|  |  |
| --- | --- |
|  | Moving Picture, Audio and Data Coding by Artificial Intelligence  www.mpai.community |

|  |  |
| --- | --- |
|  | **Public Document** |
| N46 | 2020/11/18 |
| Source | Andrea Basso, Leonardo Chiariglione, Miran Choi, Fabrizio Davide, Thierry Fautier, Michelangelo Guarise, Steven Hawley, Michel Hospital, Guy Paillet, Roberto Iacoviello, Valeria Lazzaroli, Ajay Luthra, Massimiliano Lunghi, Marco Mazzaglia, Gary Olson, Danilo Pau, Paolo Ribeca, Mariangela Rosano, Luigi Troiano |
| Title | MPAI Use Cases |
| Target | For public release |

Contents

[1 Introduction 3](#_Toc56235217)

[2 Still Pictures 5](#_Toc56235218)

[2.1 Media & Entertainment 5](#_Toc56235219)

[2.1.1 ME.SP-01: Multires pictures 5](#_Toc56235220)

[2.2 Transportation 6](#_Toc56235221)

[2.3 Telco 6](#_Toc56235222)

[2.4 Information Technology 6](#_Toc56235223)

[2.5 Aerospace 6](#_Toc56235224)

[2.5.1 AS.SP-01: Large-scale, Multi-spectrum Satellite Imagery 6](#_Toc56235225)

[2.6 Manufacturing 7](#_Toc56235226)

[2.6.1 MF.SP-01: Quality Inspection 7](#_Toc56235227)

[2.7 Healthcare 8](#_Toc56235228)

[2.7.1 HC.SP-01: Radiology Imaging 8](#_Toc56235229)

[2.7.2 HC.SP-02: Intravenous bottle 8](#_Toc56235230)

[2.7.3 HC.SP-03: Reduced data acquisition in MRI scans 9](#_Toc56235231)

[2.7.4 HC.SP-04: Correlative analysis in support of infectious disease research 10](#_Toc56235232)

[2.8 Food & Beverage 11](#_Toc56235233)

[2.8.1 FB.SP-01 - Food information 11](#_Toc56235234)

[2.9 Science & Technology 12](#_Toc56235235)

[2.9.1 ST.SP-01: Images from space 12](#_Toc56235236)

[2.10 Other Domains 13](#_Toc56235237)

[3 Moving Pictures 13](#_Toc56235238)

[3.1 Media & Entertainment 13](#_Toc56235239)

[3.1.1 MP.ME-04: Game content distribution 13](#_Toc56235240)

[3.1.2 ME.MP-05: AI-enhanced traditional video coding 14](#_Toc56235241)

[3.1.3 ME.MP-06: Fully AI-based video coding 16](#_Toc56235242)

[3.1.4 ME.MP-07: AI based optimisation of existing or new codecs 17](#_Toc56235243)

[3.1.5 ME.MP-08 AI compression of high-quality video and images 18](#_Toc56235244)

[3.1.6 ME.MP-09: Distributed multiplayer online gaming in Next Generation Games 19](#_Toc56235245)

[3.1.7 ME.MP-10: AI- pan&scan 8K 21](#_Toc56235246)

[3.1.8 ME.MP-11: AI-services for next generation TV 22](#_Toc56235247)

[3.1.9 ME.MP-12: Tracking game player’s movements 23](#_Toc56235248)

[3.1.10 ME.MP-13: Visual Object and Scene Description 24](#_Toc56235249)

[3.2 Transportation 26](#_Toc56235250)

[3.2.1 TP.MP-01: AI-assisted driving 26](#_Toc56235251)

[3.2.2 TP.MP-02: Highspeed Train Safety 30](#_Toc56235252)

[3.3 Telco 31](#_Toc56235253)

[3.3.1 TC.MP-01: Videoconferencing 31](#_Toc56235254)

[3.4 Information Technology 32](#_Toc56235255)

[3.4.1 IT.MP-01: Source verification 32](#_Toc56235256)

[3.5 Aerospace 33](#_Toc56235257)

[3.6 Manufacturing 33](#_Toc56235258)

[3.7 Healthcare 33](#_Toc56235259)

[3.7.1 HC.MP-02: Correct Pose 33](#_Toc56235260)

