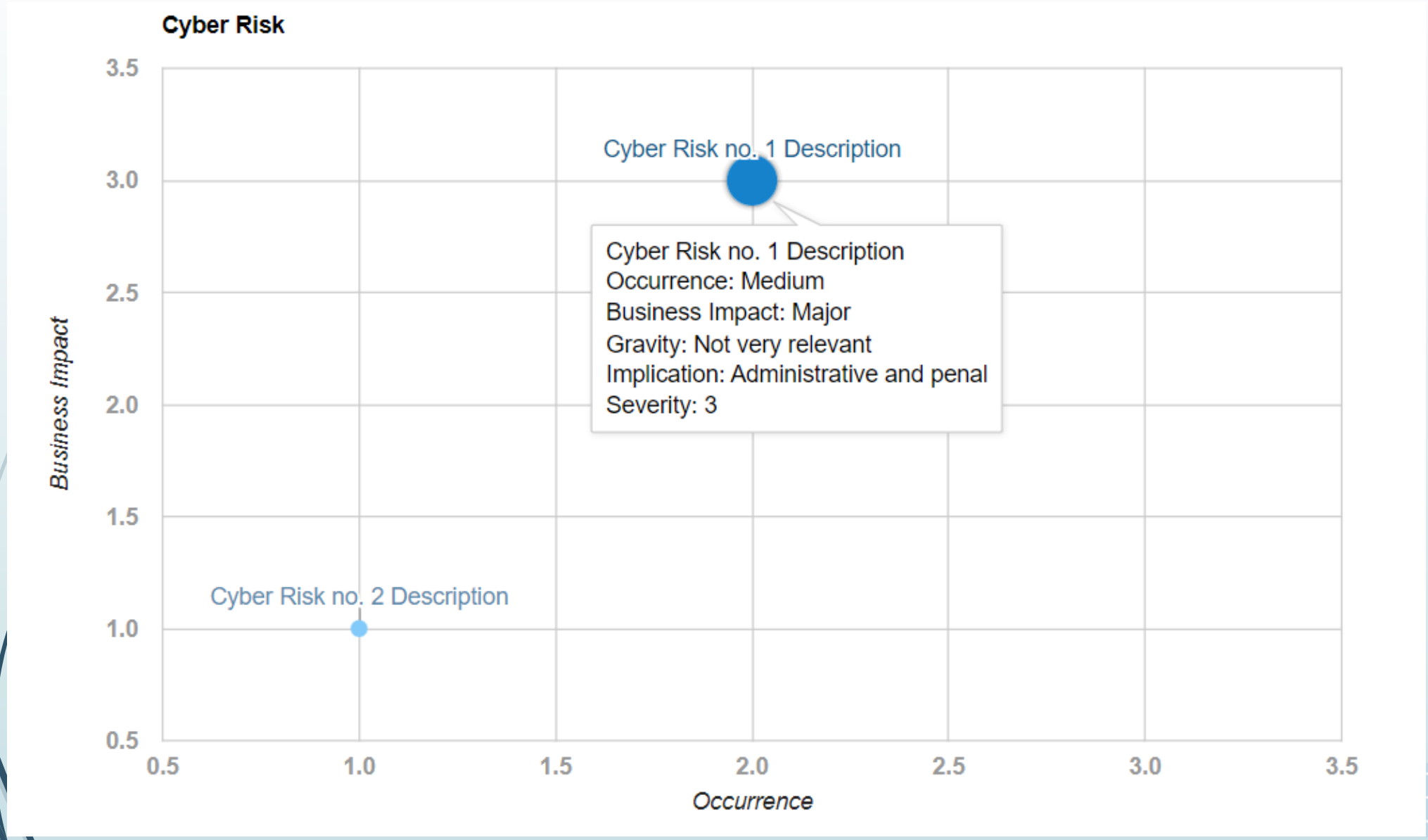
The report includes various information, the most important are:

Financial KPIs, in this section the main financial KPIs of the company are collected and presented in tabular form.

Risk Analysis, information section in which it is possible to see the risk analysis compiled in the survey section in graphic format.

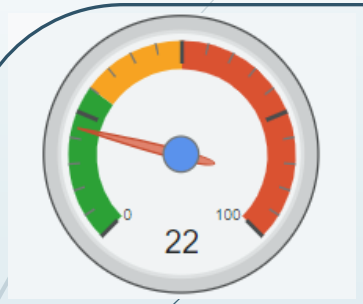
Red Flags, this section presents the instantly computed scores that satisfy a condition.

Demo- Risk Analysis - Example



Demo – Instant Score - Example

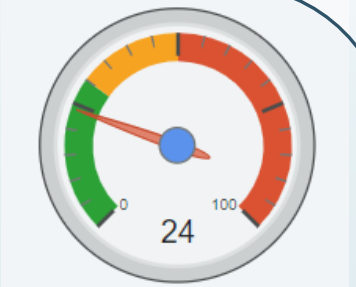
Red Flags: Last section but one of the most important of the report.



Default Probability (%)

22% chance of default in the next 36 months

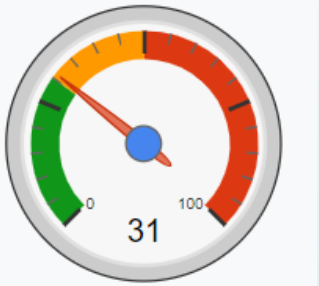
LOW RISK



Organisational Model Index (%)

24% the risk of inadequacy of the organizational model

GOOD STRUCTURE



Business Discontinuity (%)

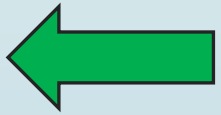
31% the risk of business discontinuity

GOOD RISK MANAGEMENT

INSTANT SCORE

- *The ML/AI used in this DEMO Software is indicative of the MPAI-CUI potential.*
- *This demo is a work by Arisk Srl*

http://cui.mpai.community/



2/25/2022

49

The Context-based Audio Enhancement standard (MPAI-CAE)

Marina Bosi

MPAI-CAE How to Best Experience Audio

- While we are surrounded by sounds and sounds can elicit a gamut of emotions - from imminent danger to the gentle whispers of a lullaby, often the audio experience can be inaccurate and distorted
  - Think for example of an audio conference where it is difficult to follow the thread because of a noisy environment and/or multiple speakers talking at once;
  - or trying to enjoy music while bicycling in town: we want to continue to be aware of any danger from the surroundings but still have the option of listening to high quality audio
  - or trying to preserve oral/ audio cultural heritage so that it can be enjoyed for generations to come
- MPAI-CAE addresses different scenarios where the improvement of the user experience is achieved by AI context information in a variety of situations
- Four specific use cases have been identified and currently standardized in MPAI CAE Version 1

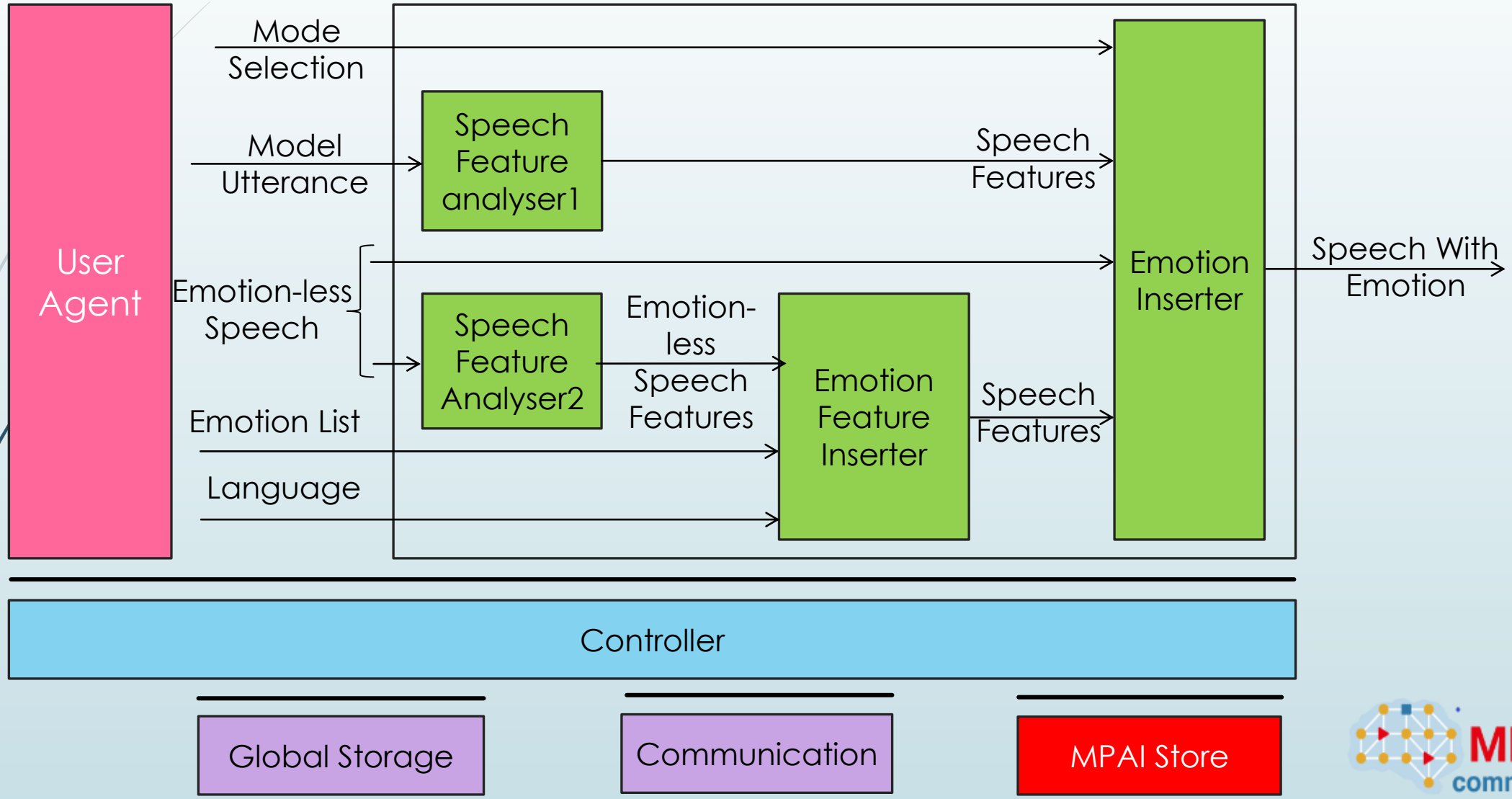
MPAI-CAE Version 1 Use Cases

1. **Emotion-Enhanced Speech (EES)**: add a specified emotion to an emotion-less speech.
2. **Audio Recording Preservation (ARP):** extract preservation information from video tape of the reading head.
3. **Speech Restoration System (SRS):** restore a damaged segment of a digital vocal track where replacements for the damaged vocal elements will be synthesized.
4. **Enhanced Audioconference Experience (EAE):** improve conference call experience.

Emotion-Enhanced Speech (CAE-EES): Overview

- Converts an **emotionless** speech segment to one **with emotion**.
- Both input and output speech segments are contained in files.
- The desired emotion is expressed
  - … as a **label** (e.g., "angry") belonging to a standard list of emotions
  - … or derived by **extracting speech features** from a model utterance.

Emotion-Enhanced Speech (CAE-EES): Workflow



CAE-EES Workflow: Explanation

Two Pathways for addition of **emotional charge** to neutral utterance (**Emotionless Speech**).

Pathway 1 (upper and middle left of workflow):

1. **Model Utterance** is input with **Emotionless Speech**
2. Features of **Model Utterance** can be captured and transferred to **Emotionless Speech**.

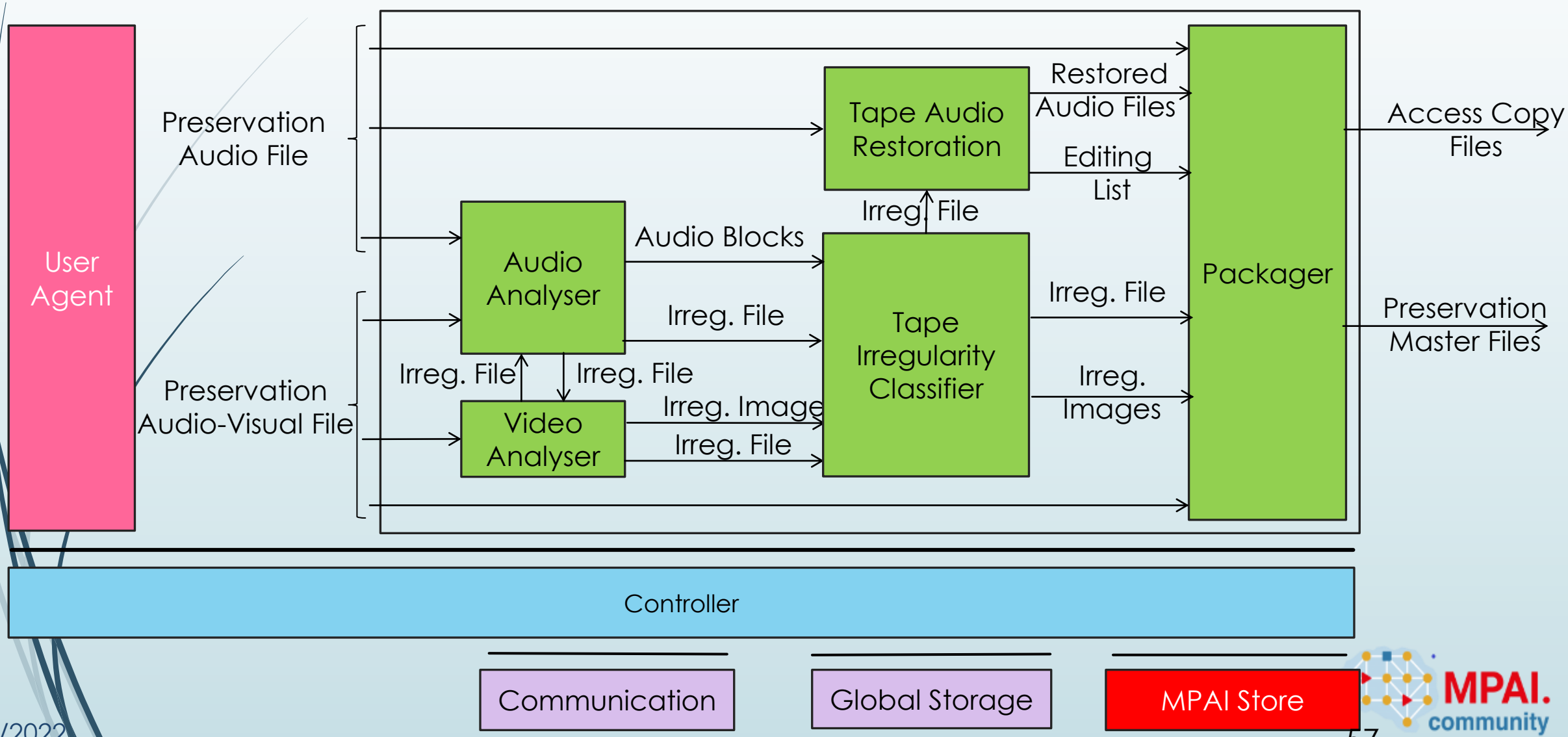
Pathway 2 (middle and lower left of workflow):

1. **Emotionless Speech** is input with **Emotion List** specifying desired emotions.
2. **Speech Feature Analyser2** extracts **Emotionless Speech Features** describing initial state
3. … and sends them to **Emotion Feature Inserter**
4. … which produces **Speech Features** that will yield emotional charge specified by **Emotion List**
5. … and sends them to **Emotion Inserter**
6. … which uses them to synthesize **Speech With Emotion**.

MPAI-CAE Audio Recording Preservation (ARP)

- Audio archives need to digitise their records, especially analogue magnetic tapes.
- Preservation requires important resources: people, time, funding.
- The magnetic tape carrier may hold important information: multiples splices and annotations; display several types of irregularities (e.g., corruptions of the carrier, tape of different colour or chemical composition).
- International guidelines (e.g., International Association of Sound and Audio-visuals Archives; World Digital Library; Europeana), but no international standards.
- AI can help cultural heritage by making cultural heritage preservation sustainable by drastically changing the way it is preserved and accessed for added value.

MPAI-CAE Audio Recording Preservation (ARP)

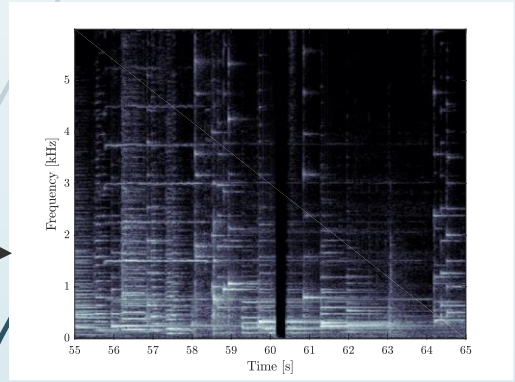


MPAI-CAE ARP

Preservation Audio & Audio Visual Files

Tape Audio Restoration

Access Copy Files



Audio Analyser

Packager

Tape Irregularity Classifier

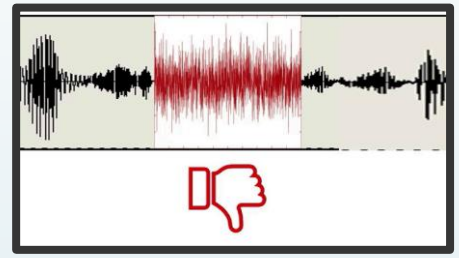
Video Analyser

Preservation Master Files

MPAI-CAE ARP

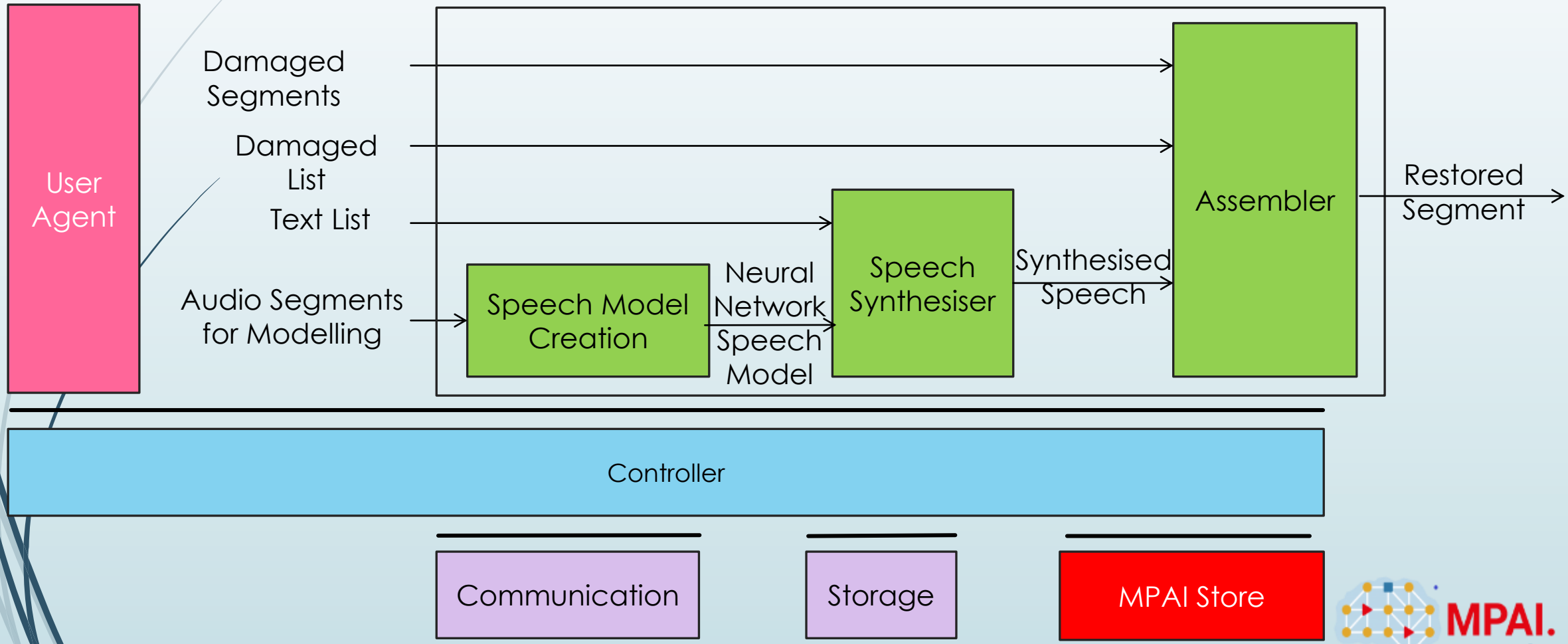
1. **Irregularities** can be found on the **audio signal** and **video of the tape** acquired during the digitization process of an open-reel tape (e.g., damages of the carrier, splices, marks).
2. The detected irregularities are **classified** and **selected** by artificial intelligence algorithms.
3. The selected irregularities are used to **restore** the audio signal and also included in the **copies for preserving and accessing** the open-reel recordings' content.

Speech Restoration System (CAE-SRS): Overview



- **Problem**: damaged or missing voice segments
- Audio restoration has aimed to repair damaged vocal audio
  - … by filtering extraneous noise
  - … or filling minor gaps by extrapolating from surrounding audio
  - … perhaps using AI
- But can't handle substantial vocal gaps: nothing to repair!
- **Proposal**: replace damaged segments
  - Learn model of relevant voice from undamaged voice samples
  - Then (if script available) regenerate damaged segments
  - Fine-tune if necessary
- **Major advantage**: restores even missing segments

Speech Restoration System (CAE-SRS): Workflow



CAE-SRS Workflow: Explanation

Entire **Damaged Segment** can be replaced by synthesized segment, or parts can be synthesized and integrated.

Sequence:

1. **Speech Model Creation** receives **Audio Segments for Modelling**
    1. Recordings used to train a **Neural Network Speech Model** in **Speech Model Creation**
2. **Neural Network Speech Model** passed to **Speech Synthesiser**
    1. … which also receives a **Text List** as input
        a. Each element: string specifying text of a damaged section of **Damaged Segment** (or text of entire **Damaged Segment**).
3. **Speech Synthesiser** produces synthetic replacement for each damaged section (or for entire **Damaged Segment**) and passes to **Assembler**.
4. **Assembler** receives entire **Damaged Segment**, plus **Damaged List** (locations of any damaged sections; null if entire **Damaged Segment** to be replaced).
5. **Assembler** produces **Restored Segment**
    1. … in which repaired sections have been replaced by synthetic sections
    2. … or in which entire **Damaged Segment** has been replaced.