[3.8 Food & Beverage 34](#_Toc56235261)

[3.8.1 FB.MP-01: Checking the health of vines 34](#_Toc56235262)

[3.9 Science & Technology 35](#_Toc56235263)

[3.10 Other Domains 35](#_Toc56235264)

[4 Audio 35](#_Toc56235265)

[4.1 Media & Entertainment 35](#_Toc56235266)

[4.1.1 ME.AU-01: Purpose-dependent Audio Reproduction 35](#_Toc56235267)

[4.1.2 ME.AU-02: Context-based Audio Enhancement (MPAI-CAE) 36](#_Toc56235268)

[4.1.3 ME.AU-03: Context-aware audio experience enrichment 40](#_Toc56235269)

[4.1.4 ME.AU-04: Normalization of volume across channels/streams 41](#_Toc56235270)

[4.2 Transportation 42](#_Toc56235271)

[4.2.1 TP.AU.01: Personal communications at high-speed 42](#_Toc56235272)

[4.2.2 TP.AU.02: Traffic sound 43](#_Toc56235273)

[4.3 Telco 43](#_Toc56235274)

[4.3.1 TC.AU-01: High-fidelity voice 43](#_Toc56235275)

[4.4 Information Technology 44](#_Toc56235276)

[4.5 Aerospace 44](#_Toc56235277)

[4.6 Manufacturing 44](#_Toc56235278)

[4.7 Healthcare 44](#_Toc56235279)

[4.7.1 HC.AU-01: 3D Audio 44](#_Toc56235280)

[4.8 Food & Beverage 45](#_Toc56235281)

[4.9 Science & Technology 45](#_Toc56235282)

[4.10 Other Domains 45](#_Toc56235283)

[5 Audio and Video 45](#_Toc56235284)

[5.1 Media & Entertainment 45](#_Toc56235285)

[5.2 Transportation 45](#_Toc56235286)

[5.3 Telco 46](#_Toc56235287)

[5.4 Information Technology 46](#_Toc56235288)

[5.5 Aerospace 46](#_Toc56235289)

[5.6 Manufacturing 46](#_Toc56235290)

[5.7 Healthcare 46](#_Toc56235291)

[5.7.1 HC.AV-01: Vision-to-Sound Transformation 46](#_Toc56235292)

[5.8 Food & Beverage 47](#_Toc56235293)

[5.9 Science & Technology 47](#_Toc56235294)

[5.10 Other Domains 47](#_Toc56235295)

[6 Event sequence 47](#_Toc56235296)

[6.1 Media & Entertainment 47](#_Toc56235297)

[6.2 Transportation 47](#_Toc56235298)

[6.2.1 TP.ES-01: Predictive Vehicle Maintenance 47](#_Toc56235299)

[6.3 Telco 48](#_Toc56235300)

[6.4 Information Technology 48](#_Toc56235301)

[6.5 Aerospace 48](#_Toc56235302)

[6.6 Manufacturing 48](#_Toc56235303)

[6.7 Healthcare 48](#_Toc56235304)

[6.8 Food & Beverage 48](#_Toc56235305)

[6.9 Science & Technology 48](#_Toc56235306)

[6.10 Other Domains 49](#_Toc56235307)

[6.10.1 ES-OD-01: Compression and Understanding of Industrial Data 49](#_Toc56235308)

[7 Other data 51](#_Toc56235309)

[7.1 Media & Entertainment 51](#_Toc56235310)

[7.1.1 OD.ME-01: Anomalous service access 51](#_Toc56235311)

[7.1.2 OD.ME-03: Multi-Modal Conversation 53](#_Toc56235312)

[7.1.3 OD.ME-04: Server-based Predictive Multiplayer Gaming in Next Generation Games 55](#_Toc56235313)

[7.2 Transportation 59](#_Toc56235314)

[7.3 Telco 59](#_Toc56235315)

[7.4 Information Technology 59](#_Toc56235316)

[7.5 Aerospace 59](#_Toc56235317)

[7.6 Manufacturing 59](#_Toc56235318)

[7.7 Healthcare 59](#_Toc56235319)

[7.8 Food & Beverage 59](#_Toc56235320)

[7.9 Science & Technology 60](#_Toc56235321)

[7.9.1 ST.OD-01: Gravitational-Wave Interferometric Data 60](#_Toc56235322)