MPAI. community

- **Enhanced Audioconference Experience (EAE):** Separating the speech sources, noise cancellation and packaging the separated speech sources with their spatial information on the transmitter side

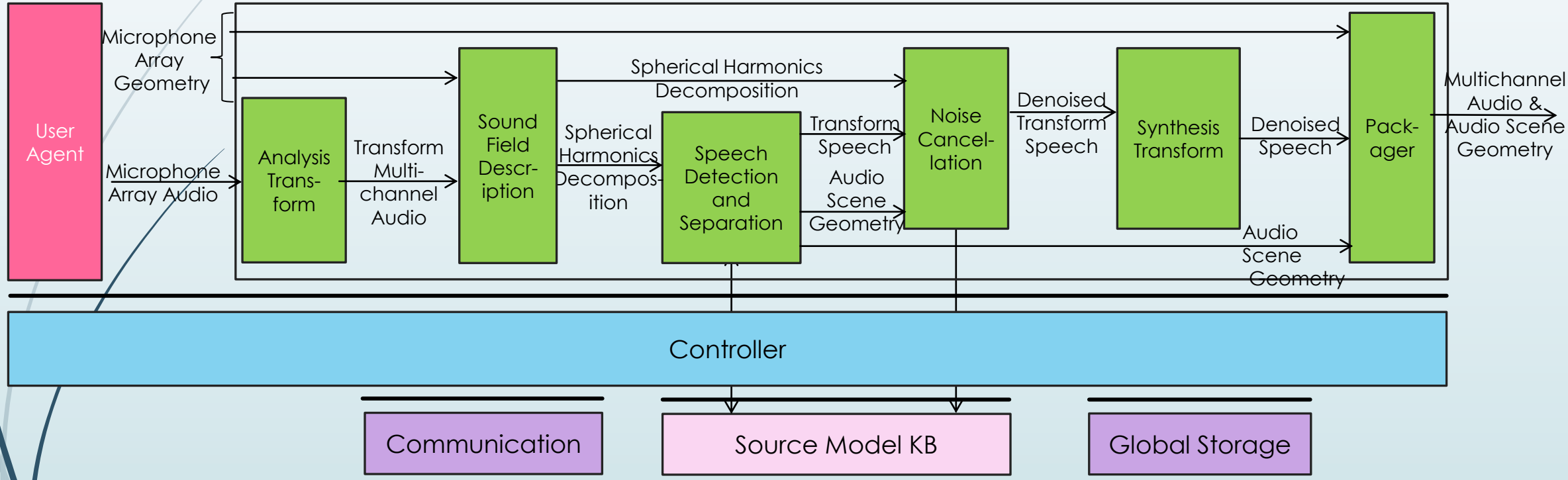
Problem: Want to capture a single speaker, but…
- Interfering speakers
- Noise
- Reverberation

Conventional solution
- Costly
- Inconvenient
- Impractical
- [Typically] manual
- Hard to extend

MPAI-CAE EAE
- Low cost
- Convenient
- Practical
- Automatic
- Flexible / Scalable

Photo by fauxels from Pexels

MPAI-CAE Enhanced Audioconference Experience



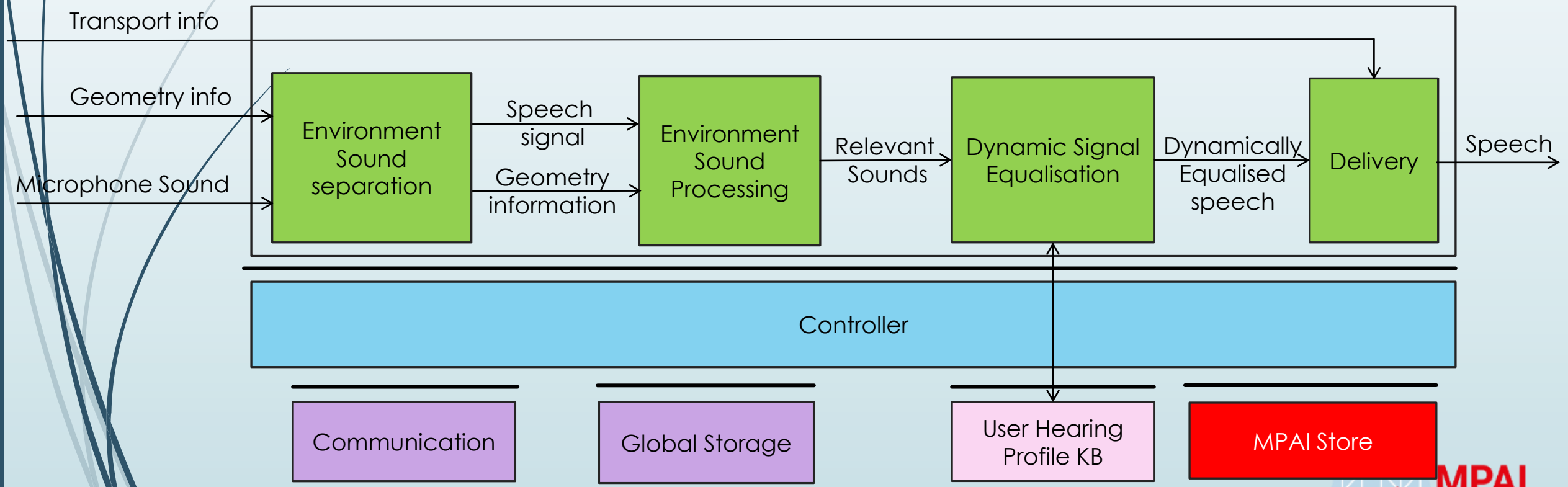
MPAI-CAE EAE: How Does it Work?

- Process signals from a microphone array to capture a compact description of the sound field in the spherical harmonic domain
- Process this representation to detect and separate speakers
- Estimate the number and directions of speakers in the room
- Reduce/eliminate background noise and reverberation in order to enhance the quality and intelligibility of the separated speech signal
- Package speech signals as well as their spatial attributes for reconstruction at the receiver side of the audioconference

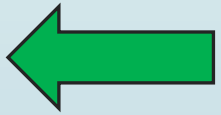
MPAI. community

Next Generation CAE - Candidate MPAI-CAE Version 2 WD 0.1: Stand Alone Use Case

- **Audio-On-the-Go** (AOG): improve sound quality on the go without losing contact with the acoustic surrounding



http://cae.mpai.community/



2/25/2022

67

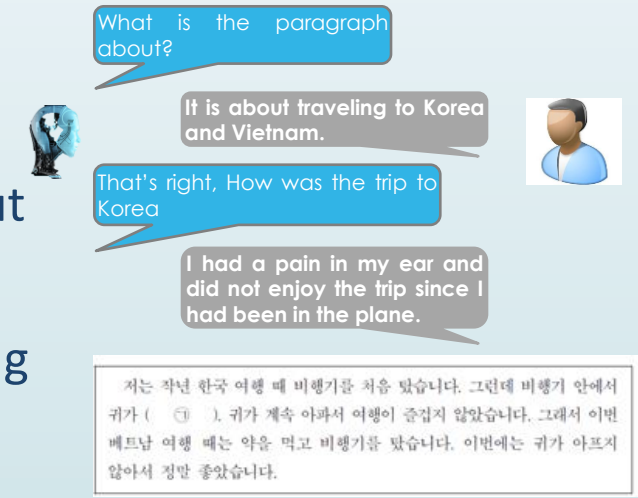
The Multimodal Conversation standard (MPAI-MMC)

Miran Choi

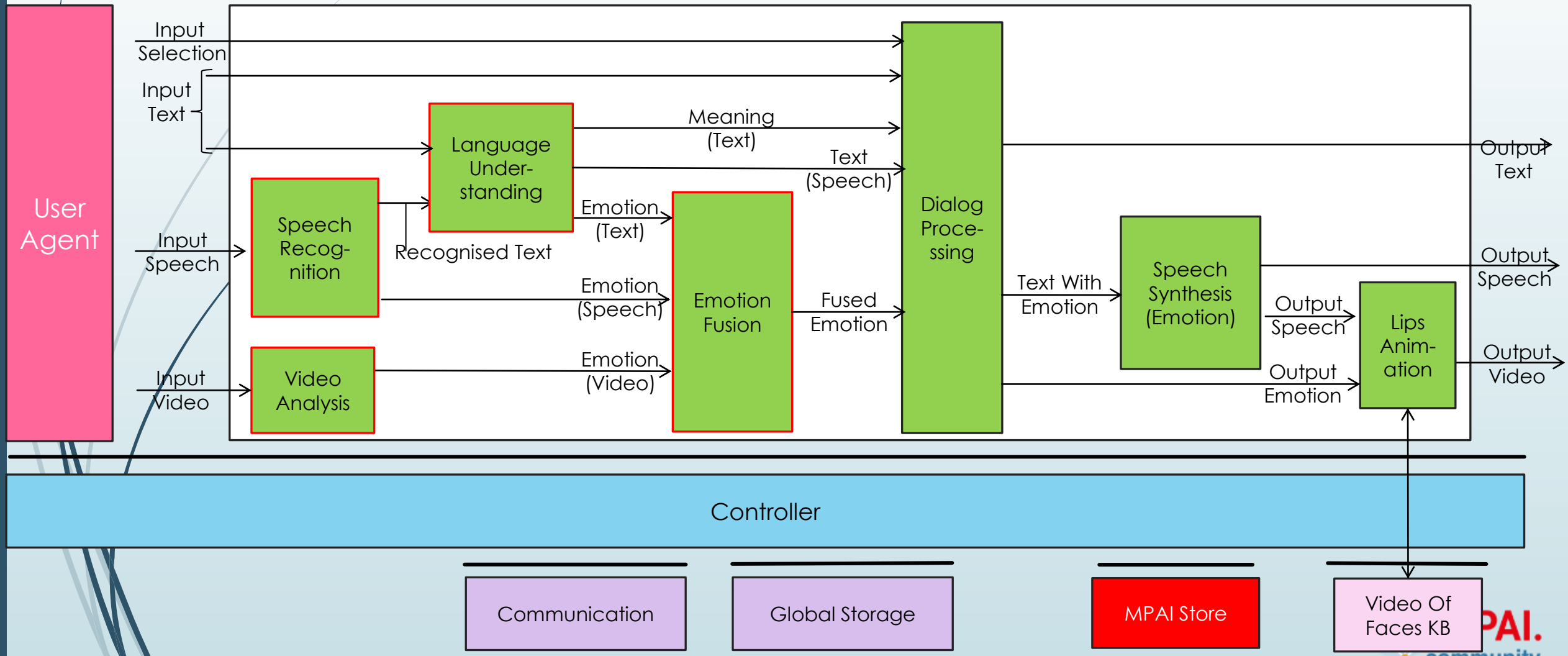
MPAI-MMC - Summary

1. Common Characteristic: human-machine conversation that emulates human-human conversation in completeness and intensity using AI

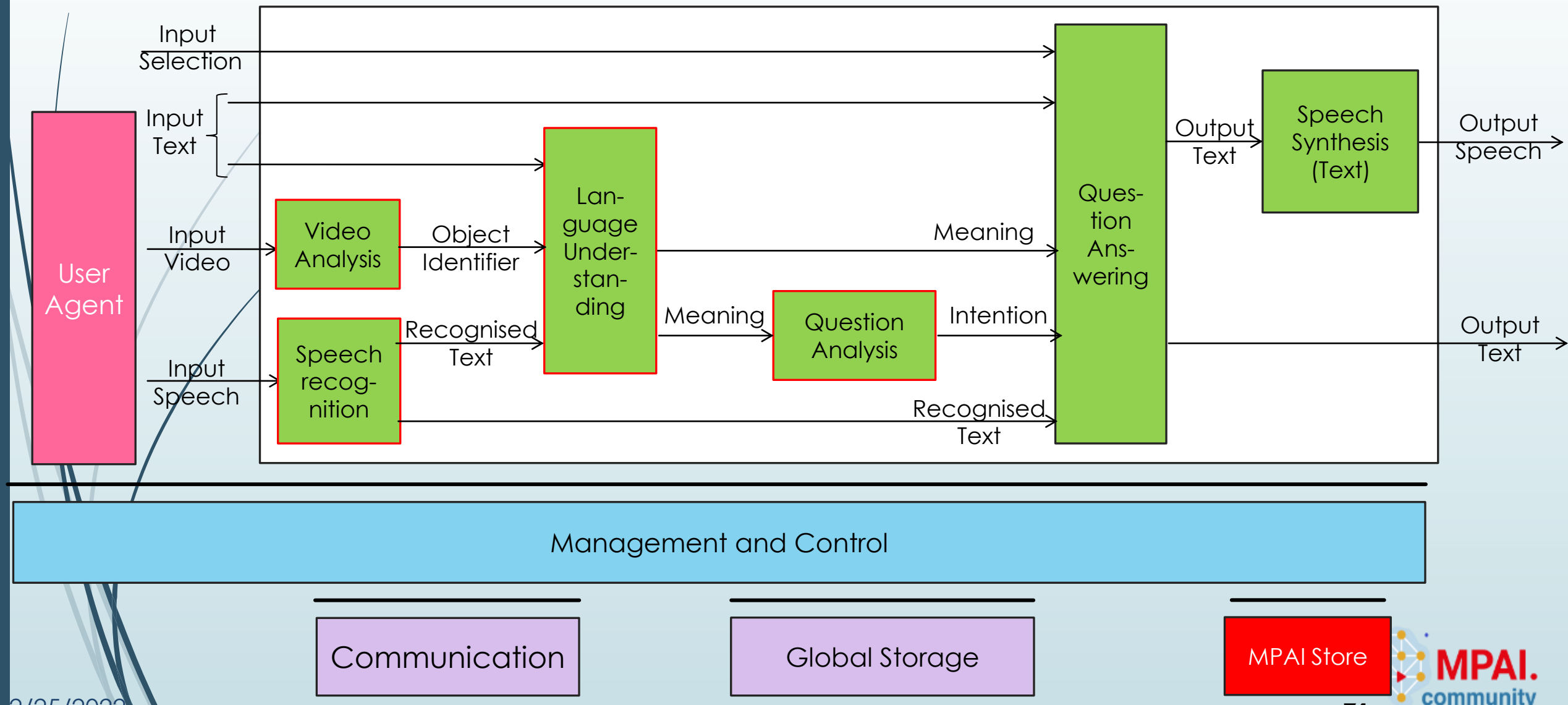
2. MPAI-MMC Use Cases:
    1. Conversation with Emotion (CWE): AV conversation with a machine impersonated by a synthetic voice and an animated face.
    2. Multimodal Question Answering (MQA): Request for information about a displayed object.
    3. Automatic Speech Translation: Translate a spoken sentence preserving the speaker's speech features in the translated speech:
        1. Unidirectional Speech Translation (UST).
        2. Bidirectional Speech Translation (BST).
        3. One-to-Many Speech Translation (MST).



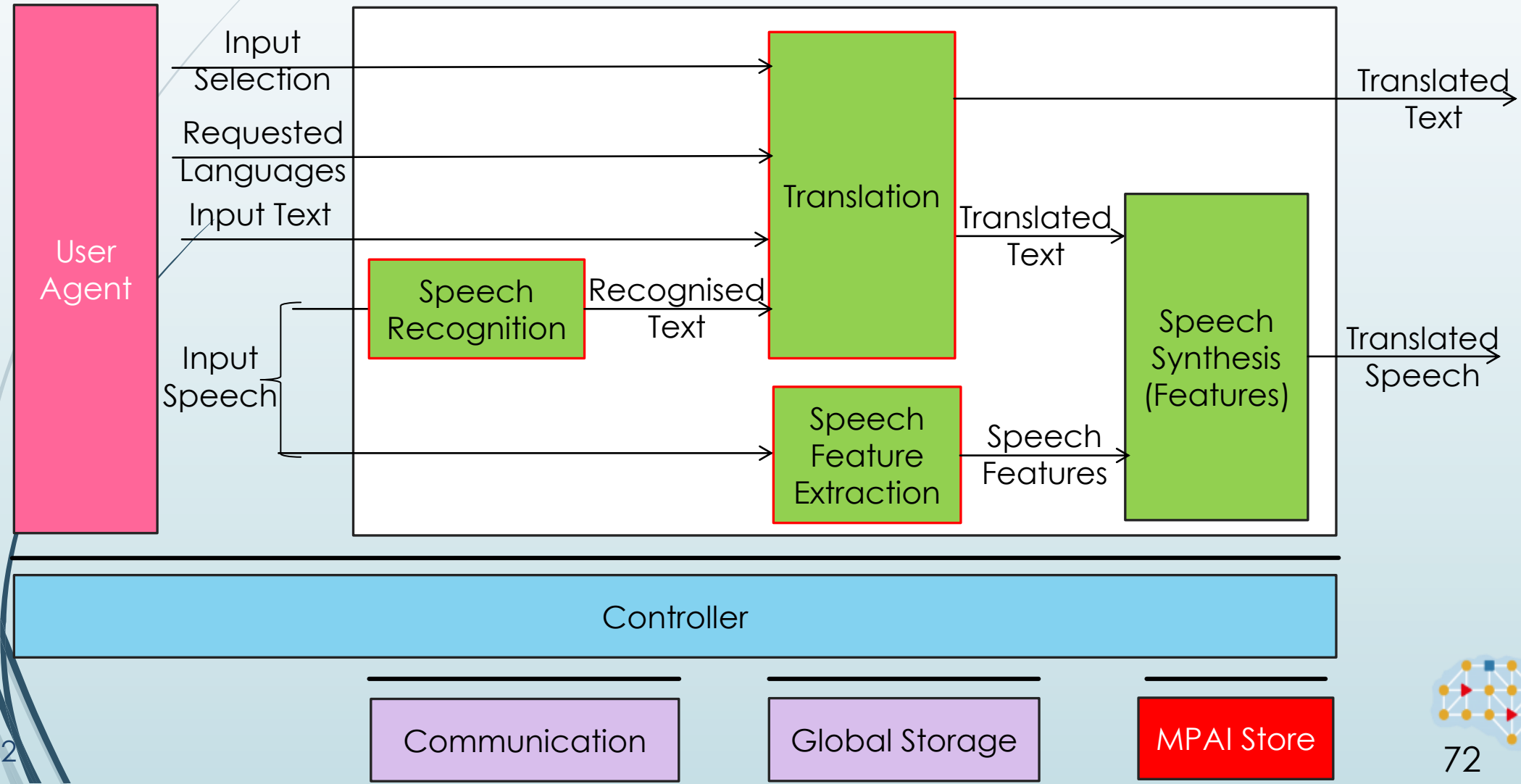
Conversation with Emotion



Multimodal Question Answering



Unidirectional Speech Translation



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 (some 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

Text with Emotion – Syntax

```
{
  "$schema":"http://json-schema.org/draft-07/schema",
  "definitions":{
    "TextWithEmotionType":{
      "type":"object",
      "properties":{
        "text":{"type":"string"},
        "emotionDegree":{"type":"string"},
        "emotionName":{"type":"string"},
        "emotionSetName":{"type":"string"}
      }
    }
  },
  "type":"object",
  "properties":{
    "primary":{"$ref":"#/definitions/TextWithEmotionType"},
    "secondary":{"$ref":"#/definitions/TextWithEmotionType"}
  }
}
```

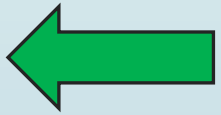
Text with Emotion – Semantics

Name	Definition
TextWithEmotionType	Indicates the Emotion that the text carries.
emotionDegree	Indicates the Degree of the Emotion expressed as human readable words: "Low", "Medium", "High".
emotionName	Indicates the name of the Emotion.
emotionSetName	Name of the Emotion Set which used to describe the Emotion: Basic, Extended or Proprietary Emotion Set.

Future Plan for MPAI-MMC

Activities
Version 2 - Extended data formats for existing AIMs - Human to Connected Autonomous Vehicle Interaction - Mixed-reality Collaborative Spaces
- Version 1 - Reference Software - Conformance Testing - Performance Assessment

http://mmc.mpai.community/



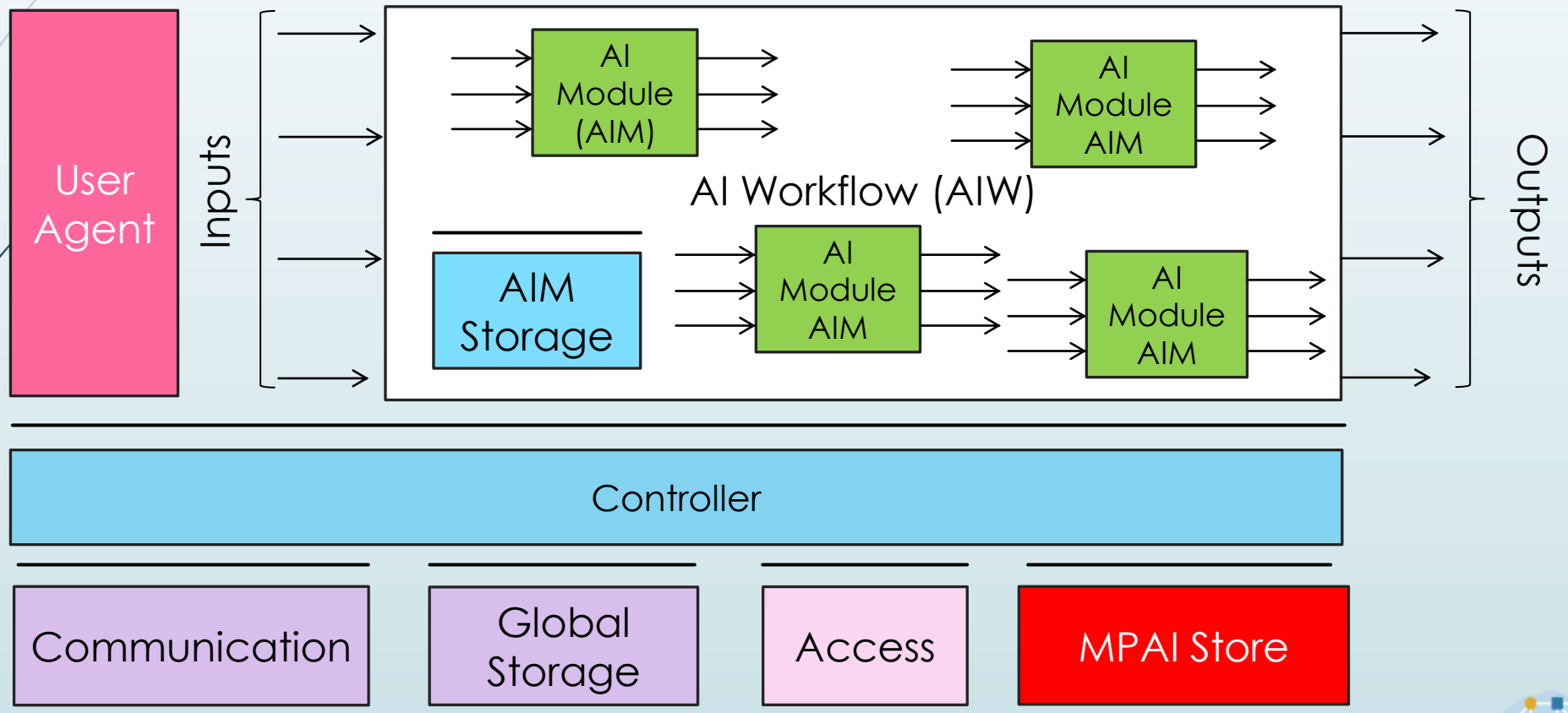
Reference Software, Conformance and Performance

Panos Kudumakis

What is an MPAI standard?

- A collection of 4 documents with associated software and data sets:
  1. **Technical Specification** guiding users to make implementations.
  2. **Reference Software Specification** with attached software implementation and data sets, normatively equivalent to the technical specification.
  3. **Conformance Testing Specification** with datasets and/or the methods to generate them, the tools, and the procedures for testing the technical correctness of an implementation.
  4. **Performance Assessment Specification** with datasets and/or the methods to generate them, the tools, and the procedures to assess how well an implementation "performs". Performance is a multi-dimensional notion meaning, e.g., that the implementation is unbiased.

The AI Framework



Technical Specifications/1

- Technical Specifications (TS) contain normative clauses to be followed by a user wishing to develop a conforming implementation.
- There are two types of TS:
  - System-oriented specs concern support for AI operation, e.g., the MPAI-AIF standard,
  - Application oriented specs concern specific domains, e.g., MPAI-CAE and MPAI-MMC.
- An application standard contains applications aka use cases (UC), e.g., Multimodal Conversation TS contains 5.
- Each TS is identified by 3 characters (e.g., MMC); each UC by 3 characters, e.g., CWE.
- For each UC, the TS specifies the AI Workflow (AIW) implementing the UC with the AIMs

Technical Specifications/2

- AIWs/AIMs are specified by
  1. The function executed by the AIW/AIM.
  2. The syntax and semantics of the input and output data of the AIW/AIM.
  3. The topology of the AIMs composing the AIW.
- The TS includes the syntax and semantics of all data formats used by all AIMs and AIWs in a single section because some data formats are shared across AIMs/AIWs. Some are also shared by different standards.
- MPAI plans on developing a standard collecting all AIMs/AIWs used by multiple standards.

Reference Software Specification

- A Reference Software Specification contains
  - Conditions of use
  - Descriptions of the Reference Software Implementation
  - Relevant data sets used by the AIWs and the AIMs defined in the Use Cases of the MPAI Technical Specification they refer to.
- Software Implementations of AIMs are made available in one or more than one of the following Software Forms:
  1. As source code providing a satisfactory user experience and/or functionality.
  2. As source code providing a more limited user experience, sufficient to assess the value of the standard.
  3. As compiled code providing a satisfactory user experience and/or functionality.
  4. As source code software wrapping access to a third-party service enabling a conforming AIM Implementation (Wrapper AIM).

Reference Software Disclaimers

- RS is a working Implementation of the Standard, not suitable for any other purposes.
- RS is not a ready-to-use product, it only exposes correct I/O interfaces.
- RS is released according to the MPAI modified BSD licence.
- A compiled AIM of the RS may not be used in commercial products or services.
- Sample input data is provided if required to operate the Software.
- If RS uses a knowledge base (KB), access to a conforming KB is provided.
- MPAI makes no guarantee that AIMs/AIWs pass Performance Assessment.
- Users shall check whether they have the right to use referenced 3rd party software.
- Wrapper AIM (W-AIM)
  - Provided only in source code for the part calling the service.
  - Accompanied by description and references to documentation of 3rd party services.
  - Submitter maintains W-AIM for 12 mo. after 1st publication of RS.
  - If the 3rd party discontinues service within 12 mo., best effort to find a similar third-party service.
  - If Wrapper-AIM does not function after 12 mo., users should update W-AIM by themselves.