[7.9.2 ST.OD-02: Multi-omics 60](#_Toc56235323)

[7.9.3 ST.OD-03: Spatial metabolomics 62](#_Toc56235324)

[7.9.4 ST.OD-03: Integrated -omics and image data 63](#_Toc56235325)

[7.9.5 ST.OD-06: Integrative AI-based analysis of multi-source genomic/sensor experiments 64](#_Toc56235326)

[7.10 Other Domains 70](#_Toc56235327)

[7.10.1 OD.OD-02: Machine vibrations 70](#_Toc56235328)

[7.10.2 OD.OD-03: Building vibrations 71](#_Toc56235329)

# Introduction

This document is a collection of use cases organised in a matrix whose columns are currently 5 data types for which coding is considered useful. Each chapter deals with each of the 5 data types – Still Pictures (SP), Moving Pictures (MP), Audio (AU), Event Sequence (ES) and Other Data. More data types may be spun off from Other Data as work progresses.

The rows of the matrix are currently 10 areas that collect specific use cases. introduced to help derive requirements for the AI-supported coding of data types. The areas are: Media & Entertainment (ME), Transportation (TP), Telco (TC), Information Technology (IT), Aerospace (AS), Industrial Manufacturing (IM), Healthcare (HC), Food & Beverage (FB), Science & Technology (SC) and Other Domains (OD).

Use cases

Each chapter has 10 subchapters corresponding to the 10 areas and each area has a number of section each corresponding to the use cases.

It is important to note that, in general, an MPAI standard is developed using the requirements derived from an aggregation of use cases. An MPAI use case is not necessarily due to become a candidate MPAI standard. A use case could be expanded or merged with other related use cases before it can become a candidate MPAI standard.

Table 1 depicts the data type-area matrix.

*Table 1 – The matrix of media types and areas*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Still Picture (SP) | Moving Pictures (MP) | Audio (AU) | Event Sequence (ES) | Other Data (OD) |
| Media & Entertainment (ME) |  |  |  |  |  |
| Transportation (TP) |  |  |  |  |  |
| Telco (TC) |  |  |  |  |  |
| Information Technology (IT) |  |  |  |  |  |
| Aerospace (AS) |  |  |  |  |  |
| Manifacturing (IM) |  |  |  |  |  |
| Healthcare (HC) |  |  |  |  |  |
| Food & Beverage (FB) |  |  |  |  |  |
| Science &Technology (SC) |  |  |  |  |  |
| Other Domains (OD) |  |  |  |  |  |

An MPAI use case is a component of the MPAI standard development process. Potential standards are identified in a bottom-up process.

These are the elements that are expected to appear in an MPAI use case. Some elements are mandatory and some optional. Optional elements should also be filled out to support a positive dec­ision by the General Assembly.

*Table 2 – Use case structure*

|  |  |
| --- | --- |
| **Title** | Concise title of use case |
| **Proponent** | Proponent’s name and affiliation |
| **Description** | Explains and delimits the scope of the use case |
| **Comments** | General comment on why and how AI can support the use case |
| **Examples** | Illustrate how the use case can cover different contexts, especially if the use case has a broad coverage |
| **Requirements** | Preliminary requirements to clarify the use case (full requirements identific­ation is part of the subsequent Functional Requirements stage) |
| **Object of standard** | Provides general identification of what is normative in the proposed use case if a standard will be developed |
| **Benefits** | Advantages offered by the standard over existing solutions and new oppor­tun­ities offered to industry and/or end users |
| **Bottlenecks** | Technical issues that may limit use of the standard or whose passing over will facilitate use of the standard |
| **Social aspects** | Cases where using the standard may have social impacts (optional) |
| **Success criteria** | Proposed measures of the standard’s success. These should include outcomes (short term) and impact (longer term) |

Table 3 gives lists the key features and their meaning.

*Table 3 – Feature table*

|  |  |
| --- | --- |
| Volume | The quantity of data handled (Low/Med/High), if possible, bits/s etc. |
| Maturity | The current state of the relevant technology (Research/Inception/Adoption) |
| AI relevance | How critical is using AI as opposed to other technologies (Low/Med/High)? |
| Users | Who are the intended users of this standard? |
| Participants | Entities that should take part in the development of the standard: research groups, industries, other stakeholders etc. |
| Interest | How interesting is the project for MPAI (Low/Medium/High) |

# Still Pictures

## Media & Entertainment

### ME.SP-01: Multires pictures

**Proponent**: Luigi Troiano

**Description**: Pictures are generally managed and distributed in multiple formats, concerning the different usage they meant to be used for, such as digital display or printing. AI can be used to reproduce an image at multiple resolutions using the same minimal core encoding. Preliminary studies in AI-driven super-resolution may lead to innovative solutions in the field of media.

**Comments:**

**Examples:**

**Requirements:**

**Object of standard:**

**Benefits:** To obtain a desired resolution keeping the same minimal encoding

**Bottlenecks:**

**Social aspects:**

**Success criteria:**

**Feature table:**

|  |  |
| --- | --- |
| Volume | Medium |
| Maturity |  |
| AI Relevance | High |
| Users | Press, Media and Advertising Agencies, Photographers, Editors, Publishers |
| Participants |  |
| Interest |  |

## Transportation

Currently, no use case available

## Telco

Currently, no use case available

## Information Technology

Currently, no use case available

## Aerospace

### AS.SP-01: Large-scale, Multi-spectrum Satellite Imagery

**Proponent**: Luigi Troiano

**Description**: Satellite imagery refers to the same area using ultrawide scale resolution and multi-spectrum acquisition. Transmission, storage and distribution of this information is resource intensive. However, the area may include a variety of elements, such as vegetation, buildings, infrastructures, vehicles, people, and others that may require different encoding for their reproduction. The compression scheme may also include the analytic purpose of satellite imagery.

**Comments:**

**Examples:**

**Requirements:**

1. High accuracy (sensibility vs. specificity) in defect detection

**Object of standard:**

**Benefits:** Selective compression according to different elements and analytics

**Bottlenecks:**

**Social aspects:**

**Success criteria:**

**Feature table:**

|  |  |
| --- | --- |
| Volume | High |
| Maturity |  |
| AI Relevance | Med |
| Users | Governmental, environmental and space agencies |
| Participants |  |
| Interest |  |

## Manufacturing

### MF.SP-01: Quality Inspection

**Proponent**: Luigi Troiano

**Description:** One way to inspect the quality of a product is by analysing a high-res picture and identifying possible defects. Images may be required to be stored for documentation and archival purposes, even in compliance with legal regulations. Assuming that in mass production we have a vast amount of objects that are all supposed to be almost the same, this may open the possibility that the whole collection of images in a production batch or multiple batches may serve as basis for high rate compression and anomaly detection.

**Comments:**

**Examples:**

**Requirements:**

**Object of standard:**

**Benefits:** High compression; Encoded defects

**Bottlenecks:**

**Social aspects:**

**Success criteria:**

**Feature table:**

|  |  |
| --- | --- |
| Volume | Med |
| Maturity |  |
| AI Relevance | Medium |
| Users |  |
| Participants |  |
| Interest |  |

## Healthcare

### HC.SP-01: Radiology Imaging

**Proponent**: Luigi Troiano

**Description**: Image analysis in medical diagnostics requires that the introduction of image processing artifacts be minimized. Because of this, acquired images are often stored and distributed in RAW format as compression may alter the diagnostic outcome, which is often based on accurate tissue measurements. This opens the possibility to develop AI-based compression techniques that can reduce the impact of compression on the original image.

**Comments:**

**Examples:**

**Requirements:**

**Object of standard:**

**Benefits:**

**Bottlenecks:**

**Social aspects:**

**Success criteria:**

**Feature table:**

|  |  |
| --- | --- |
| Volume |  |
| Maturity |  |
| AI Relevance |  |
| Users |  |
| Participants |  |
| Interest |  |

### HC.SP-02: Intravenous bottle

**Proponent**: Danilo Pau (STM)

**Description**: Currently, a nurse regularly monitors the level of sodium chloride liquid for intravenous adminis­tration in a bottle. If the saline in the bottle is fully consumed, and the bottle is not replaced or the process stopped immediately, the difference between the patient’s blood pressure and the empty saline bottle could cause an outward rush of blood into the bottle.