Conformance Testing/1

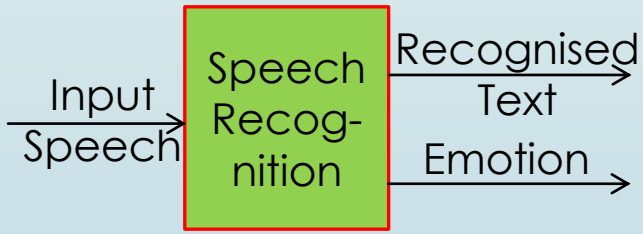
- TS sets the Conformance Testing (CT) rules to determine whether an implementation is technically correct.
- In digital media an "encoder" produces data that a "decoder" can decode. Therefore, CT can be formulated as
  - "Provide bitstreams and check that the decoder under test can correctly decode them"
  - "Feed bitstreams of the RS encoder and check that RS decoder can correctly decode them".
- Two digital media decoders may very well not decode the same bitstream in the same way because two decoders may have: different initial state & different precision levels. While different, they may very well pass the conformance testing.
- In MPAI different outputs from different implementations is the norm. AIMs may contain NNs of unspecified architectures, trained with unspecified data sets.

Conformance Testing/2

- An MPAI CT specification
  - Defines the datasets and/or the methods to generate them, the tools, and the procedures (the "Means") to test the technical correctness of an implementation.
  - Specifies the tolerance of the output of an AIM given the input data used for the Test.
- MPAI CT is based on the following process:
  - Valid Conformance Tests may only be carried out by the MPAI Store using the Means.
  - When the Datasets are not made public, the MPAI Store randomly selects a large subset of test items and feeds it to the AIM under test.
  - End users are informed of which Conformance Tests it has passed in two way: a descriptive sentence and the full table of results.
  - The submitter can verify that the AIM provides the result reported by the MPAI Store by verifying that, the submitted AIM by being subjected to the inputs corresponding to the identifiers of the records of the relevant CT Dataset, produces the same output as the MPAI Store.
  - Some Datasets may only be accessible for a time window and for a number of accesses after receiving the test results.

Conformance Testing/3

- MPAI CT reflects the state of technology at the time of publication of CT spec.
- It may be extended by new Versions of the CT Specification, as follows:
  - A new Version of the CT Specification applies to the current Version of the TS.
  - The current Version of the TS may apply to a new, current or preceding version of the CT Specification. If it applies to a preceding version, MPAI should not have discontinued support of that Version of the CT Specification.
- For example, in Conversation with Emotion, an AIM takes input speech and produce text and emotion of the input speech as output.



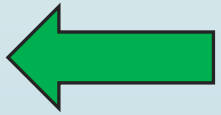
Conformance Testing/4

- In this case CT is defined as the ability of an implementation to produce Unicode Text and Emotion expressed as one of the MPAI standard Emotions.
- For a user, knowing that the data are syntactically and semantically correct is irrelevant if Recognised Text is not what was contained in the Input Speech and the Emotion is declared as Angry when in the Input Speech, it was Happy.
- Imposing that an AIM implementation Conforms only when the output is perfect is not realistic, because no implementation can be perfect in all cases.
  - The word error rate is known way to give a grade to a speech recogniser.
  - Nothing equivalent for Emotion. MPAI is considering three "emotion error rate" measures:
    - Use human testers,
    - Train a network to measure the distance between Emotions
    - Define an emotion space with suitable metrics.

Performance Assessment

- Researchers are engaged in the problem to determine when AI is reliable.
- MPAI defines **Performance** of an implementation as the set of the following attributes:
  - **Reliability**: implementation performs as specified by the standard e.g., within the application scope, with stated limitations, and for the period of time specified by the Implementer.
  - **Robustness**: the implementation can cope with data outside of the stated application scope with an estimated degree of confidence.
  - **Replicability**: the assessment made by an entity can be replicated, within an agreed level, by another entity.
  - **Fairness**: the training set and/or network is open to testing for bias and unanticipated results so that the extent of applicability of the system can be assessed.
- Performance can have **Grades**, possibly depending on the specific domain.
- **Performance Assessment** is the specification defining datasets and/or the method to generate them, the tools, and the procedures used to assess the performance of an implementation.

https://mpai.community/



The Governance of the MPAI Ecosystem (MPAI-GME) standard

Paolo Ribeca

Interoperability

- Standards are defined to guarantee interoperability
- MPAI defines Interoperability as the ability to replace an AIW or an AIM Implementation with a functionally equivalent Implementation.
- MPAI defines 3 Interoperability Levels of an AIW executed in an AIF:
  - Level 1 – The AIW is Implementer-specific satisfying the MPAI-AIF Standard.
  - Level 2 – The AIW is specified by and tested against an MPAI Application Standard.
  - Level 3 – The AIW is specified by and tested against an MPAI Application Standard, and certified by an MPAI-appointed Performance Assessor.

The foundations of the MPAI Ecosystem

- Developers: develop components → require interoperability to bring their components to the market.
- Integrators: assemble components → require interoperability to be able to assemble third party components.
- Consumers: use assembled components → require that components and their assembly be trusted.
- We need an entity guaranteeing:
  - Interoperability.
  - Trust.
  - Availability.

The need for the MPAI Store

To verify that implementations of MPAI components:

- Are secure.
- Conform with the MPAI standards defining the components, i.e.,
  - Implement the claimed function
  - Expose the specified interfaces.
  - Offer a minimum level of quality.
- Can be used reliably, according to the MPAI definition of Performance:
  - Reliability: performs within the application scope, with the stated limitations.
  - Robustness: handles data outside the stated application scope with an estimated degree of confidence.
  - Replicability: performance assessment can be replicated within the stated Grade.
  - Fairness: performance assessment does not detect bias or unanticipated results.

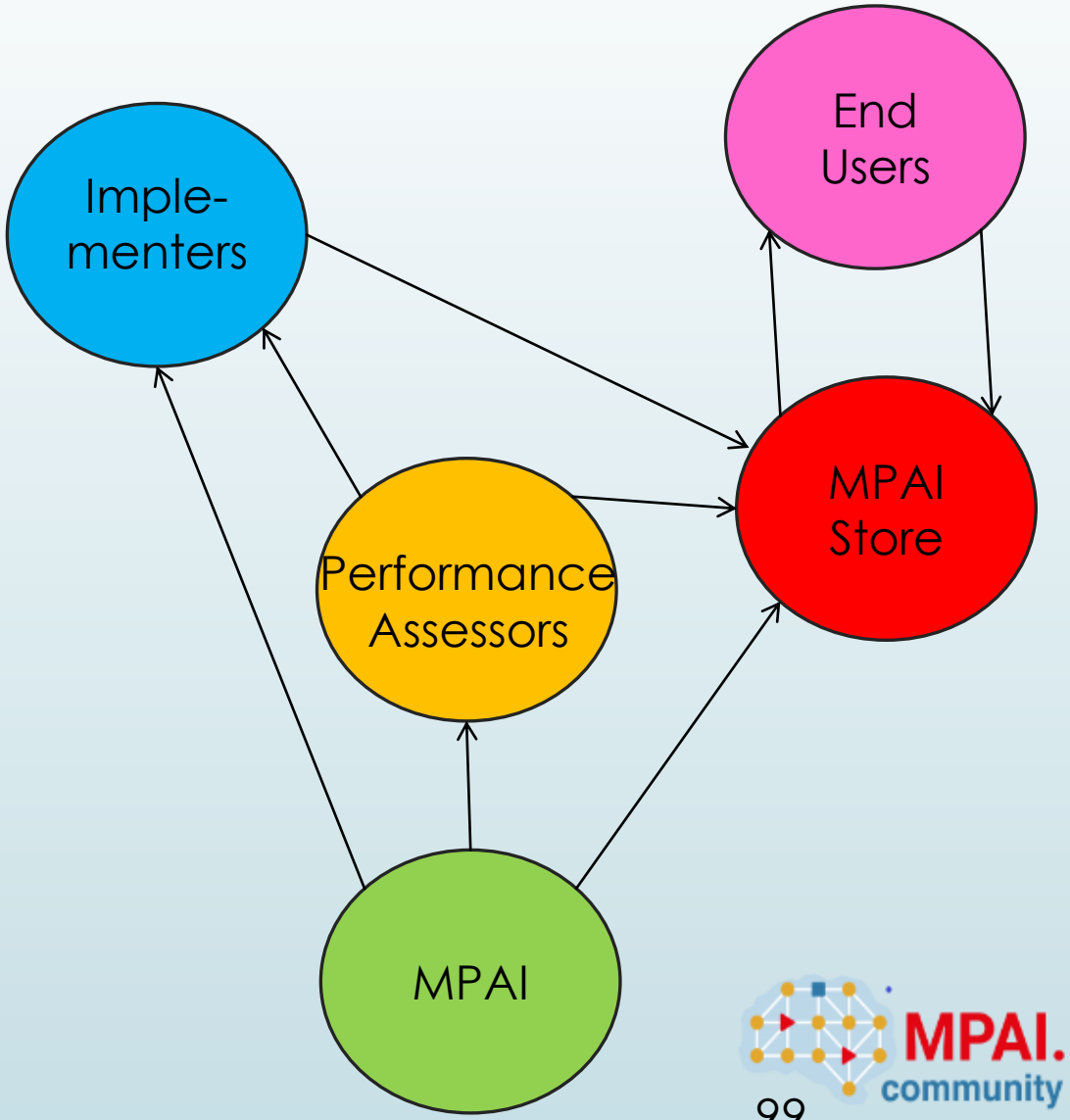
The functions of the MPAI Store

The MPAI Store implements the requirements above as follows:

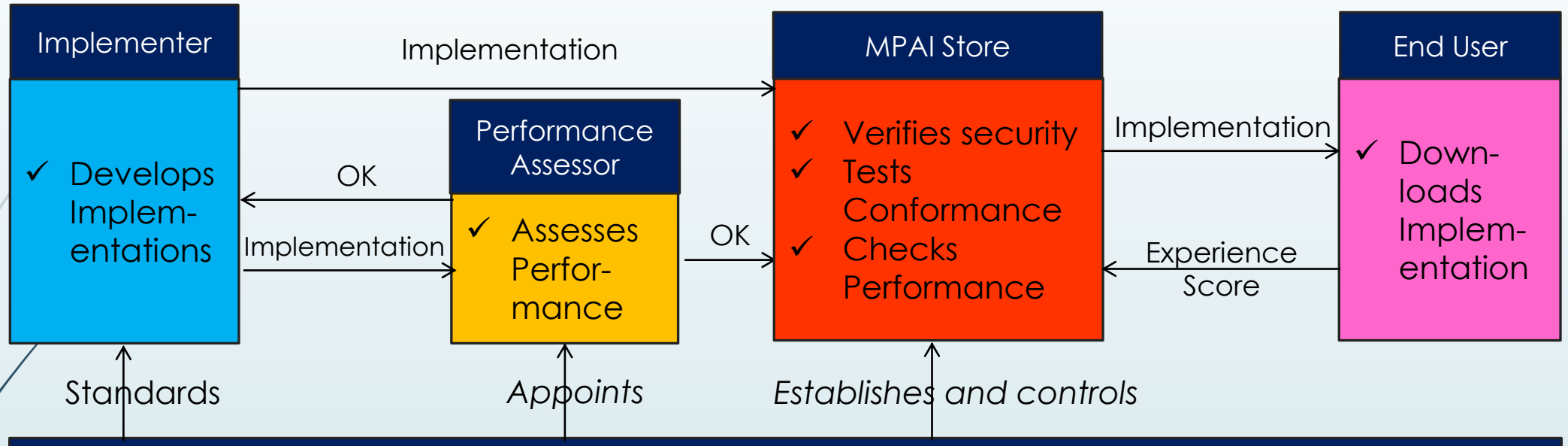
- Receiving submissions of MPAI implementations.
- Verifying security of the implementations.
- Testing conformance of implementations, including function, interfaces and quality.
- Collecting results of performance assessment
- Publishing a catalogue of implementations with clearly identified interoperability level
- Guaranteeing:
  - High-availability component download as per MPAI-AIF specifications.
  - Centralised licensing system with suitable access interfaces for both developers/integrators and end users.
  - Component integrity by offering signature services.
- Managing a reputation system for implementations.

The MPAI Ecosystem

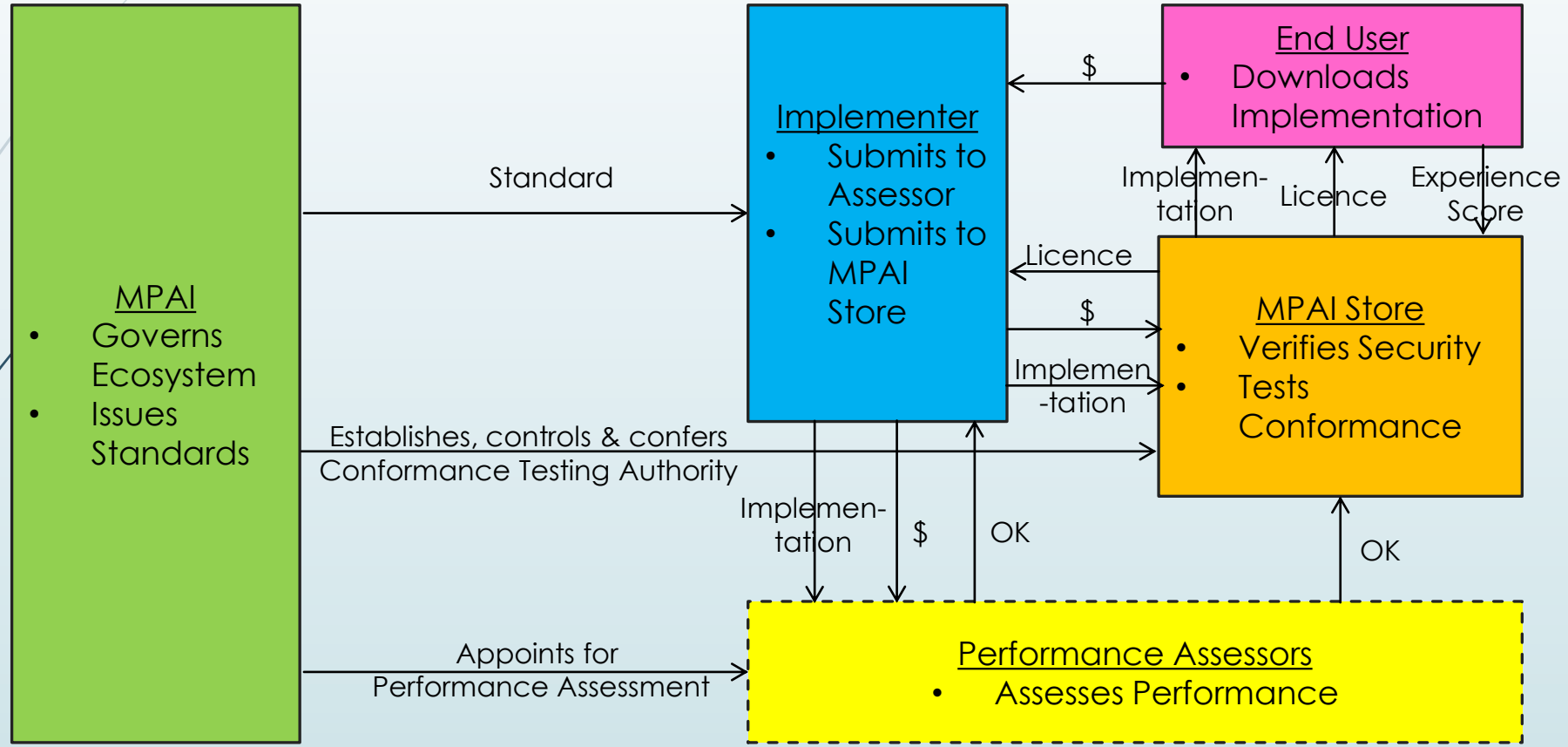
- The direct and indirect results of MPAI activities give rise to an Ecosystem
  - MPAI — Standards
  - Implementers — Implementations
  - MPAI Store — Test & distribution
  - Performance Assessors — Reliability
  - End Users — Consumption
- Governance of the MPAI Ecosystem sets its rules of operation.



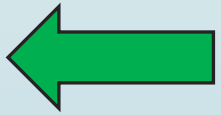
The Governance of the MPAI Ecosystem



How the MPAI Ecosystem works



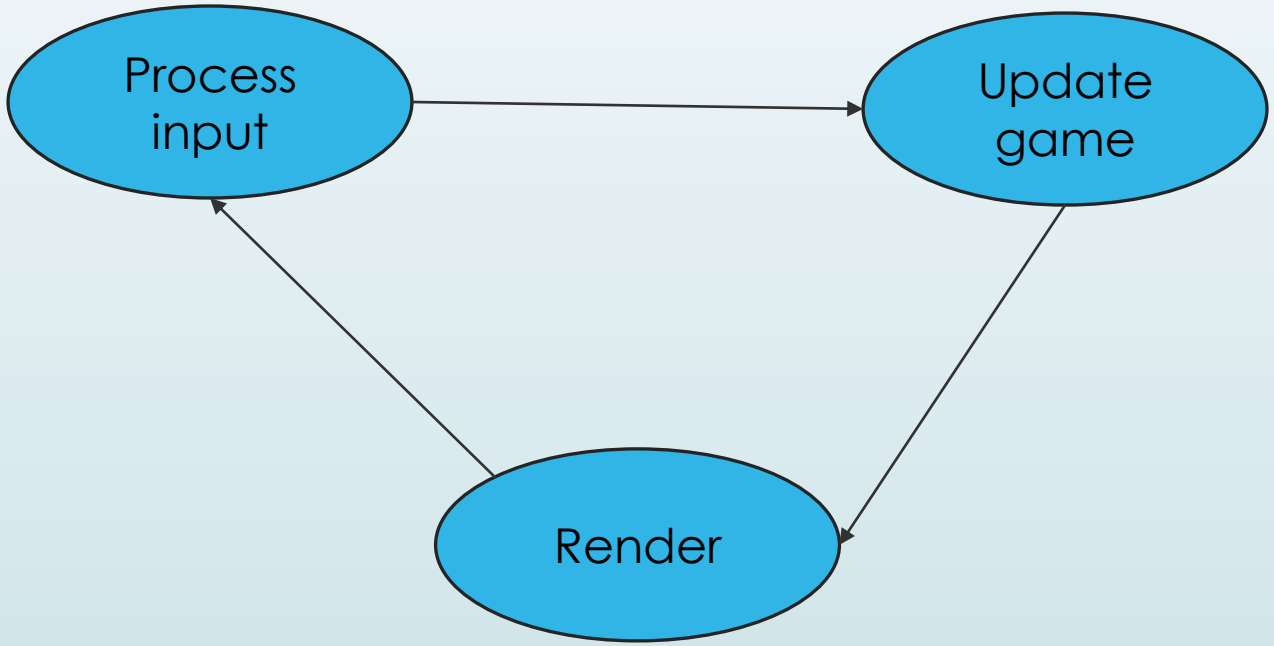
http://gme.mpai.community/



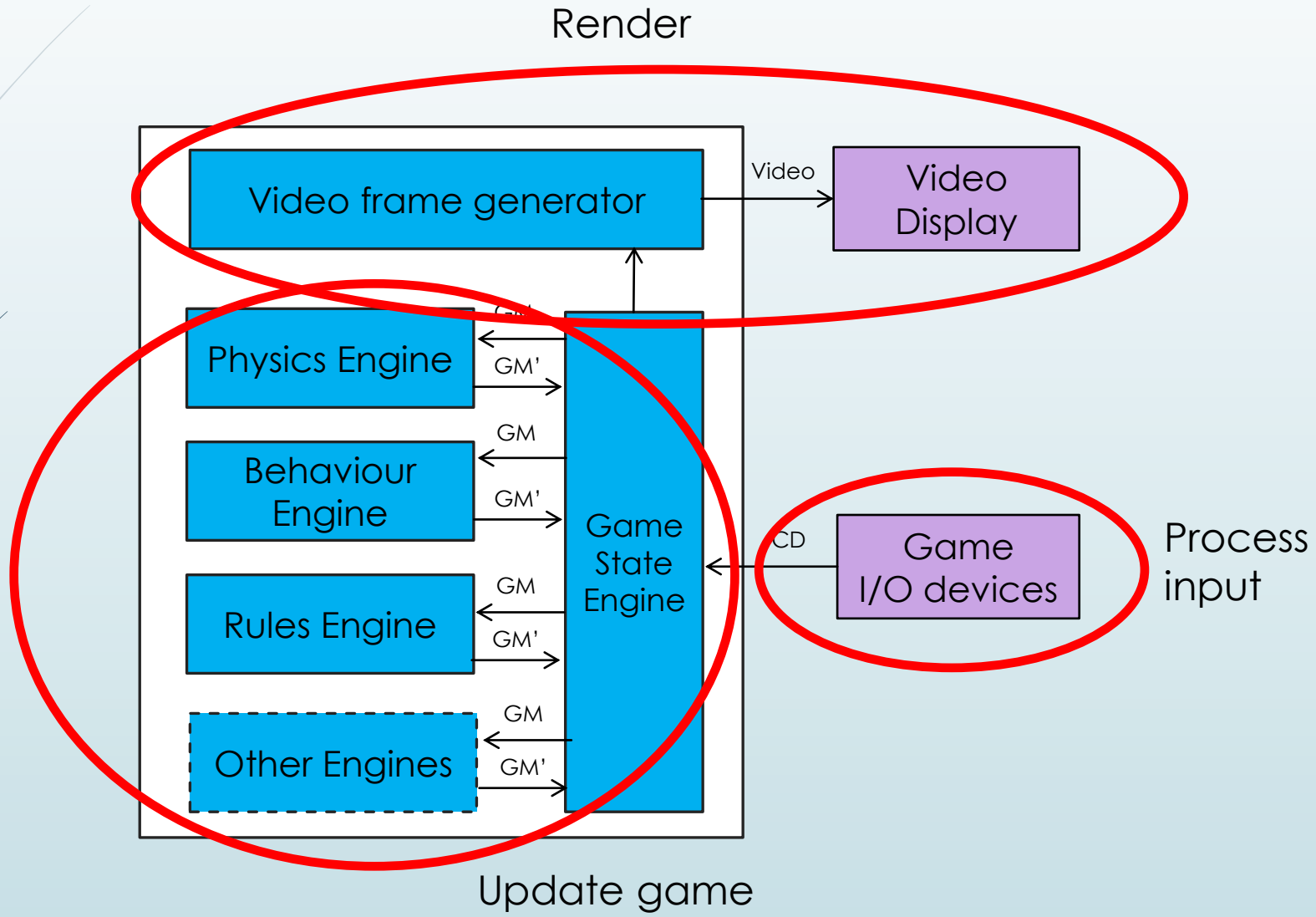
The Server-based Predictive Multiplayer Gaming (MPAI-SPG) Project

Marco Mazzaglia

The game loop



Offline gaming

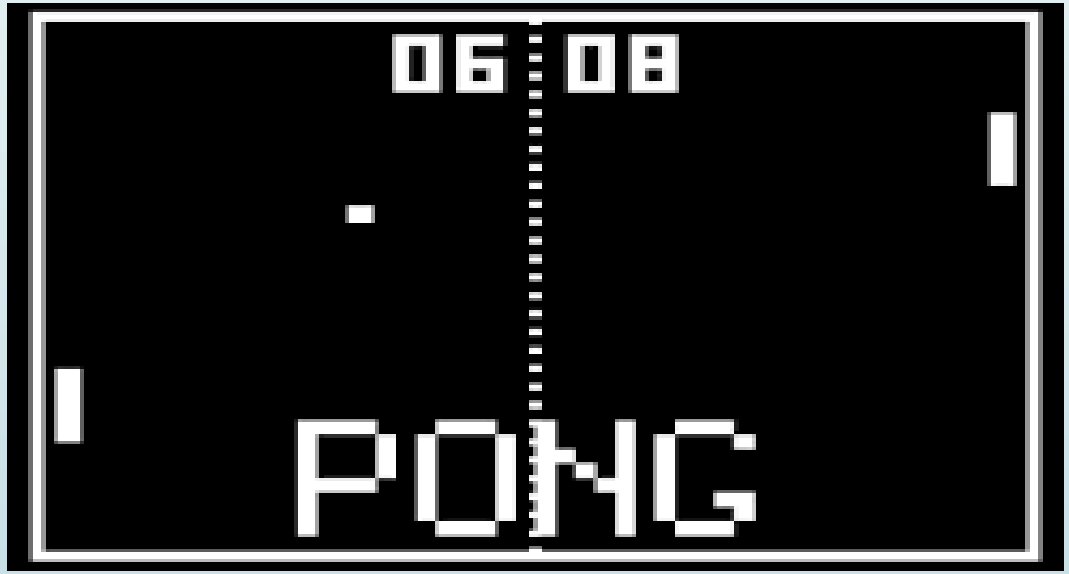


Game State (GS)

The Game State is the collection of different variables and data of the game.

It is a representation, a snapshot of the system in a moment of the execution of the game.

If we consider the video game Pong, a primitive Game State of Pong will be:



Game state: a sample

Game State
ball (xPosition, yPosition, vectorX, vectorY)
matchStatus (state, scorePlayer1, scorePlayer2, victoryScore)
player1 (xPosition, yPosition, vectorX, vectorY)
player2 (xPosition, yPosition, vectorX, vectorY)

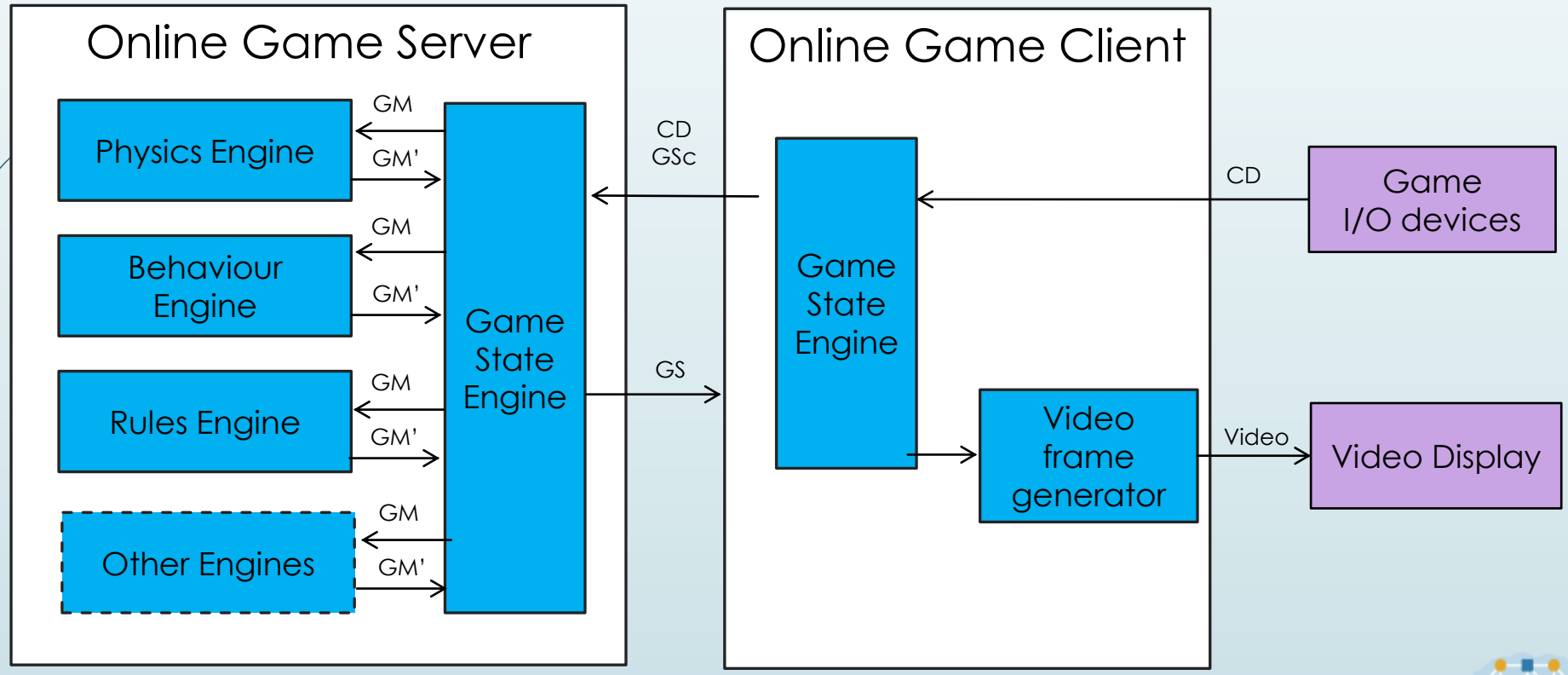
Controller Data (CD)

The controller data are the inputs created by different controllers. They can be digital inputs (joytick), analog inputs (paddle or analog sticks) and so on…

This is an information that changes the status of the system when it is received by the components of the game engine.

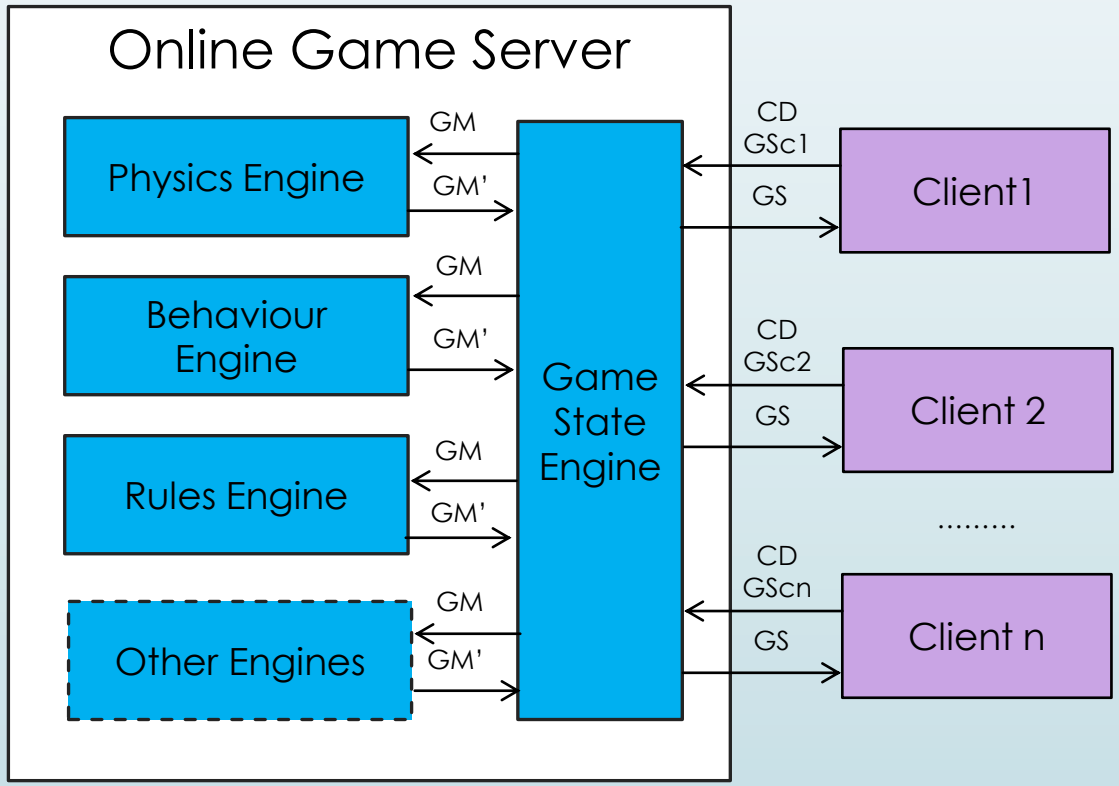
Online gaming (single player)

CD = Controller Data	GS = Game State	GSc = Client Game State
GM = Game Message from GSE to Other Engines	GM' = Game Message from Other Engines to	Video = video streaming to the thin client monitor



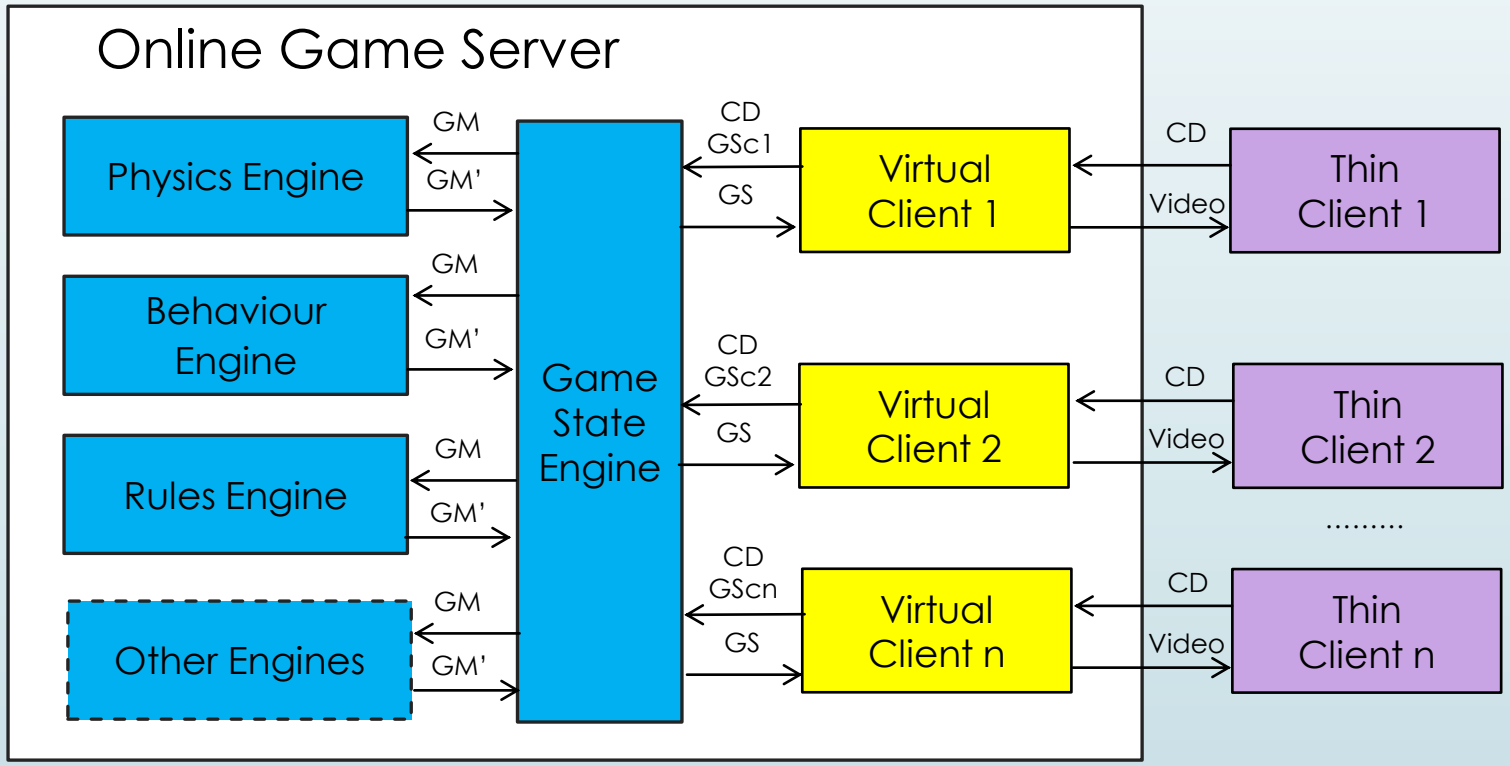
Online gaming (two or more players)

CD = Controller Data	GS = Game State	GSc = Client Game State
GM = Game Message from GSE to Other Engines	GM' = Game Message from Other Engines to	



Multiplayer cloud gaming

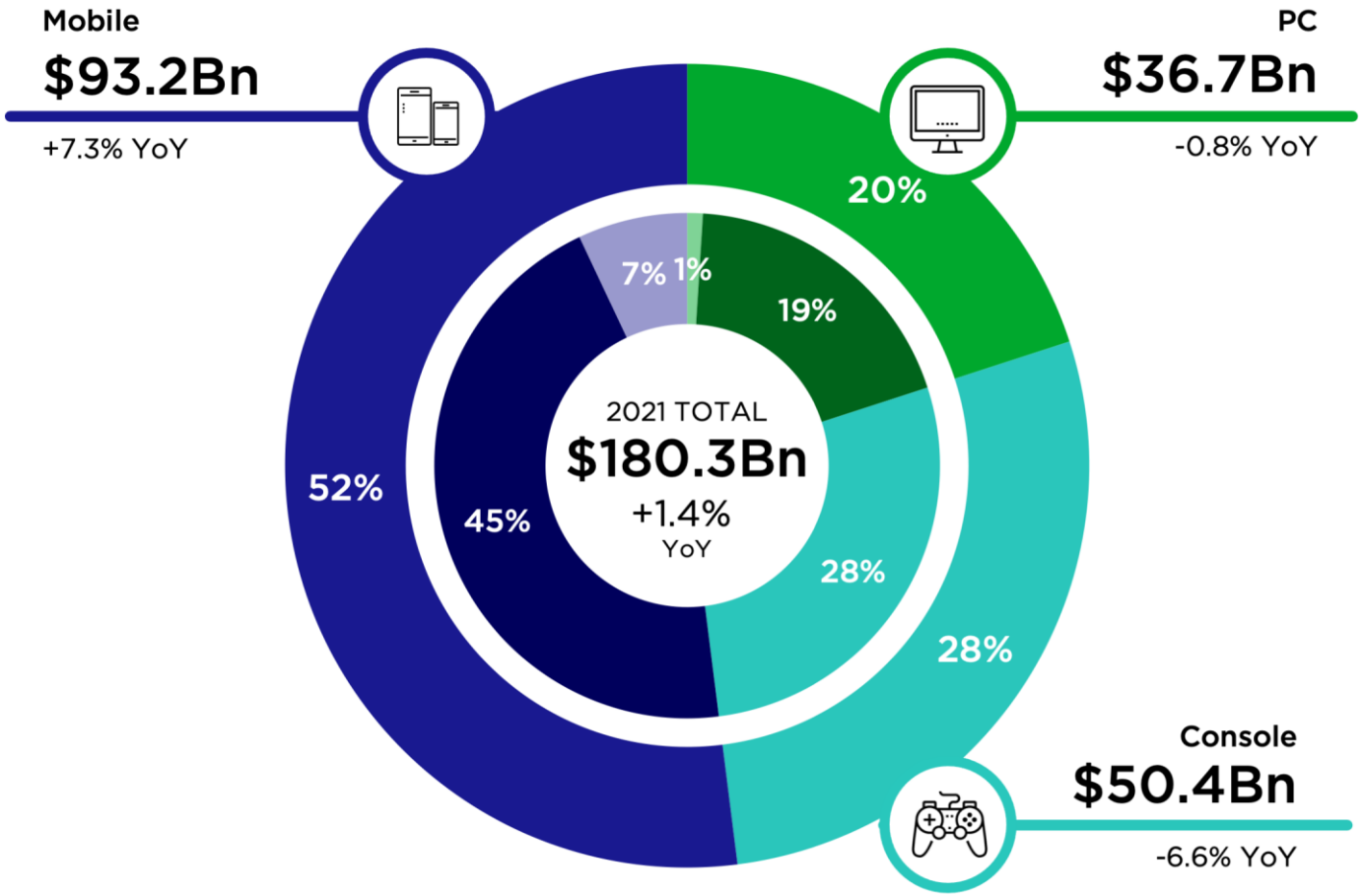
CD = Controller Data	GS = Game State	GSc = Client Game State
GM = Game Message from GSE to Other Engines	GM' = Game Message from Other Engines to	Video = video streaming to the thin client monitor



WHY MPAI-SPG?

2021 Global Games Market

Per Device & Segment With Year-on-Year Growth Rates

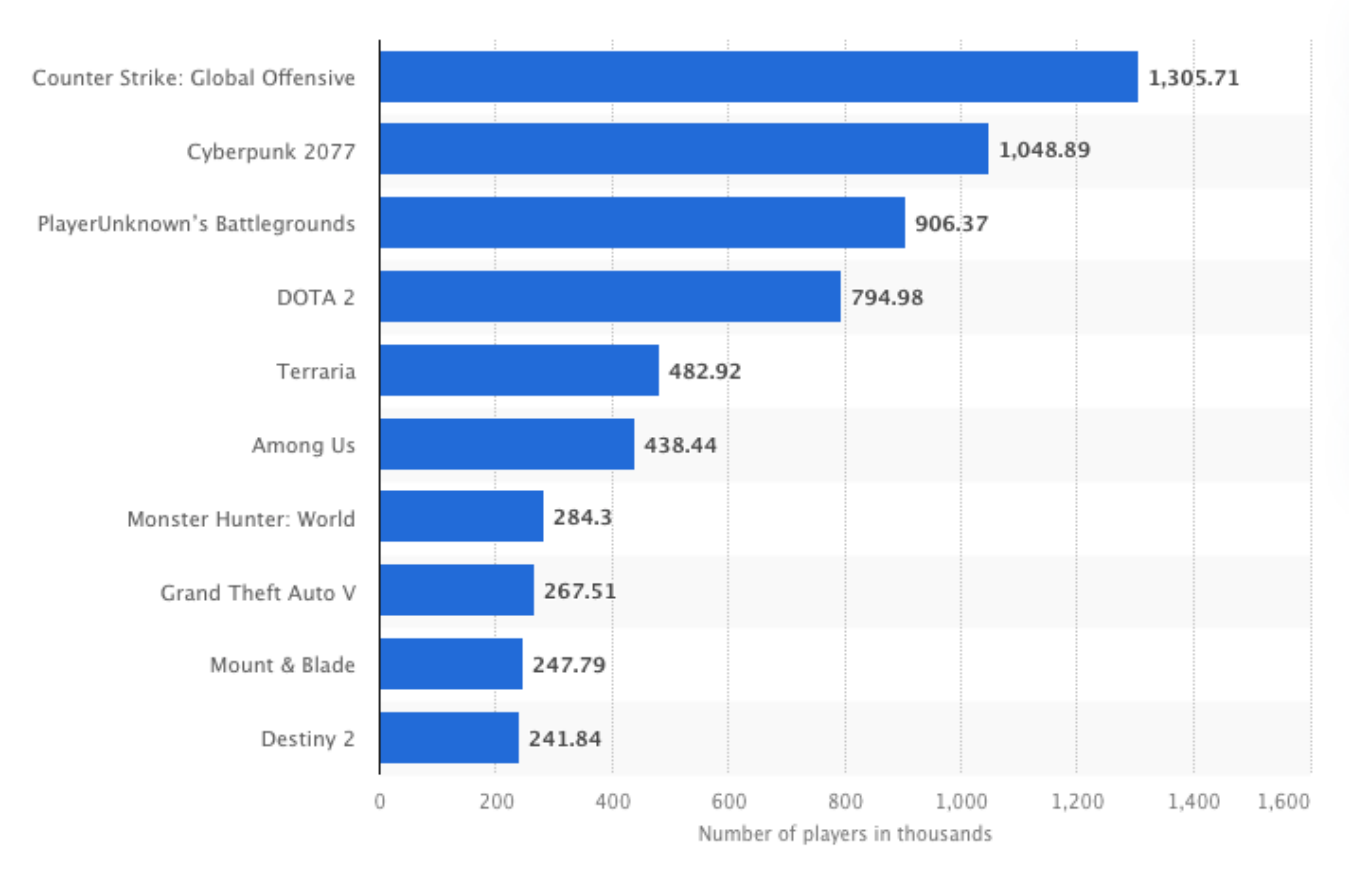


Source: ©Newzoo | Global Games Market Report | January 2022

newzoo.com/globalgamesreport

In the Game Industry online games are a profitable sector that involves all the platforms (game consoles, personal computers and mobile devices like smartphones and tablets)

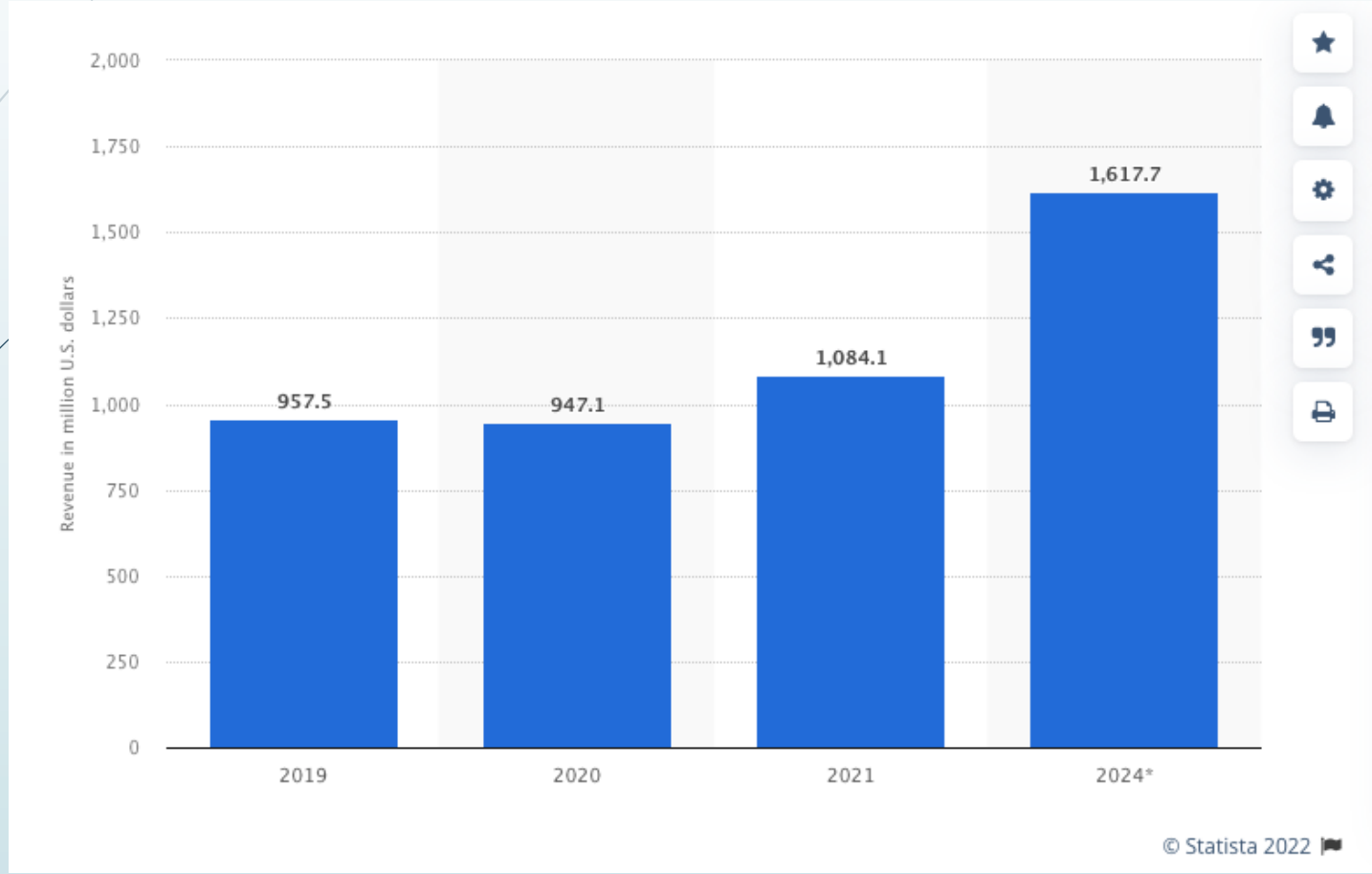
In Steam, the digital video game platform and store of Valve, 9 out of the first 10 most played games of 2020 are online games.



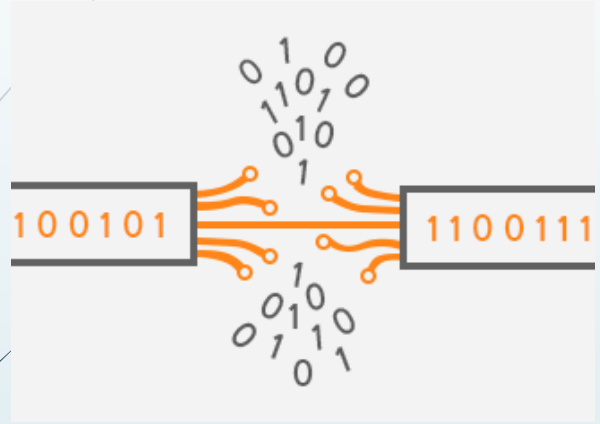
https://www.statista.com/statistics/656278/steam-most-played-games-peak-concurrent-player/

The online games are so important that have brought in the development of a new branch of the Game Industry: the e-Sports.

e-Sports also are attracting a huge number of companies not closely related to the Game Industry as Sponsors, increasing the incomes.