In an AI-enabled version of this use case, a camera takes pictures and computes descriptions to check if the saline in the bottle is partially or fully consumed.

**Comments:**

**Examples:**

**Requirements:**

**Object of standard:**

**Benefits:**

**Bottlenecks:**

**Social aspects:**

**Success criteria:**

**Feature table:**

|  |  |
| --- | --- |
| Volume |  |
| Maturity |  |
| AI Relevance |  |
| Users |  |
| Participants |  |
| Interest |  |

### HC.SP-03: Reduced data acquisition in MRI scans

**Proponent**: Gary Olson (GHO Group)

**Description**: Results of ongoing research show that AI can drive a MRI scan to actually acquire less data from the body of the patient than a regular MRI scan would do.

**Comments**: This is a different paradigm than used in compression so far. Instead of data acquisition first and compression later, here less data are acquired

**Examples:**

**Requirements:**

**Object of standard:**

**Benefits:**

**Bottlenecks:**

**Social aspects:**

**Success criteria:**

**Feature table:**

|  |  |
| --- | --- |
| Volume |  |
| Maturity |  |
| AI Relevance |  |
| Users |  |
| Participants |  |
| Interest |  |

[1] Devin Coldewey; Blind test shows AI-enhanced MRI scans are just as good but 4 times faster; TechCrunch <https://techcrunch.com/2020/08/18/blind-test-shows-ai-enhanced-mri-scans-are-just-as-good-but-4-times-faster/>

### HC.SP-04: Correlative analysis in support of infectious disease research

**Proponent**: Gary Olson (GHO Group)

**Description**: Understanding the propagation characteristics of an infectious disease requires correlation across multiple data, image and video sources. Researchers need AI tools to help with the massive amount of global data being collected for diagnosis and population propagation.

**Comments**: The images and video are typically ultra-high resolution, current compression algorithms can potentially compromise the accuracy needed. An AI based compression schema can facilitate collaborative working by enabling information sharing over open networks.

**Examples:**

**Requirements:**

**Object of standard:**

**Benefits:**

**Bottlenecks:**

**Social aspects:**

**Success criteria:**

**Feature table:**

|  |  |
| --- | --- |
| Volume |  |
| Maturity |  |
| AI Relevance |  |
| Users |  |
| Participants |  |
| Interest |  |

## Food & Beverage

### FB.SP-01 - Food information

**Proponent**: Danilo Pau (STM)

**Description**: A user employs an image sensor to classify foods (being prepared, or being taken by a waiter to the user’s table) to know more about the food that is being prepared or eaten etc. The user sends an AI description of the still picture to a remote application which responds to user’s queries.

**Comments:**

**Examples:**

**Requirements:**

**Object of standard:**

**Benefits:**

**Bottlenecks:**

**Social aspects:**

**Success criteria:**

**Feature table:**

|  |  |
| --- | --- |
| Volume |  |
| Maturity |  |
| AI Relevance |  |
| Users |  |
| Participants |  |
| Interest |  |

## Science & Technology

### ST.SP-01: Images from space

**Proponent**: Luigi Troiano

Images from space, such as those acquired by probes and satellites for the exploration of outer space or other planets, are transmitted to the ground and then processed. The fusion and tessellation of images is between this type of processing. The need to transmit large piers of images remotely with a limited transmission bandwidth, poses the demand for AI-based compression techniques that can, on the one hand, guarantee efficient transmission and, on the other hand, enable image fusion and tessellation tasks, also taking into account the limited processing capabilities on board.

**Comments:**

**Examples:**

**Requirements:**

**Object of standard:**

**Benefits:**

**Bottlenecks:**

**Social aspects:**

**Success criteria:**

**Feature table:**

|  |  |
| --- | --- |
| Volume |  |
| Maturity |  |
| AI Relevance |  |
| Users |  |
| Participants |  |
| Interest |  |

## Other Domains

Currently, no use case available

# Moving Pictures

## Media & Entertainment

### MP.ME-04: Game content distribution

Note: the use case is obsolete. It has been used to develop the MPAI-EVC area of work.

**Proponent**: Leonardo Chiariglione (CEDEO)

**Description**: A user plays a game with a simple device that receives sequences of pictures from a remote application. The user sends commands to the application which generates sequences of pictures that depend on the history of user commands and may include pictures that are synthetic (graphics), natural (video) or a combination of the two. Picture sequences are encoded, compressed and sent to the user.

**Comments**:

Classical gaming is a clear use case for MPAI, but we need to know more about

1. How important is the role of natural video today and what are the trends
2. How the natural and synthetic pictures are generated and composed
3. If the 3G graphics information is made accessible separately
4. How can AI be applied to major 3D graphic compression codecs.
5. How extracted features associated with videos should be encoded

A second issue is whether it makes sense to have a generalised AI-based moving picture codec that encompasses 3D Graphics and moving pictures.

**Object of standard**:

1. Audio-video coding scheme that considers the specific nature of video gaming content.

**Benefits:**

**Bottlenecks:**

**Social aspects:**

**Success criteria:**

**Feature table:**

|  |  |
| --- | --- |
| Expected benefits | More efficient compression because we can better adapt compression tools to the content |
| Volume | High stream of data |
| Quality Criteria | Natural part could be traditional objective/subjective criteria, synthetic part TBD |
| Maturity | Research to Inception |
| Relevance of AI | AI is expected to provide better performance that traditional video coding |
| Users | Video gaming service providers |
| Players | Academia/industry research groups, video gaming companies (as requirements providers) |
| Level of interest | High |

### ME.MP-05: AI-enhanced traditional video coding

Note: the use case is obsolete. It has been used to develop the MPAI-EVC area of work.

**Proponent**: Leonardo Chiariglione (CEDEO)

**Description**: Improved compression ratio has been observed by extending existing video coding architectures through AI (e.g. CNN-based video coding tools (rate-distortion optimization, debloc­king filters, interpolation filters, chroma from luma prediction methods), learned entropy coding, end-to-end image compression, decision trees based encoder speed up, fuzzy network band­width prediction, and video network resource allocation via reinforcement learning.

MPAI could capitalise on the big investments made by industry in video coding and augment the performance of video codecs by adding AI components to the scheme(s).

**Comments**:

While improvements have been observed, it is not clear how much the total improvement would be if all these new tools were employed, much less how this could be done.

Therefore, before starting work on this area, a Call for Evidence would be needed to ascertain the level of benefits that the use of AI technologies could bring. Other considerations, such as industry needs for such a standard, may also weigh in.

The starting point of such an AI-enhanced video codec is an issue because typical high-performance video codecs have many patent holders. They would have to be convinced to develop and agree on a framework licence that allows MPAI members to extend the starting point with AI-based tools. This might be a major task because adoption of a framework licence is possible if done ex ante, but hard to achieve if done ex post.

Therefore, the starting point should be selected as one of the following codecs (examples in brackets)

1. Unencumbered by known IP (EVC baseline)
2. Encumbered by IP help by a limited number of IPR holders (EVC)
3. Unencumbered by current IP that will expire in a limited period of time (AVC)

The design of framework licences should probably take into account how AI may influence IP.

MPAI could develop an AI-enhanced traditional video codec because, compared to other organisations, an MPAI standard

* Negative: is constrained by a framework licence and will likely to be less performing than a standard developed based on FRAND declarations.
* Positive: would come with a framework licence that can be very close to an actual licence, while other standards would come with a set of FRAND declarations, probably much larger than we have seen so far
* Positive: could later be extended with more tools and new framework licences.

**Requirements**

1. Friendly to architectures that have vector processing units, e.g. CPU + GPU (or FPGA)
2. Capable to exploit a variable number of CPU cores
3. Capable to encode with varying degrees of delay
4. Capable to shift complexity between encoder and decoder
5. Ideally 50% more compression

**Object of standard**: a video coding scheme that extends a video coding scheme by adding more performing AI-based coding tools

**Object of standard:**

**Benefits:**

1. Extension of existing video coding architecture adding more efficient AI-based tools.
2. Computational cost of the new tools TBD

**Bottlenecks:**

**Social aspects:**

**Success criteria:**

**Feature table:**

|  |  |
| --- | --- |
| Volume | High |
| Maturity | Research |
| AI Relevance | Use of AI is expected (and has already been partly shown) to bring improvements compared to traditional techniques |
| Users | Same as current users of video compression |
| Participants | Academia/industry research groups |
| Interest | High |

### ME.MP-06: Fully AI-based video coding

**Proponent**: Leonardo Chiariglione (CEDEO)

**Description**: Video coding standard that is entirely based on AI technologies, without being constrained by traditional video coding architectures.