Notwithstanding the results, there are still two problems that plague online gaming:



the problem of latency

or packet loss

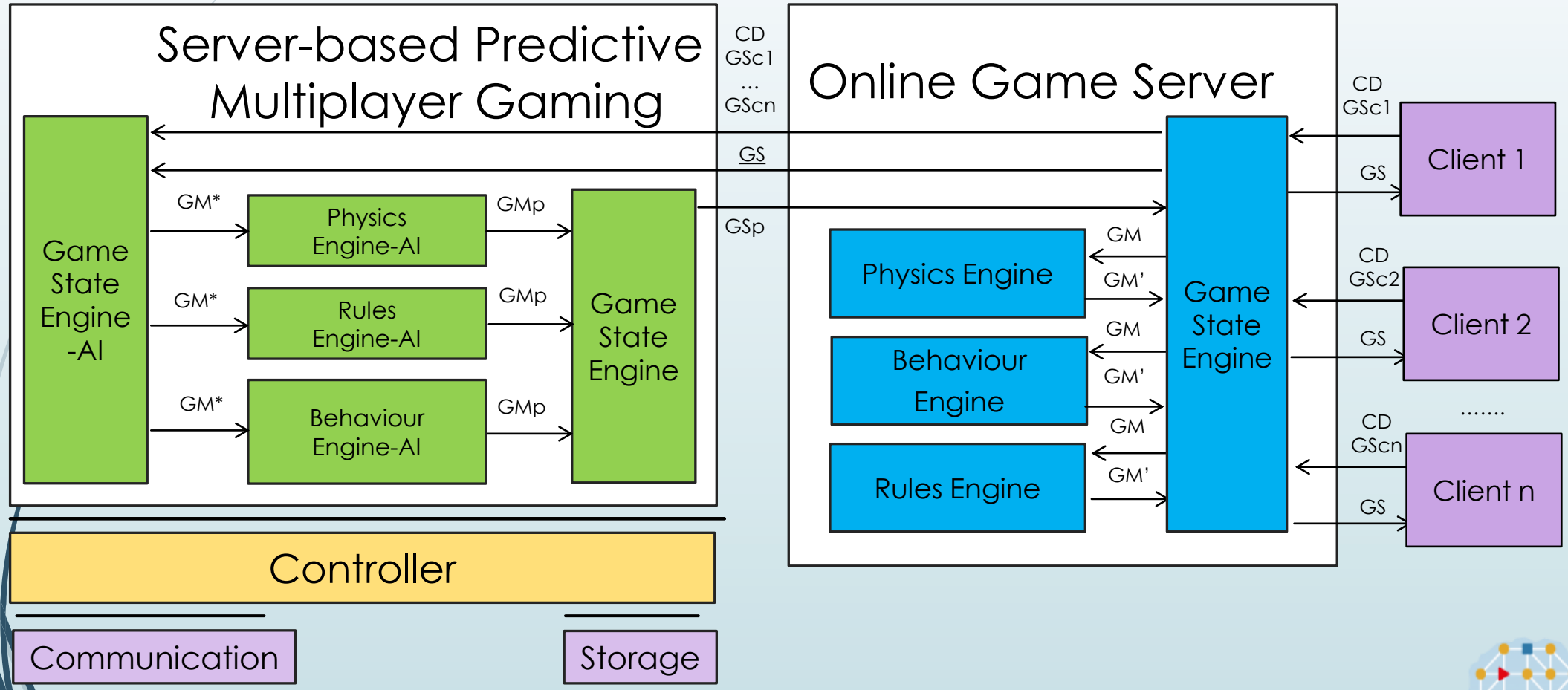
Players trying to cheat

to win games

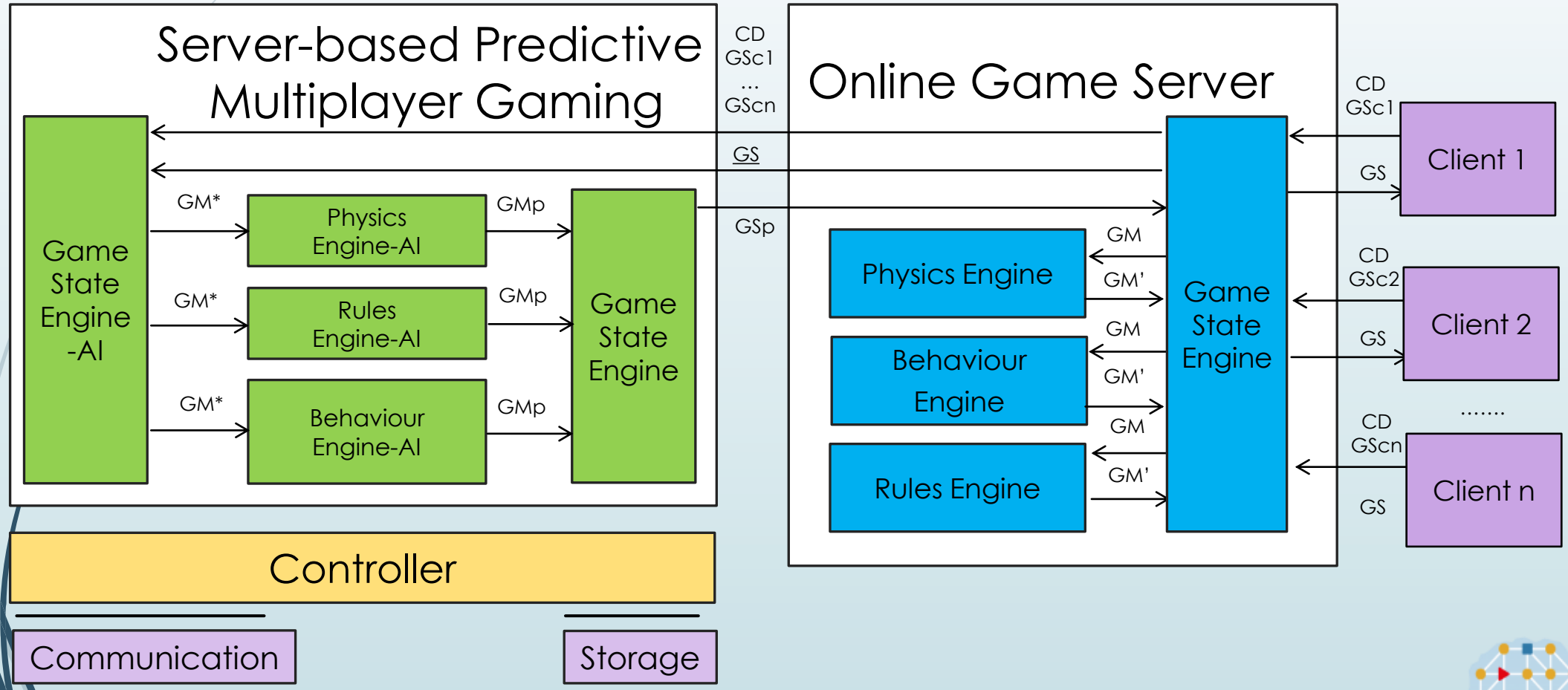
MPAI-SPG works in order to solve or mitigate these two important issues.

HOW?

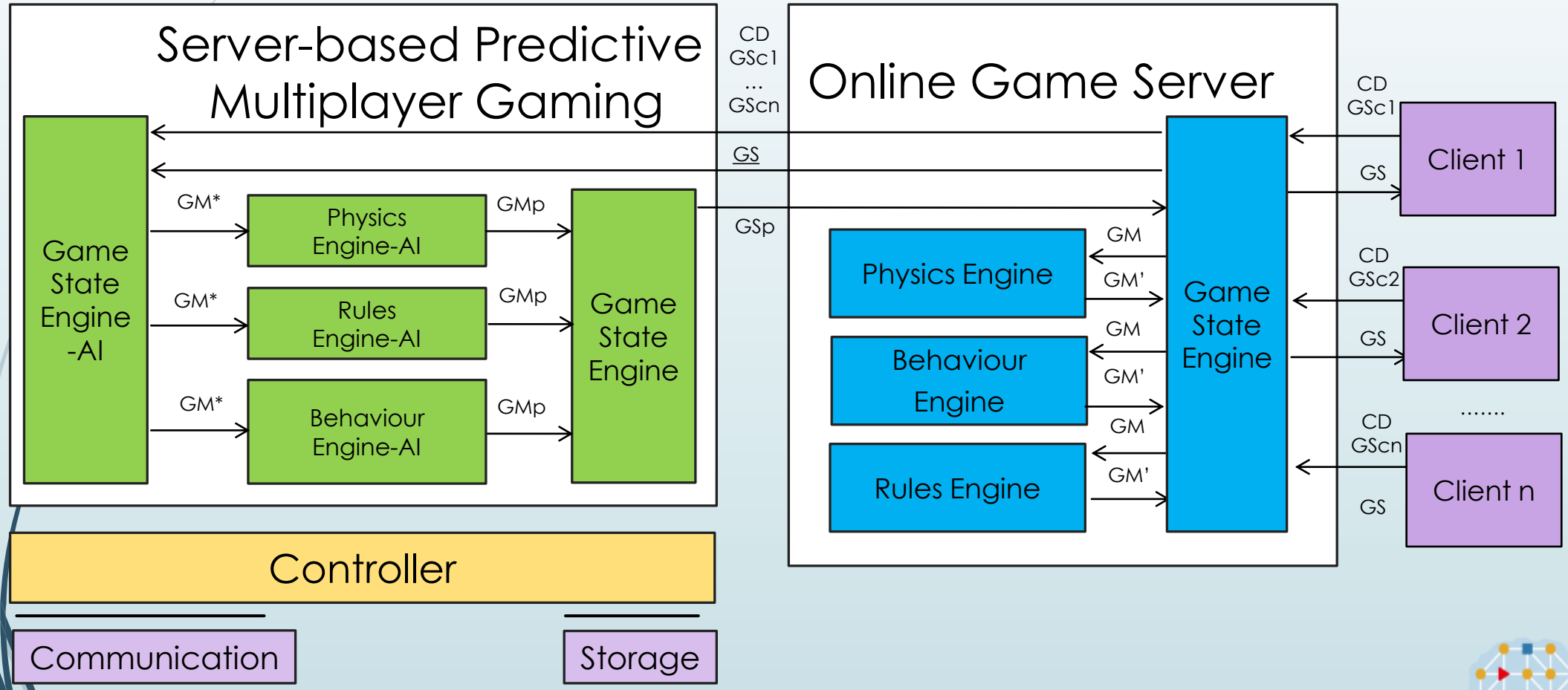
The idea starts to generalise the structure of a game engine and then to make a digital twin where each main component is a neural network (green boxes) able to learn the behaviour of the physical component.



The goal is to create a prediction server based, considering all the game states of all previous matches both simulated by CPU (machine learning) and played until now by human players.

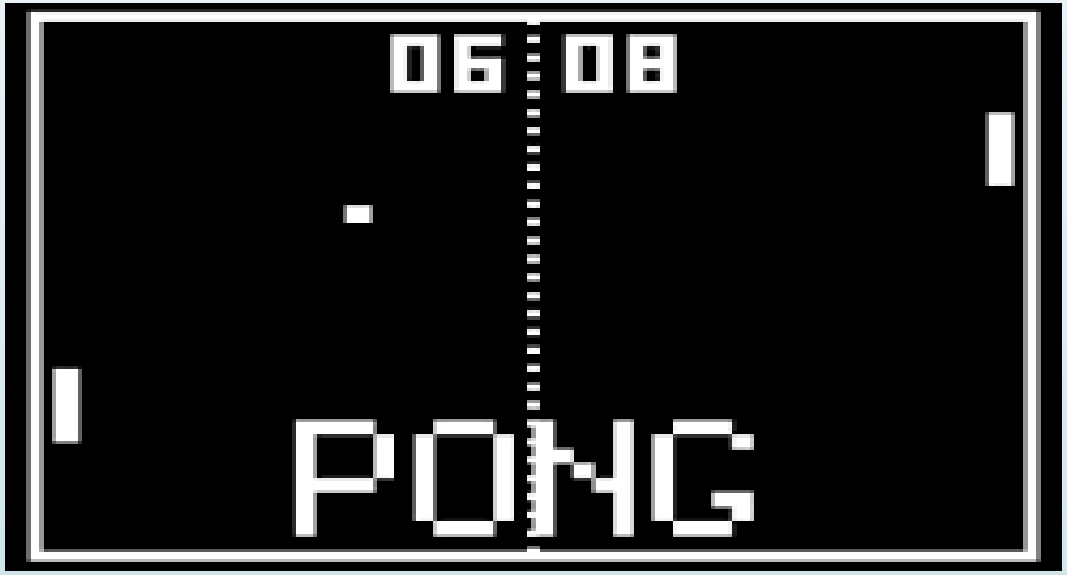


When the game is running, each Game State of the server is calculated starting from the information - Controller Data (**CD**) and Client Game States (**GScn**) – collected from the clients.



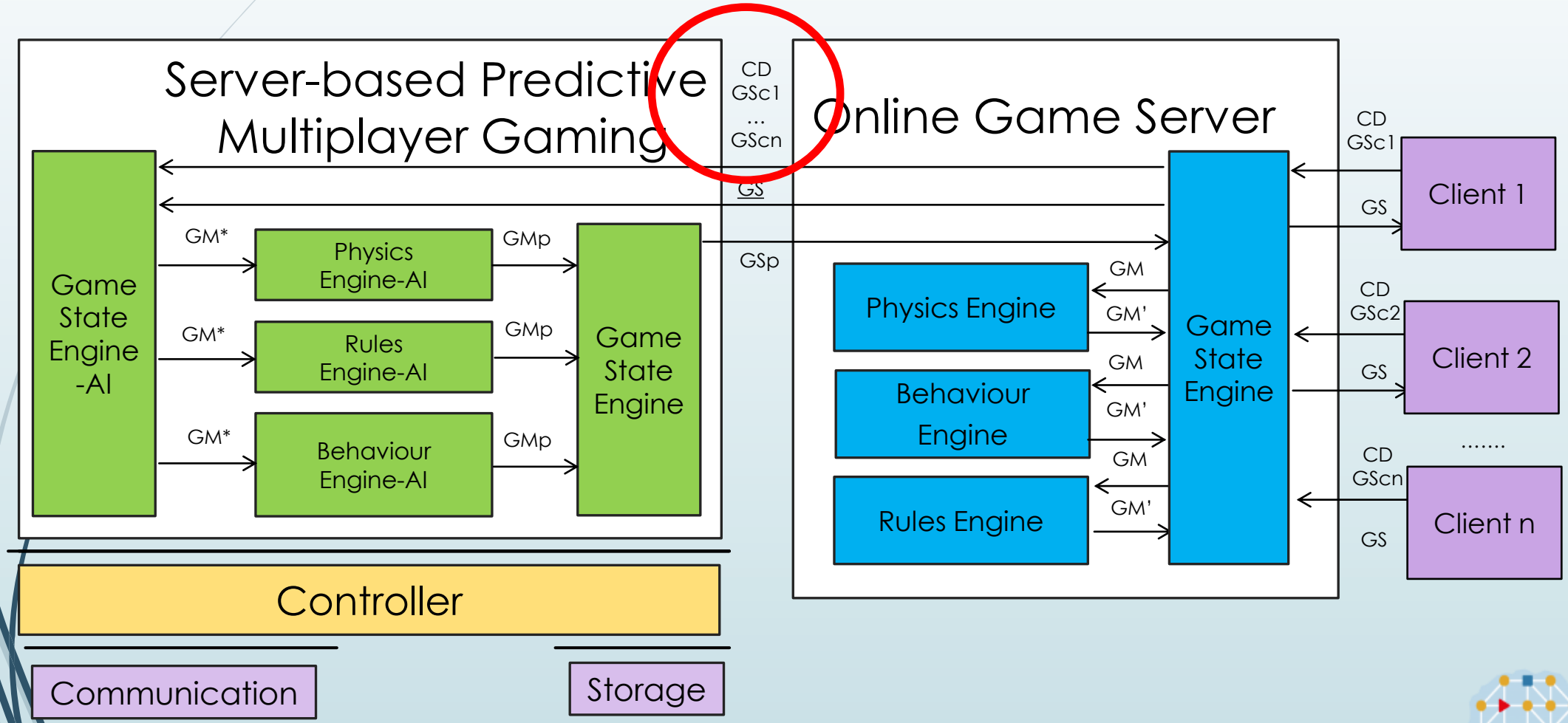
How the system works

To show how MPAI-SPG works, we are going to use a prototype of PONG, designed in a server authoritative environment.



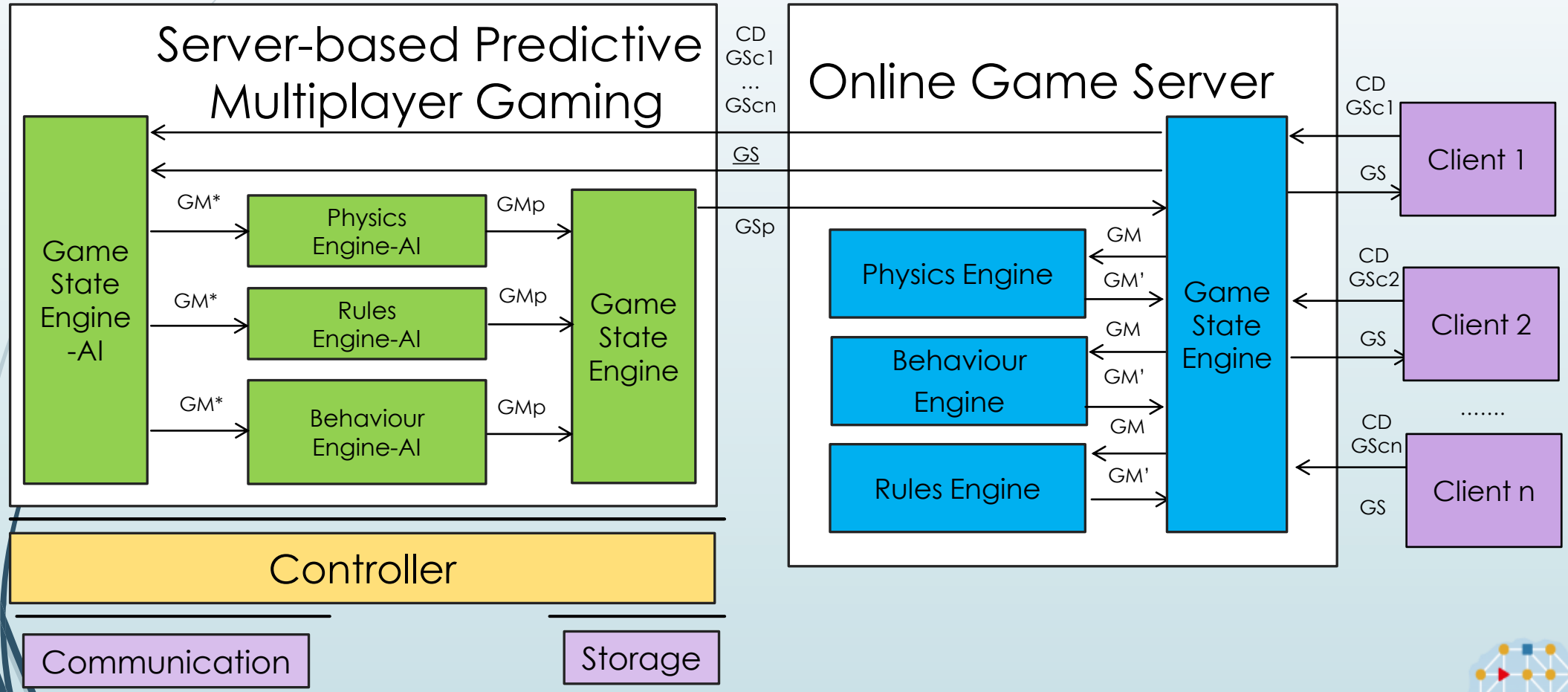
How the system works – packet loss

In case of packet loss, recurrent neural networks complete the missing information and send the complete Predicted Game State (GSp) to the Online Game Server to continue the game.



How the system works – cheating

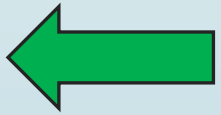
In case of cheating, the behaviour is similar; in this case the information sent to recurrent neural networks is complete but the Game State predicted will be different from the Game State calculated by the Online Game Server. MPAI-SPG generates an alert. It will be forwarded to the Online Game Server that will decide how to manage this information (client alert, a temporary ban, kicking out the game client from the game).



WHAT?

MPAI-SPG will be used as added component to be configured as plug-in in a generic game engine. In the game development, each project will add MPAI-SPG and the online server instance will exchange information with MPAI-SPG in order to check information or get the missing information needed to complete a Game State.

http://spg.mpai.community/



The AI-Enhanced Video Coding (MPAI-EVC) Evidence Project

Roberto Iacoviello

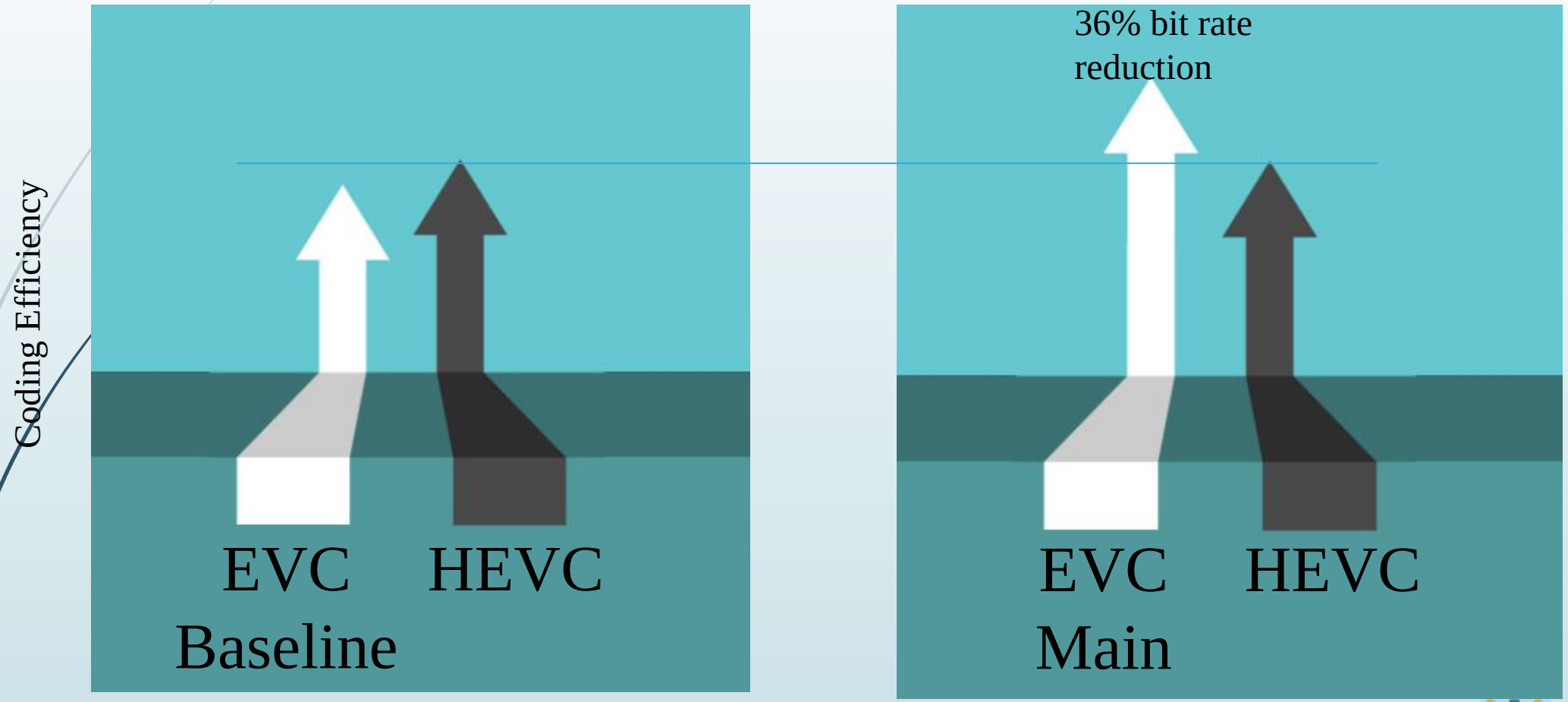
Video coding improvement by means of AI

According to Cisco, Video transmission is expected to dominate 82% of current global internet traffic in 2022

Classic block-based hybrid coding frameworks have not changed significantly for decades

The MPAI Enhanced Video Coding mission is to exploit advances in Artificial Intelligence to develop video coding standards that improve coding efficiency