**Comments**: A great deal of research has been and is being done to design new concepts of video coding architectures that make use of AI technologies. There is a belief that such new architectures and technologies will offer improved compression, with the added benefits of additional features, such quality scalability, starting from very rough to very accurate reproductions.

**Examples:**

**Requirements:**

1. The coding scheme shall operate from low-level semantic description to pixel-level reprod­uction accuracy
2. The coding scheme shall allow to provide better quality at the same bitrate by increasing encoding delay

**Object of standard:** A video coding scheme designed without constraints, such as those introduced in MP.MP-05.

**Benefits:**

1. Improved compression and support to new features (TBD)

**Bottlenecks:**

**Social aspects:**

**Success criteria:**

**Feature table:**

|  |  |
| --- | --- |
| Volume | High |
| Maturity | Research |
| AI Relevance | High |
| Users | Same as current users of video compression, but more classes of users expected |
| Participants | Research groups, companies and other stakeholders who feel that AI may satify some needs insufficiently supported by traditional video coding |
| Interest | High |

### ME.MP-07: AI based optimisation of existing or new codecs

**Proponent**: Thierry Fautier (Harmonic)

**Description**: current MPEG encoders are partly optimised based on AI techniques, this applies mostly to VOD. We want to use AI to optimise all the coding parameters for VOD and Live with a compute footprint lower than traditional methods applied today. We expect using AI will provide a better quality and will lower bitrate.

**Comments**: we need to ensure the decoders can support all possible variation of parameters. For instance, after 25 years, we still see MPEG-2 broadcast STB than do not accept resolution change. On OTT side, change of manifest is still not propertly supported by neither DASH nor HLS. So, we will need to specify the decoder behavior more stongly than done in MPEG.

**Object of standard**: One output would be a set of recommendations that facilitate the development of AI-based encoder optimisation by providing guidelines on, e.g.

* Test procedures, test sequences, minimum number of test sequences to reach a reliable model
* Parameters that are best fitted for AI control
* How to automate the learning method
* How to create a non-supervised system
* Define ways to detect novelty (there are papers about novelty detection that could work for video compression)
* Test the conditions
* Characterisation/classification of video (e.g. cartoons, sports etc.) by AI and derive parameters

Another output could be

* identify areas where decoders are underspecified
* make the decoder smart enough to detect what it should do
* standardise transitions
* propose additional signaling to the decoder

It is not clear what would the status of MPAI, which is about AI for data coding, if it makes such recommendations. One possibility it to send its findings to relevant organisation.

**Requirements:**

1. At least 20-30% better compression than using existing MPEG standard.
2. If new compression scheme, at least 70% better than existing standard (assuming compression brings 50% better efficiency)
3. CPU usage lower than currently used

**Object of standard:**

This is not expected to be a standard with a normative value but a set of recommendations.

**Benefits:**

1. Lower bitrate-improved quality
2. Lower CPU on encoder side

**Bottlenecks:**

**Social aspects:**

**Success criteria:**

**Feature table:**

|  |  |
| --- | --- |
| Volume | High |
| Maturity | Medium as already partly implemented in commercial products |
| AI Relevance | High |
| Users | Users of compression technology |
| Participants | Research groups, companies and other stakeholders who feel that AI may satisfy some needs insufficiently supported by traditional video coding |
| Interest | High |

### ME.MP-08 AI compression of high-quality video and images

Note: the use case is obsolete. It has been used to develop the MPAI-EVC area of work.

**Proponent**: Gary Olson (GHO Group)

**Description:** The proliferation of the use of video in private security and public safety generates extraordinary amounts of video. This needs to be transmitted to collection centers and/or up­loaded to cloud services. The application of an intelligent compression algorithm without compromising video and image quality can support storage, transport and display of high-quality video and images. This would enable further high-quality processing.