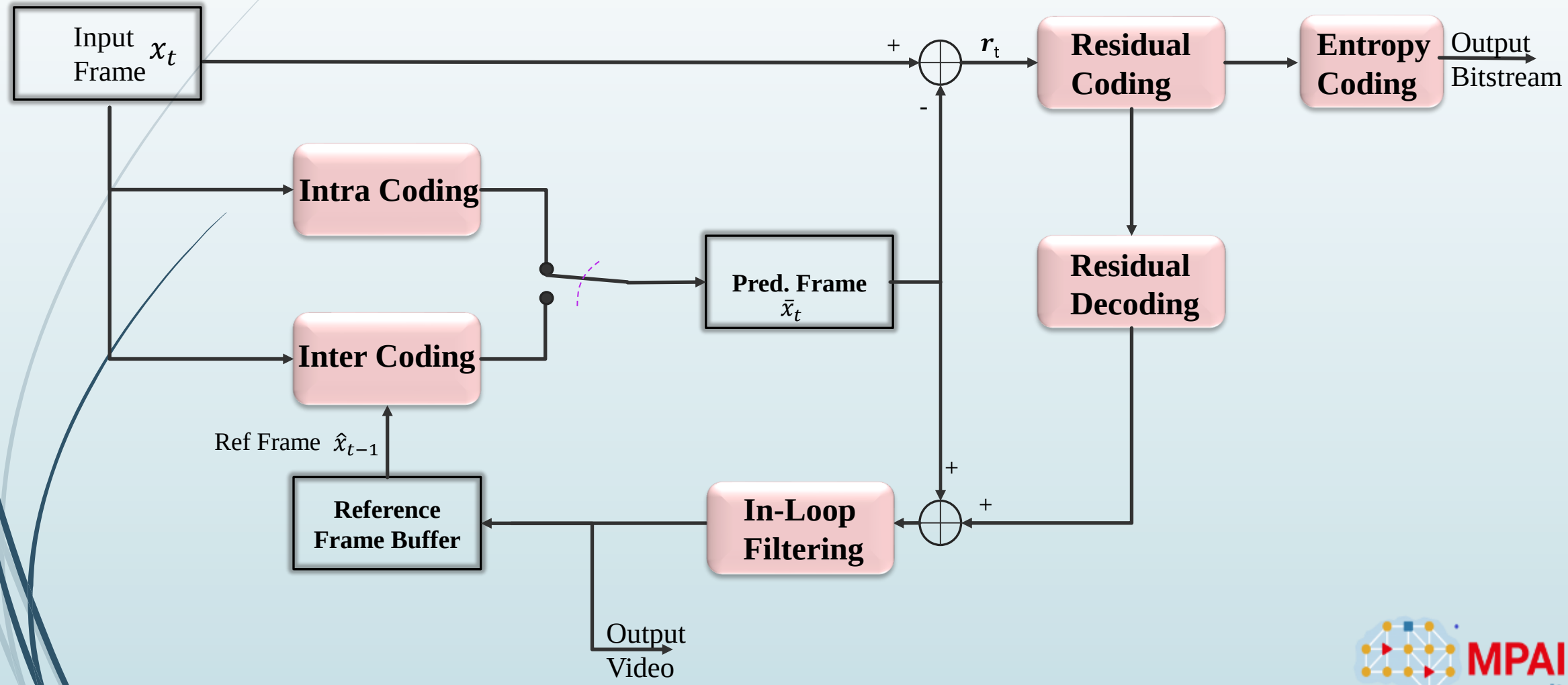
There are good video codecs: Essential Video Coding



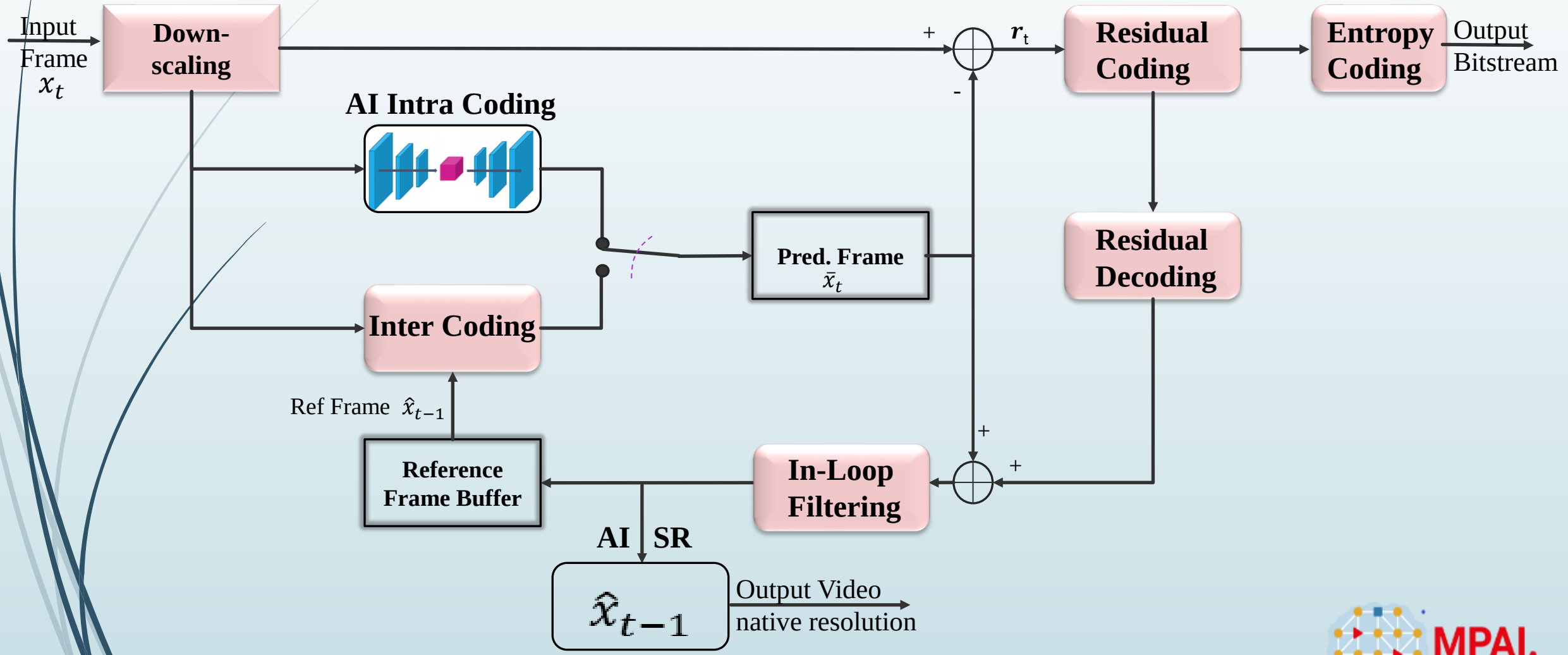
>20 years old tools

Simplified IP structure

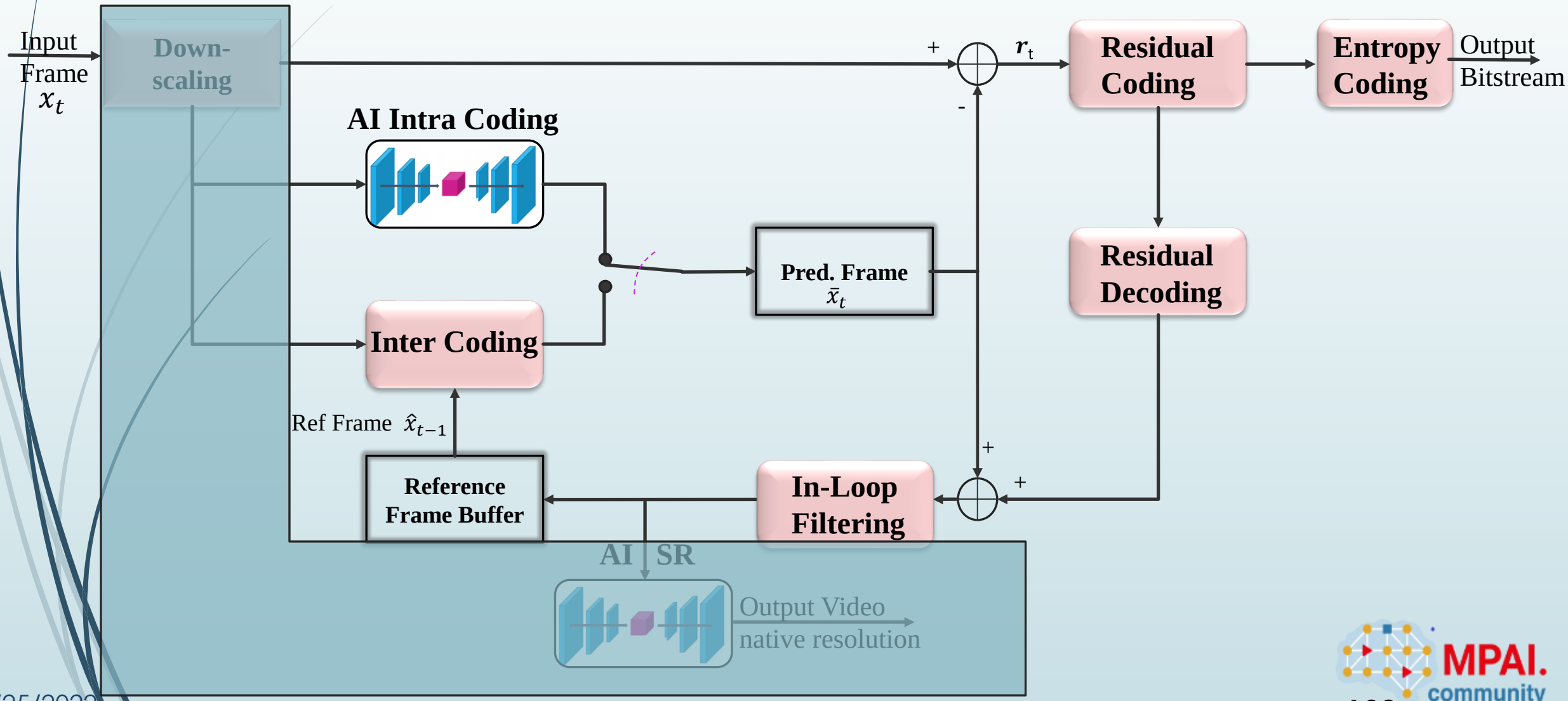
Traditional block-based hybrid framework



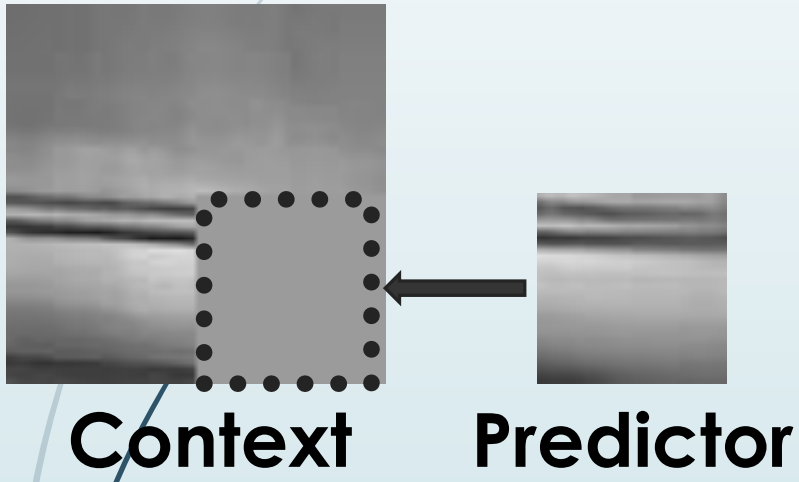
MPAI-EVC Evidence Project



AI Intra enhancement

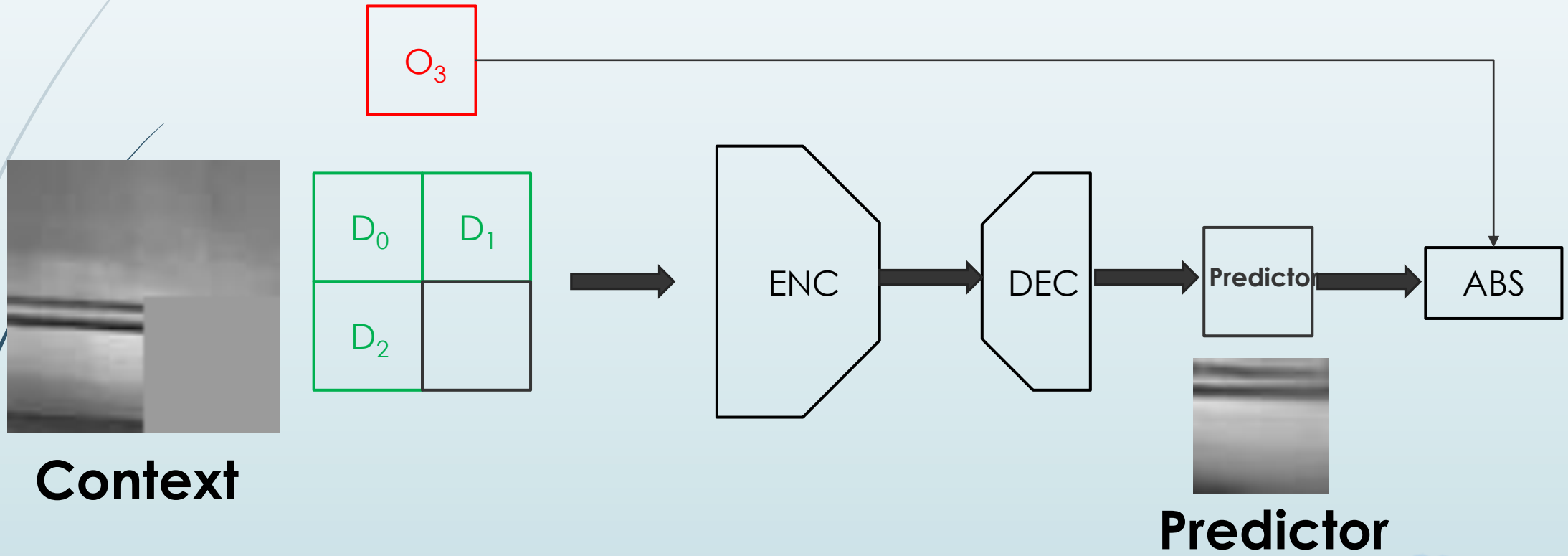


AI Intra enhancement



- For each, e.g., 32x32 CUs, 64x64 intra predictor context is sent to autoencoder
- The autoencoder returns a 32x32 prediction
- The autoencoder output always replaces the DC predictor
- The bitstream can be decoded

Reference schema



Results with Quantisation Parameter QP [22-47]

MPEG ClassB FullHD

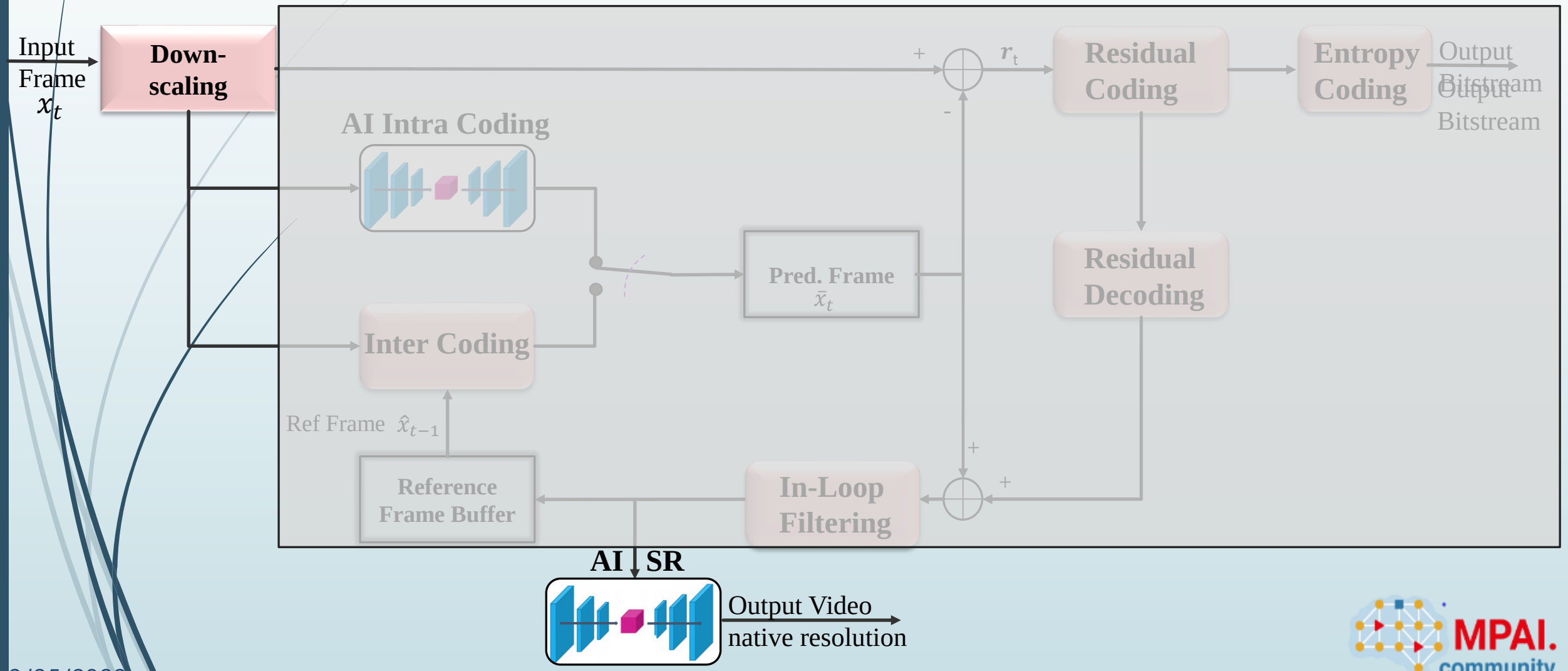
Sequence	BDRate [%]
BasketballDr	-8,90
BQTerrace	-4.90
Cactus	-5.46
Kimono1	-3.88
ParkScene	-4.17
AVG	**-5.57**

Xiph Netflix 4K

Sequence	BDRate [%]
GTAV	-5.58
MINEDRAFT	-1.12
BlueSky	-2.45
InToTree	-2.59
RedKayak	-4.86
Sunflower	-3.04
AVG	**-3.04**

Minimizing ABS, 32x32+16x16+8x8+4x4, 500 epochs

Super-Resolution



Super-resolution results

Sequence	BDRate [%]
Rome_1	0.1902
Rome_2	-18.8094
Talk_show	-21.7534
Rush_hours	4.9017
Diego_and_the_owl	8.1107
Crowd_run	-1.2430
Parkjoy	-4.2977
AVG	**-4.701**

QPs 15,30,37 and 45

Future Plans and Perspectives

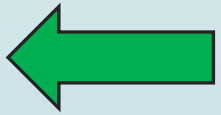
Filtering:

- ✓ reduce blockiness by filtering out some high frequencies caused by coded blocks

Inter prediction:

- ✓ estimate motion using Deep Learning architectures to refine the quality of inter-predicted blocks
- ✓ introduce new inter prediction mode to predict a frame avoiding the use of side information

http://evc.mpai.community/



2/25/2022

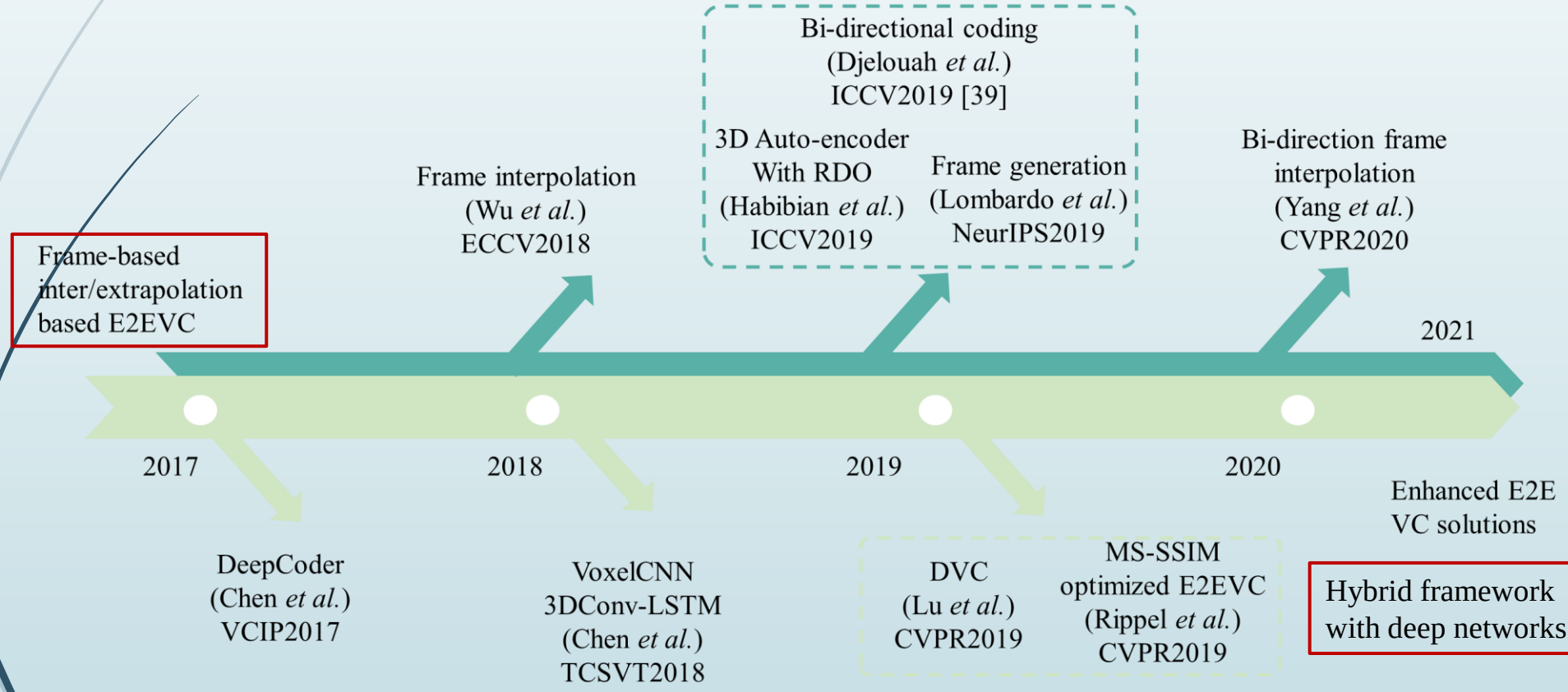
139

The AI-based End-to-End Video Coding (MPAI-EEV) project

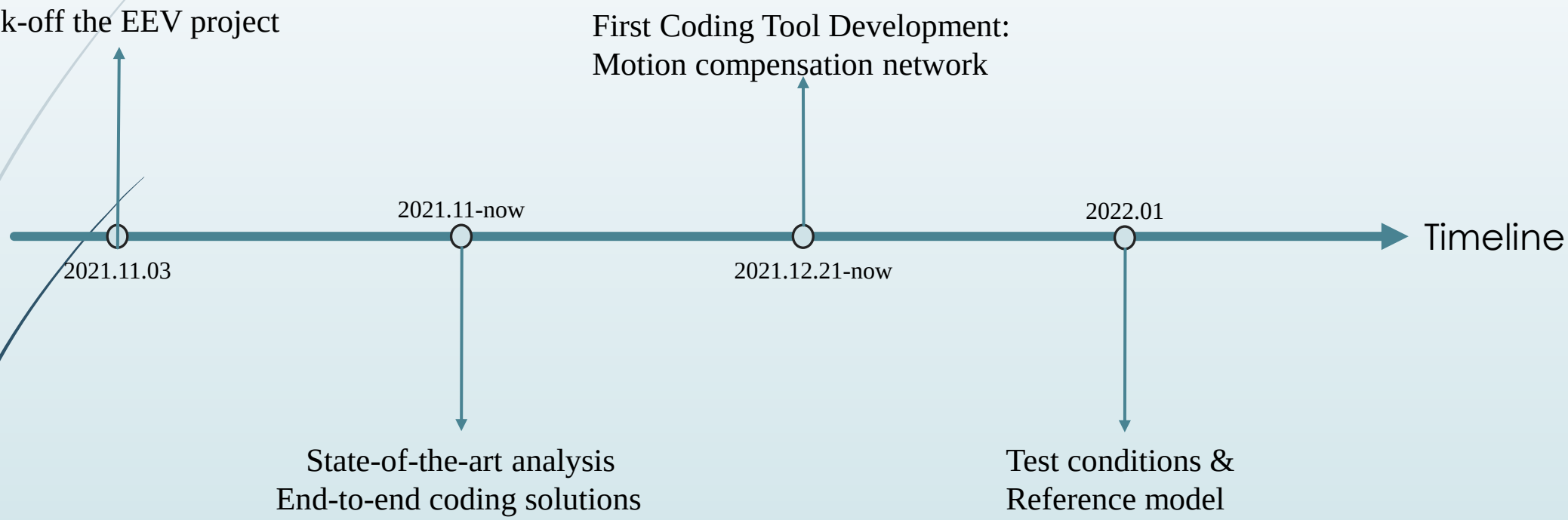
Chuanmin Jia

Video Coding Schemes using Fully Neural Models

- Frame generation-based End-to-end video coding (EEV)
  - Neural intra codec plus frame interpolation
- Hybrid framework EEV
  - Prediction plus transform-based codec using fully neural networks



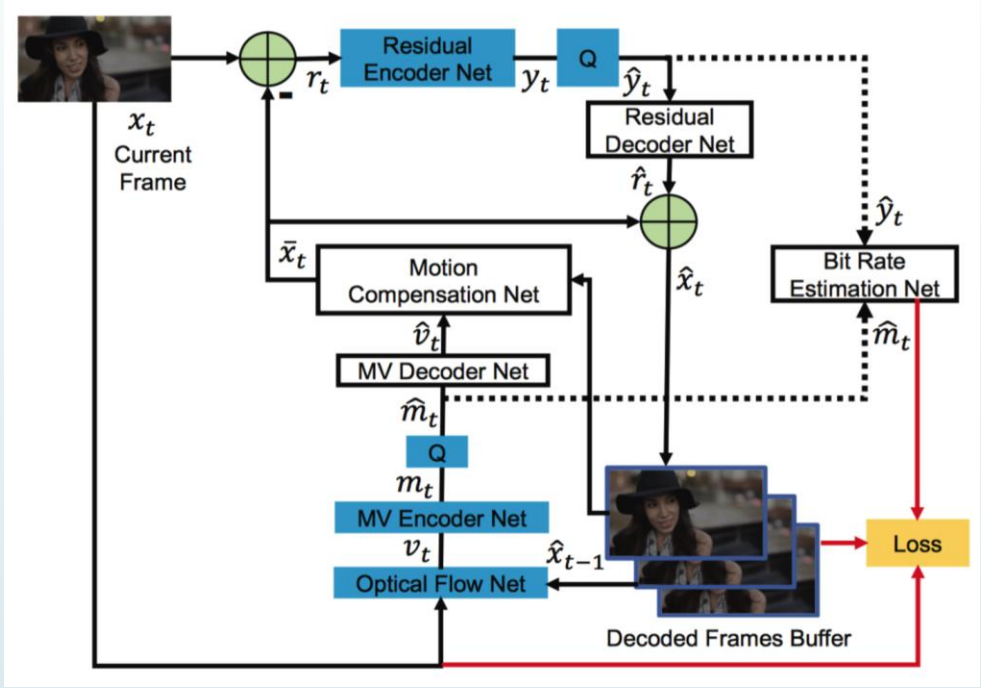
MPAI EEV Milestones



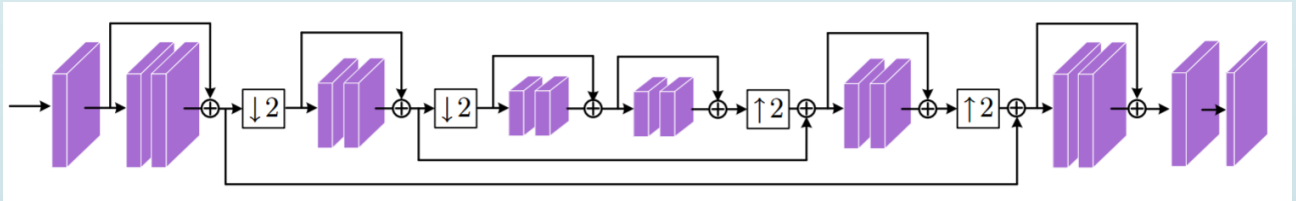
AI-based End-to-End Video Coding & OpenDVC

Reference model of EEV: OpenDVC

- Neural Motion Estimation
  - Optical Flow (motion field)
- Neural Motion Compensation
  - Predicted Frame
- Residual Compression
  - Autoencoder
- Bit-rate estimation
  - Motion field
  - Residual coding



Framework

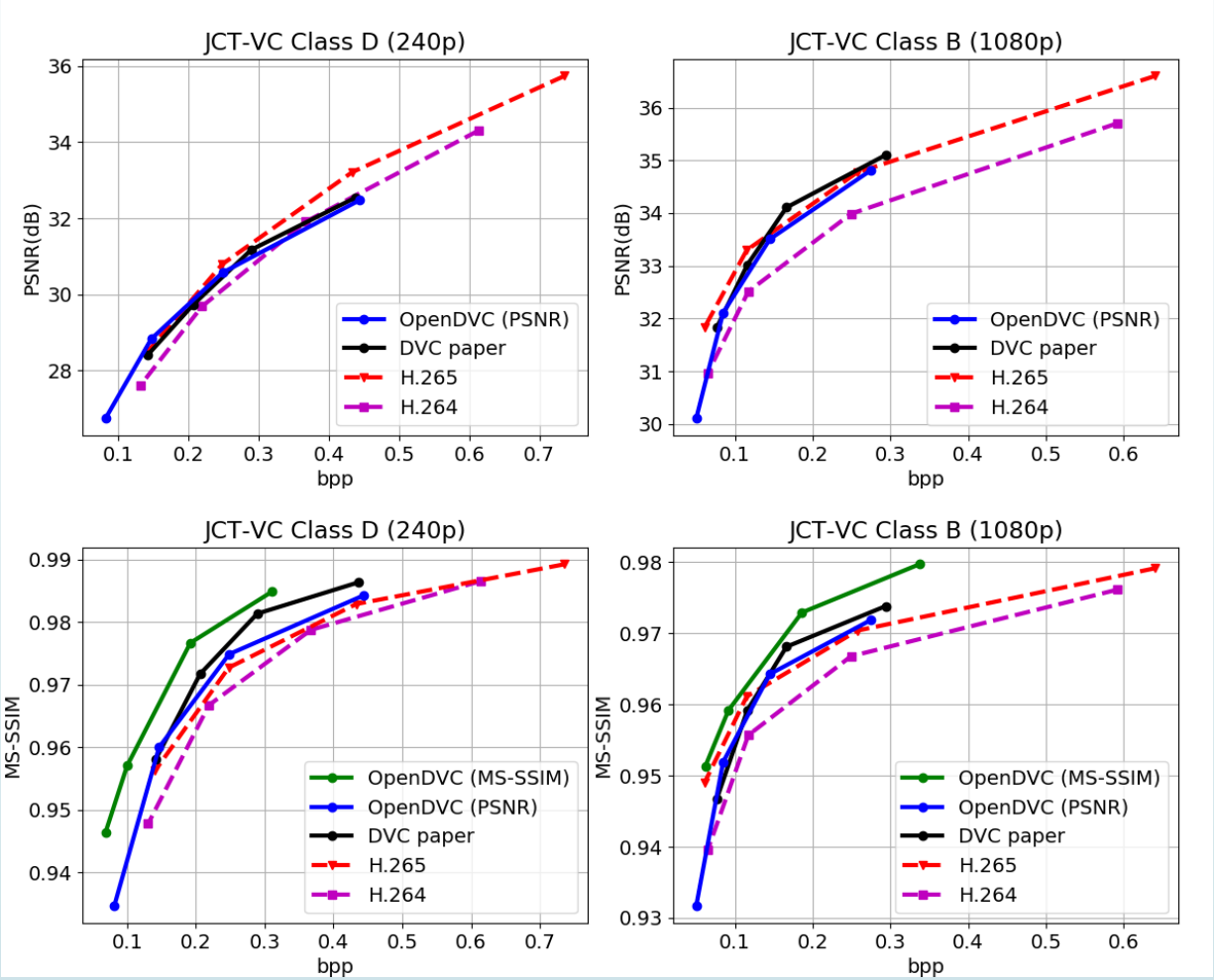


Motion Estimation Net

OpenDVC: An Open-Source Implementation of the DVC Video Compression Method, https://arxiv.org/abs/2006.15862

Open DVC Performance

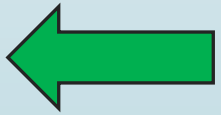
- Test data
  - JCT-VC sequences
- Performance of OpenDVC
  - Better than x264 and x265



OpenDVC: An Open Source Implementation of the DVC Video Compression Method, https://arxiv.org/abs/2006.15862

Future Plan of EEV

- Collaborative process
  - Common test conditions for MPAI-EEV
  - Training data
  - Test set
- Novel coding tools for EEV
  - Motion compensation network using non-local attention networks
- Reference model development



MPAI. community

The Connected Autonomous Vehicles (MPAI-CAV) project

Gianluca Torta

What is a Connected Autonomous Vehicle?

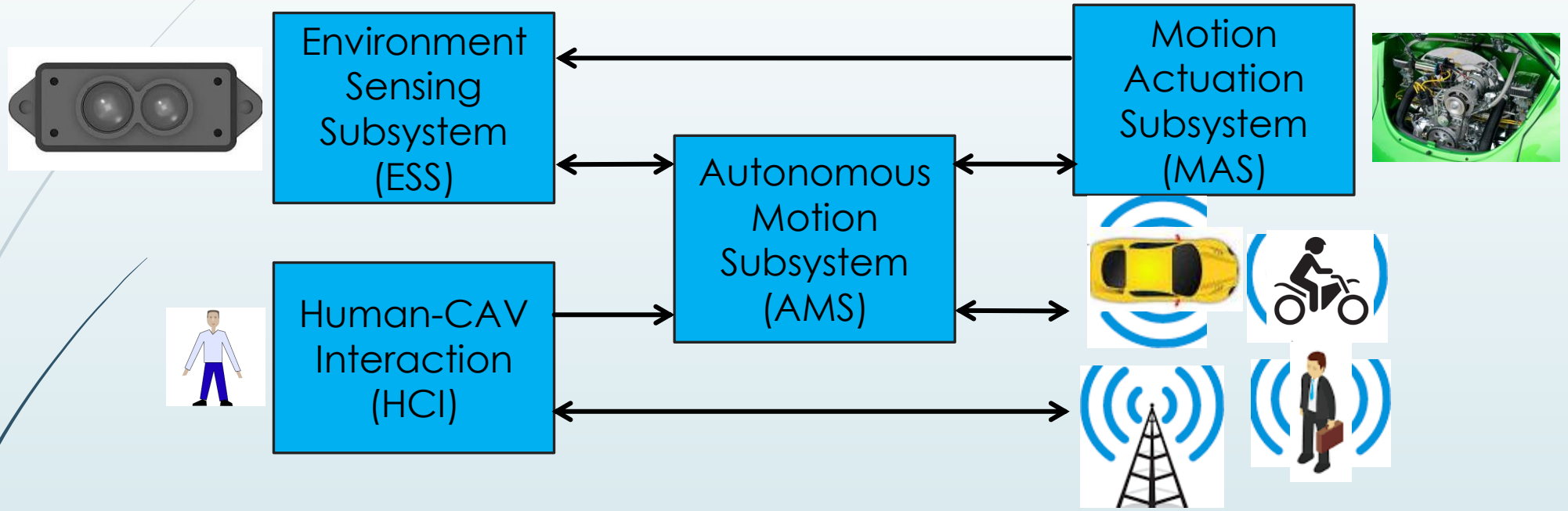
A CAV is a vehicle capable of autonomously reaching a location by:

- Understanding human utterances.
- Planning a route.
- Sensing and interpreting the environment.
- Exchanging information with other CAVs.
- Acting on the CAV's motion subsystem.

A standard for Connected Autonomous Vehicles' IT components

- Why
  - Many disparate technologies in a CAV.
  - The size of the future CAV market.
  - Users and regulators want to know if a CAV is safe, reliable and explainable.
- MPAI is the right body for a CAV standard
  - A CAV standard: multidisciplinary and AI-informed.
  - MPAI: works on data coding for all fields of endeavour.
  - MPAI standards are component- and interface-based.

The 4 subsystems of a CAV



Subsystem name	Acr.	Function
Human-CAV Interaction	HCI	Handles human-CAV interactions.
Environment Sensing Subsystem	ESS	Acquires Environment information via a variety of sensors.
Autonomous Motion Subsystem	AMS	Issues commands to drive the CAV to the destination.
Motion Actuation Subsystem	MAS	Provides Environment information, actuates motion commands

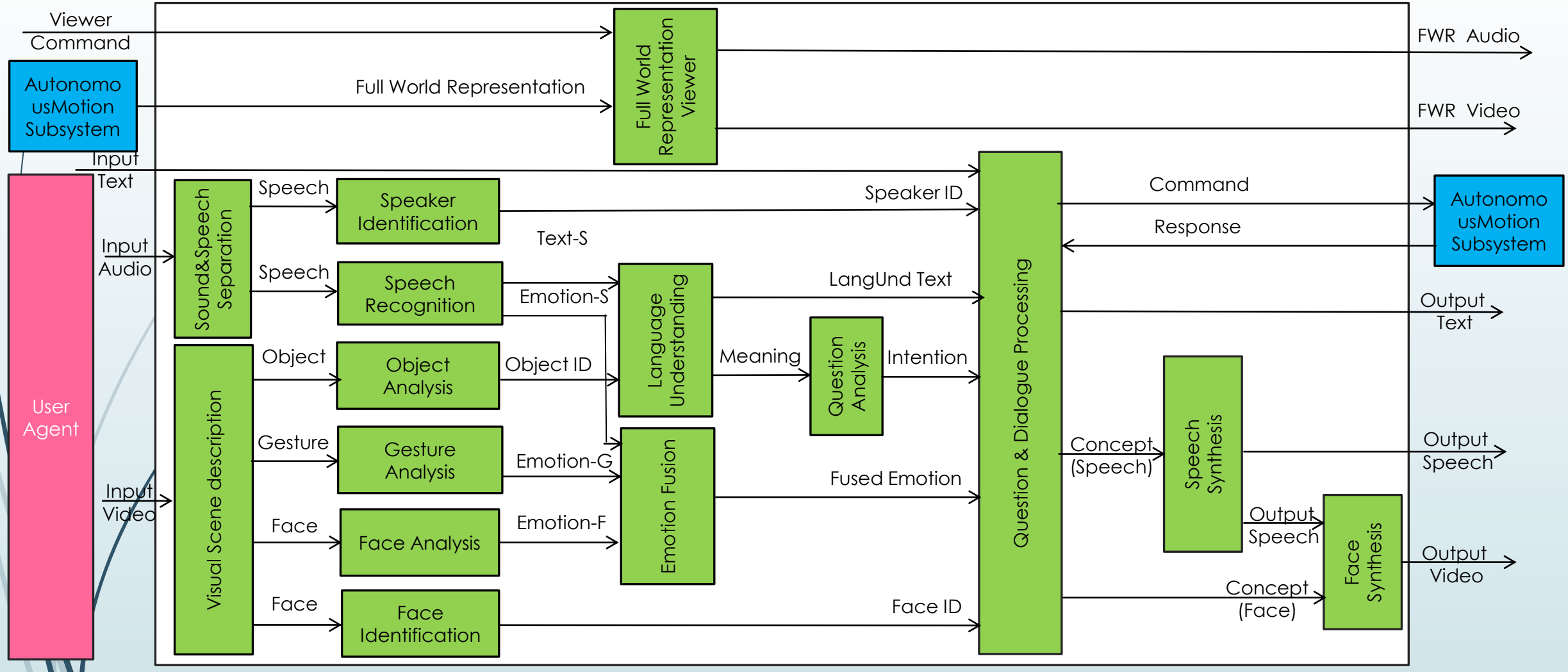
Design of Human-CAV Interaction (HCI)/1

- CAV should verify the identity of the human (out of cabin):
  - Separate speech from sound environment – recognise speaker identity.
  - Separate human shape from visual environment – recognise visual identity.
- CAV should understand what humans communicate while on board:
  - Separate speech from sound in cabin, recognise speech and its emotion.
  - Visually separate and recognise humans and objects with their location in the cabin.
  - Extract emotion from face and gestures of humans in cabin.
  - Understand questions and statements, and their associated emotions.

Design of Human-CAV Interaction (HCI)/2

- CAV should communicate to humans on board:
  - CAV should materialise as an entity with speech and face.
  - Speech should be uttered with an emotion matching humans' emotion.
  - Face should express the intended CAV "emotion".
  - Avatar's face should express same emotion and lips move in sync with speech.
  - Eyes should gaze at the human the speech is intended for.
- Humans should be able to access and navigate CAV's representation of the environment:
  - 3D visual scene
  - 3D sound scene

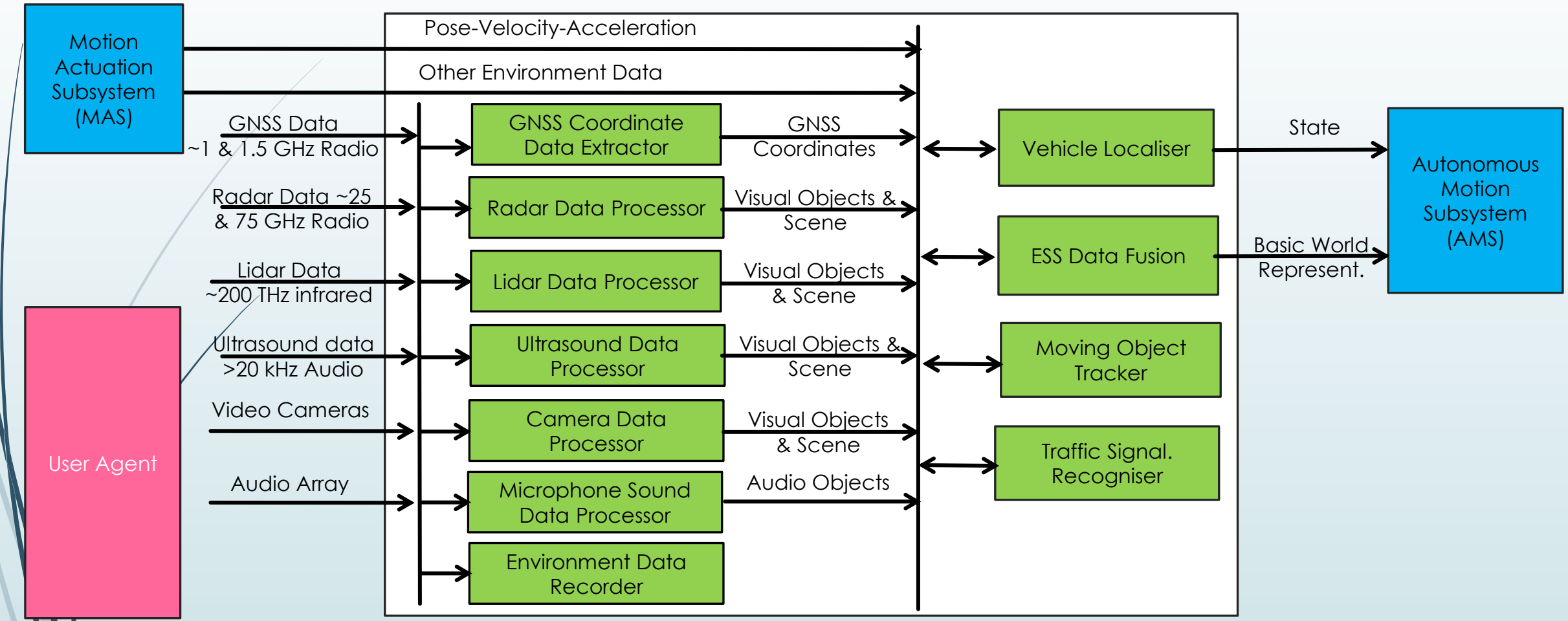
Human-CAV Interaction



Design of the Environment Sensing Subsystem (ESS)

- CAV should acquire
  - Location in Environment (pose, speed, acceleration) using:
    - Global Navigation Satellite System (GNSS).
    - Mechanical sources (odometer, speedometer, accelerometer).
    - Online maps.
  - Information about the audio and visual environment using:
    - Non-visible (Radar and Lidar) and visible range (Cameras) information.
    - Audible and non-audible (ultrasound).
  - Other environment data (temperature, humidity, etc.).
- CAV should create a (Basic) World Representation of the environment using sensed information and online maps.
- CAV should Record a selected subset of the data acquired/processed.

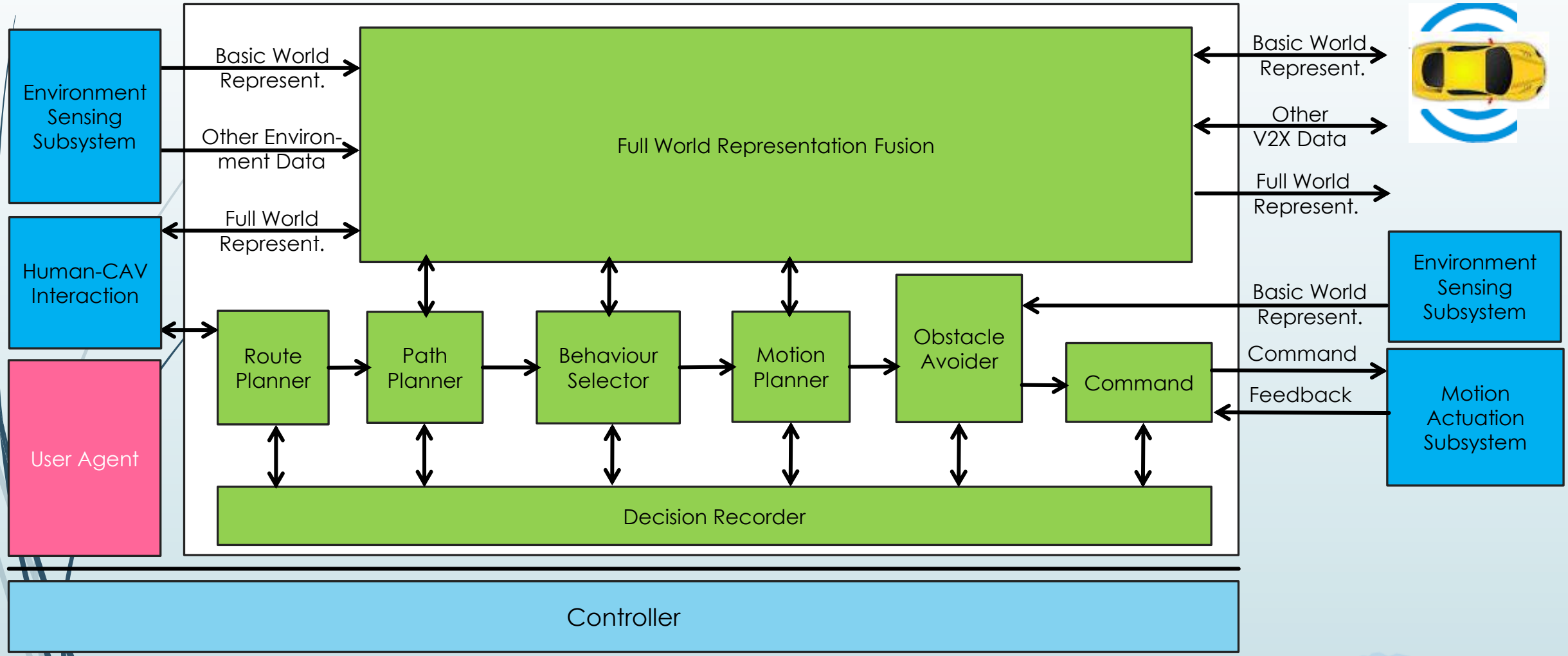
Environment Sensing Subsystem



Design of Autonomous Motion Subsystem (AMS)

- AMS should
  - Receive motion instructions from Human-CAV Interaction
  - Plan the route to the intended destination
  - Receive Basic World Representations (BWR) from
    - CAV's Environment Sensing Subsystem
    - Other CAVs in range
  - Create the Full World Representation (FWR) by fusing BWRs received
  - Issue motion command to Motion Actuation Subsystem (MAS) based on current FWR
  - Record all steps of the decision process that has led to motion command
  - Update its FWR based on feedback received from MAS

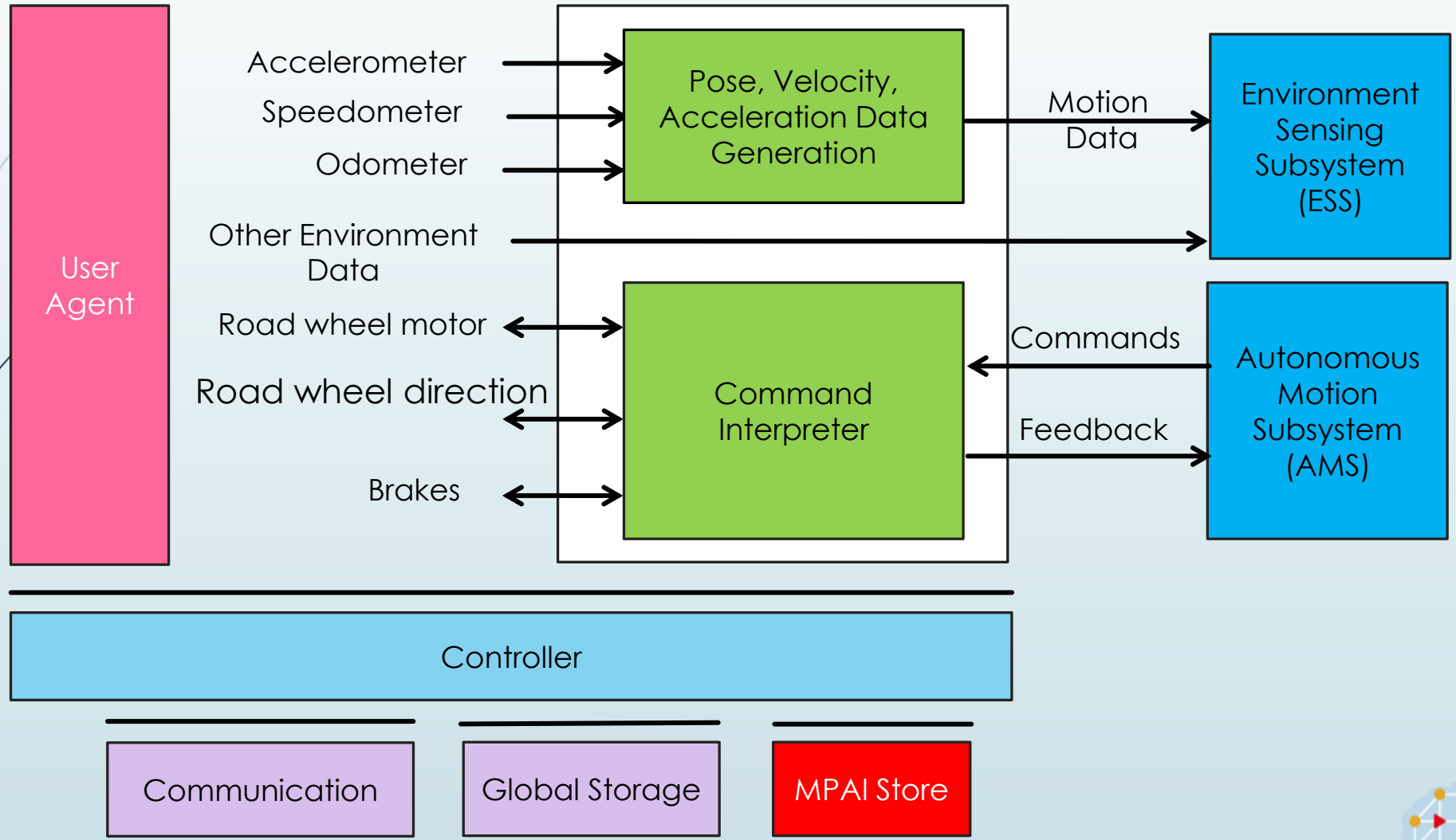
Autonomous Motion Subsystem



Design of Motion Actuation Subsystem (MAS)

- MAS should constantly
  - Send the Environment Sensing Subsystem
    - Location information (pose, speed, acceleration)
    - Other Environment Data (temperature, humidity, etc.)
  - Execute motion commands
  - Receive responses from command interpreter
  - Send feedback to Autonomous Motion Subsystem

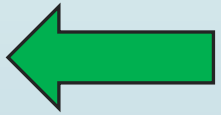
Motion Actuation Subsystem



Some CAV technologies for standardisation

- Human-CAV Interaction
  - Audio from microphone array
  - Speech and sound separation
  - 3D Audio Scene Description
  - Visual Object Separation
- Environment Sensing Subsystem
  - Basic World Representation
  - Offline Maps
  - Visual Objects and Scene from Camera, Lidar, Rada, Ultrasound and Fusion
- Autonomous Motion Subsystem
  - Command/Response
  - Full World Representation
  - Path, Pose, Route, State
  - Traffic rules and signals
  - Trajectory
- Motion Actuation Subsystem
  - Commands from/Feedback to AMS
  - Road Wheel Direction Command/Feedback
  - Road Wheel Motor Command/Feedback

http://cav.mpai.community/



The Mixed-reality Collaborative Spaces (MPAI-MCS) project

David Schultens

MCS – what it is

- Mixed-reality Collaborative Space is a virtual environment where virtual humans, twins of physical humans, are directed by their participating humans to achieve an agreed goal, for example:
  - Holding a meeting.
  - Attending a lecture.
  - Visiting a space that is fictitious or replicates a real space.
- Physical humans are embodied in speaking avatars having a specified degree of similarity – in terms of voice and appearance – with their human twins.
- An MCS is typically implemented by a set of clients connected to a server.

Avatar-based videoconference use case/1

- A group of geographically dispersed participants holds a videoconference in a virtual room equipped with table and chairs for avatars to sit.
- An avatar
  - Utters the natural voice of the participants it represents.
  - Shows a face reproducing the participant's face with the emotions displayed on it.
  - Moves its head in sync with the participant's head movements.
  - Makes gestures in sync with the participant's gestures.
  - Shows its torso and head only.

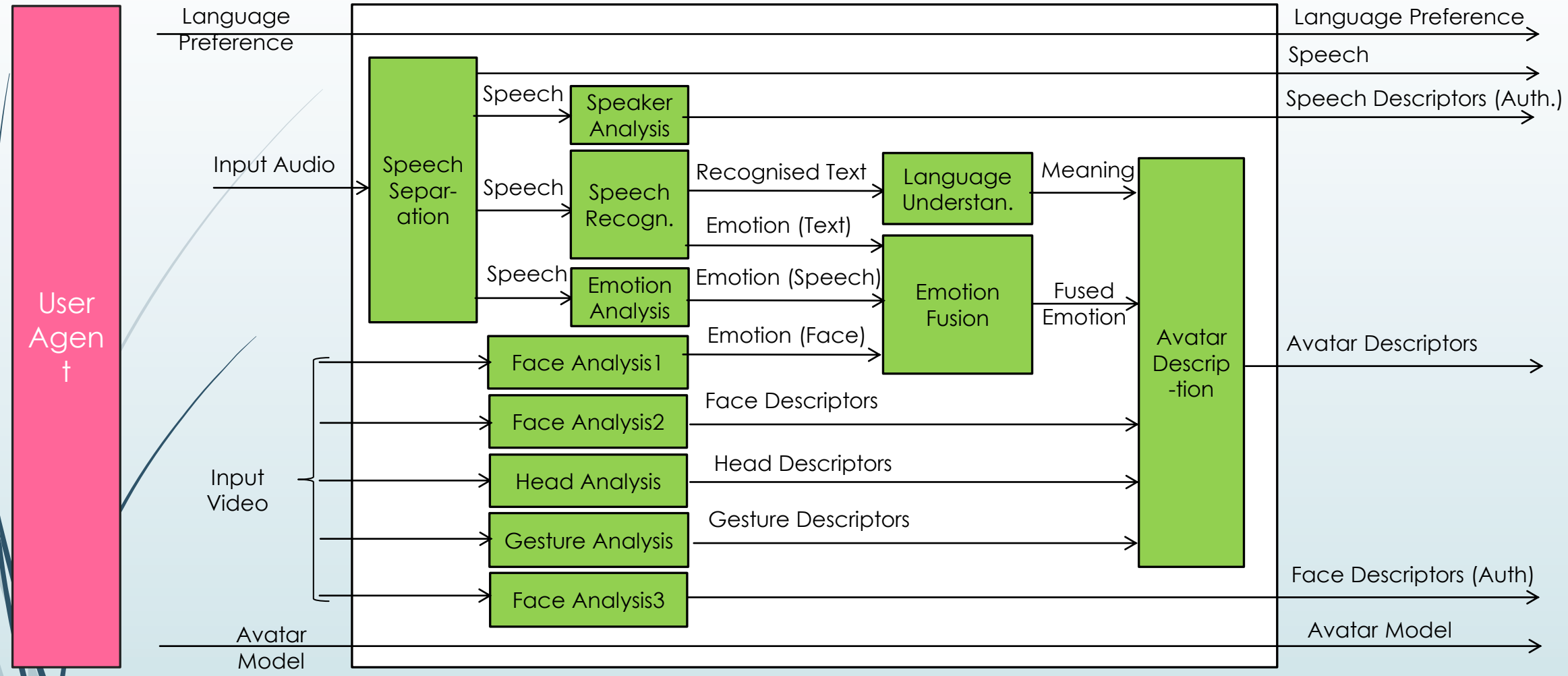
Avatar-based videoconference use case/2

- The server
  - Provides the room
  - Authenticates the avatars
  - Translates participants' speech to the selected languages.
- Avatars do not walk around, they move by teleportation.
- Participants can share and navigate audio-visual objects in the room.
- Clients receive all audio and visual objects to create their scene.
- Participants can select the viewpoint of the scene created.

Designing the transmitting client

- The participant's speech should be separated from the environment's sound
- Participant should be identification by speech and face
- Descriptor extraction
  - Emotion and meaning should be extracted from speech
  - Emotion should be extracted from face
  - Descriptors should be extracted from face, head and gesture
- Face-head-gesture descriptors should be used to create the avatar's descriptors
- Meaning of participant's utterance and fused emotion may be used to refine the avatar's appearance
- Speech, language preference and avatar model are transmitted to the server.

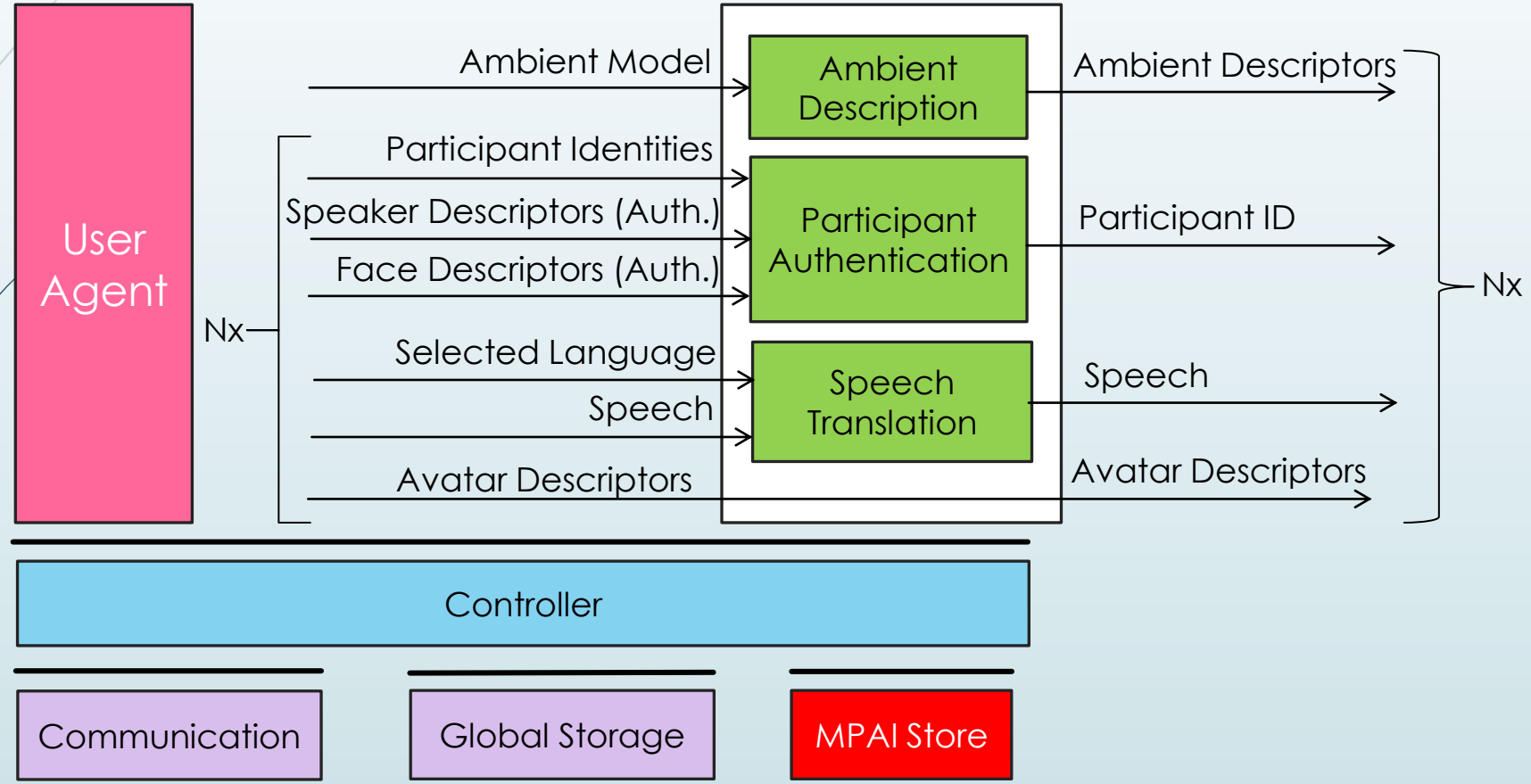
An MCS Client (transmitter)



Designing an MCS server

- The server should select a suitable room template
- Participants should be identified through speech and face
- Participants' speech should be translated as requested by participants
- Avatar descriptors, audio-visual objects and their actions should be forwarded to the clients without modification.

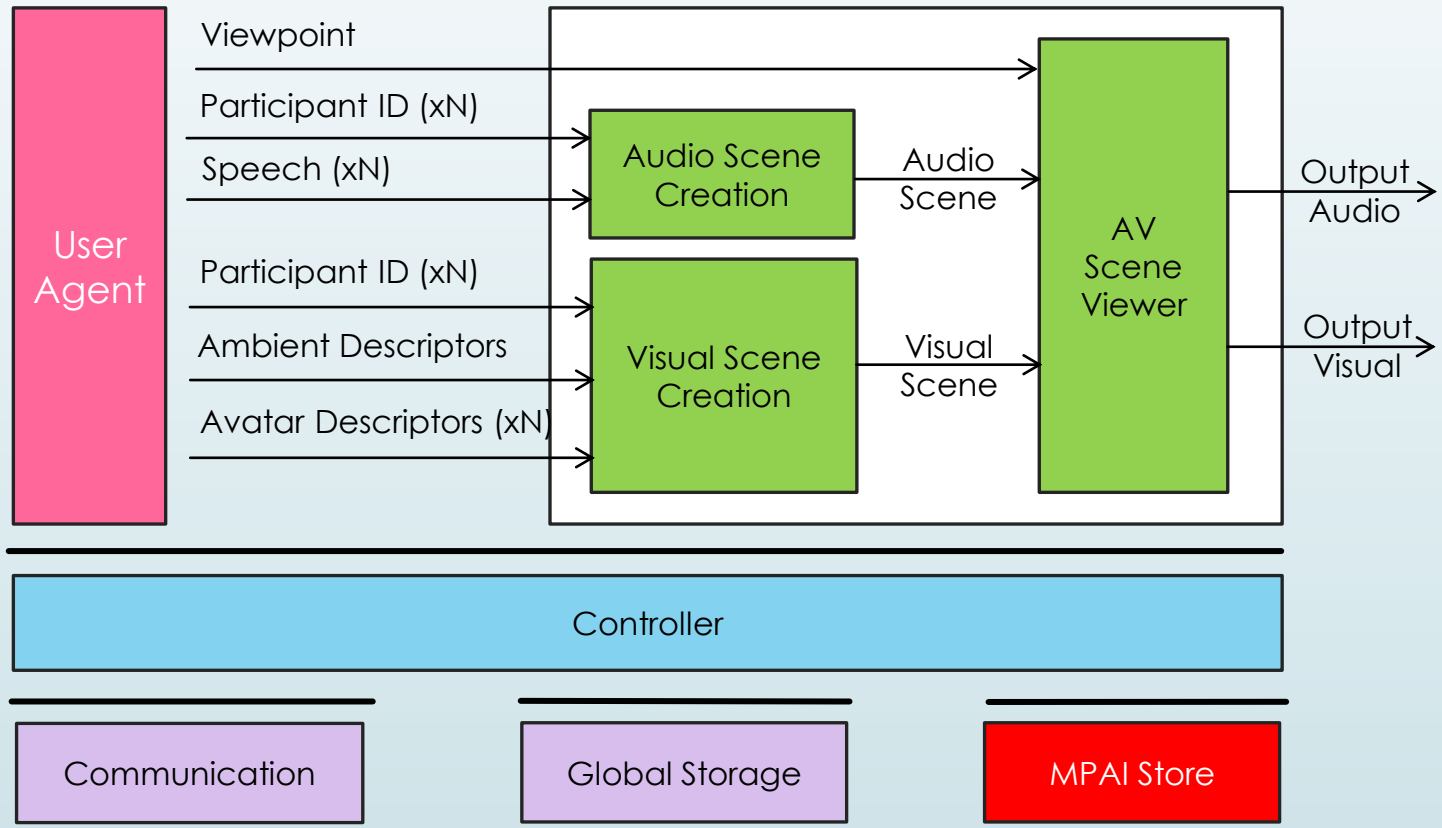
An MCS server



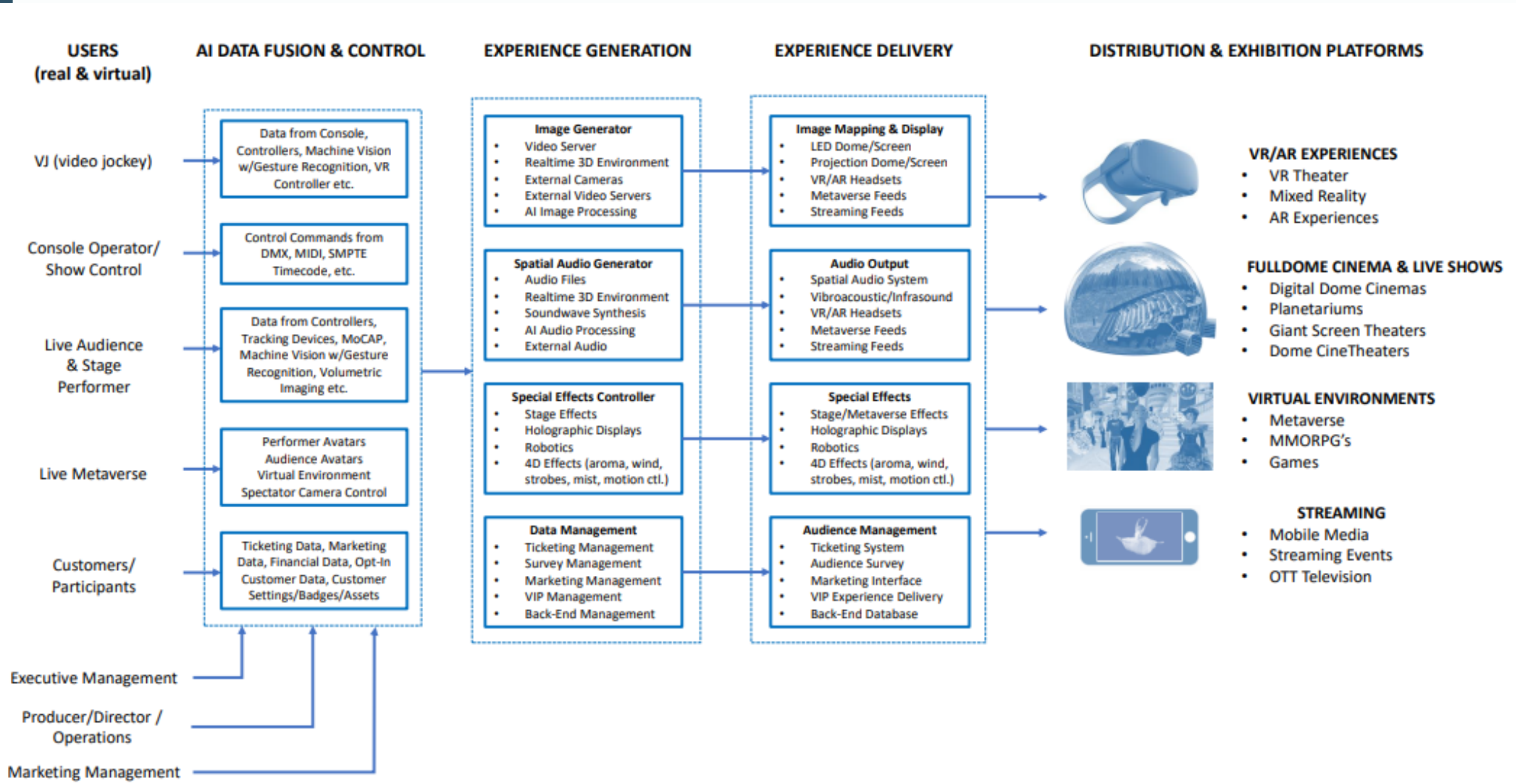
Designing the receiving client

- The following should be used to create the visual component of the scene in the form desired by a participant
  - The room
  - The avatars and avatar models
  - The visual component of the audio-visual object as modified by a participant's action.
- The following should be used to create the audio component of the scene in the form desired by a participant
  - The participants' speech positioned at the avatar location.
  - The audio component of the audio-visual object as modified by a participant's action.
- The participant should select a viewpoint.

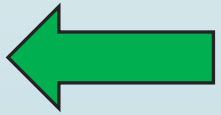
The MCS Client (receiver)



XR Theater Use Case



http://mcs.mpai.community/



2/25/2022

Conclusion of “MPAI talks to industry” presentations

Leonardo Chiariglione

MPAI in a nutshell…

- Need for **efficient**, especially **AI-based** data coding standards for the **AI age**.
- Robust **standards development process** tested during the development of its first 5 standards.
- Process designed to facilitate **timely access** to essential IP at **anticipated terms and conditions**.
- **Component- and interface**-based standards.
- **Root of trust** of the **governed ecosystem** brought about by MPAI standards.

Many are already on board…

If you wish to be part of the MPAI adventure, you should…

Be

- A legal entity or an individual representing a university technical department
  - Supporting the MPAI mission and
  - Able to contribute to MPAI standard development

Choose

- One of the two classes of membership:
  - Principal Members have right to vote (2400 € p.a.)
  - Associate Members participate in standard development (480 € p.a.)

Send

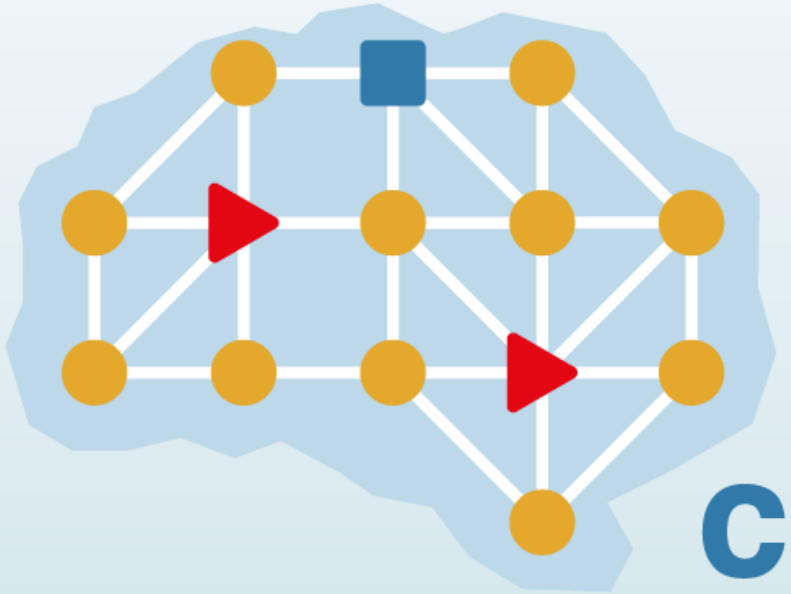
- A copy of:
  - The filled-out template for MPAI Membership applications.
  - The signed MPAI Statutes.
  - The bank transfer receipt.

  to secretariat@mpai.community

For more info visit

- mpai.community/how-to-join/join/

Join the fun, build the future



MPAI.
community

https://www.mpai.community/