**Comments:**

**Examples:**

**Requirements:**

**Object of standard:**

**Benefits:**

1. Selective compression while maintaining the integrity of the video

**Bottlenecks:**

**Social aspects:**

**Success criteria:**

**Feature table:**

|  |  |
| --- | --- |
| Volume | Very high |
| Maturity | Medium |
| AI Relevance | Cognitive tools and convolutional neural network would be applied |
| Users | Many potential applications |
| Participants | All public and private entity for which video and images on which further processing is applied is important |
| Interest | Very high |

### ME.MP-09: Distributed multiplayer online gaming in Next Generation Games

Note: Part of this use case has been taken over by ME.MP-13: Object and scene description.

**Proponents**: Fabrizio Davide (UnderD), Massimiliano Lunghi (UnderD)

**Description**: Online games in which hundreds, if not thousands, of players (real or bots) share game information at the same time are now commonplace.

Users act in a physical playground. Games are social in character and feature augmented reality and the most modern concepts of gamification (e.g., games addess user behaviour vis-à-vis consumption pattern, citizen engegement in social challenges). They are increasingly focusing on player’s interaction with the physical environment and with other players.

Interactions can be tracked via camera and microphone of the device employed (smartphone, tablet, smart glasses, fixed and/or mobile cameras, etc.). The analysis of the audio-video stream involves specialized AI and implies extrac­ting appropriate features from the streams to reconstruct the players’ action and context. The features extracted may have a local impact, i.e. on the user or an impact on the other users. In the latter case feature extraction can be done at the edge (on the game device), on the server or partly at the edge and partly on the server. The game device sends streams enriched by pre-processed features to the server.

These features will be subsequently correlated on the server side with the videos and related features sent by the other devices. Feature sharing needs standardization to unbundle the AI at the edge from the server side and to allow efficient computational strategies on the edge.

Then the server will send back to the player’s device:

1. a complete audio-visual stream for immediate rendering (Google’s approach for Stadia).
2. an audio-video stream, containing only synthetic objects, to be locally mixed, after temporal and spatial synchronization, with the audio-video stream captured locally.

**Comments**:

**Examples:**

Here are a few examples of known games that illustrate how MPAI could improve feasibility and user experience

*Example 1: Soccer*

Some friends in a pub decide to play soccer on the table. They use an app to place the goalkeeper door and the bezel of the penalty area on the table. In turn, a friend controls the goalkeeper via his smartphone and the other friend the player who takes the penalty. If the shot finds obstacles on the table, such as a glass of beer, the ball will bounce off.

*Example 2: Racing*

A boy uses an app to create a small rally circuit 4-5 meters long by scoring a few points on a sidewalk. Then he chooses a car 30 cm long and uses the smartphone to drive it around the circuit. If the car goes off the circuit, it takes penalties. Another boy may join the race with his car and competes for the top position. If a passerby crosses the circuit, the car that collides with the passerby is bounced off and penalized.

*Example 3: Hunt the creature*

The player looks at the façade of a large building through the screen of her smartphone. Suddenly a creature comes out of a window, and climbs quickly along the façade, like a spider. The player fires at the creature, trying to stop it. The creature is smart, tries to avoid hits and hides behind balconies. The player should catch it before it reaches the street and disappears in the crowd.

*Example 4: Hunt your partner*

Two users located at two mutually remote places play a game where user A climbs the façade of a large building physically located at the place of user B trying to reach to top of the building. Suddenly user A comes out of a door at the ground floor, and climbs quickly along the façade, as spiderman would do. User B fires at user A who tries to avoid being hit, hides behind balconies, comes out and climbs again. User B should hit user A before s/he can reach the top of the building.

*Example 5: Hunt the fox*

A group of users try to hit a fox roaming in the environment of one of the users, called user A. The environment can be a public place such a square or a private place such as the backyard of a house. All participants in the game see the same scene captured by user A. The winner is the one who succeeds to hit the fox. A user is disqualified it the user hits a person in the environment of user A.

*Example 6: Play* *hide and seek*

User A sees a scene (a crowded place) through her handset. User B sees the same scene and tries to hide in the crowd. Use A wins if she can recognise user B in the crowd within a given time.

**Requirements:**

1. Relevant features contained in the video stream
   1. Gesture recognition
   2. Recognition of objects
      1. Vehicles
      2. Persons
      3. animals
      4. Body parts
      5. ...
   3. Scene analysis
2. Relevant features contained in the audio stream
   1. Sounds typical of city life
      1. Traffic (car, bike, trucks, ...)
      2. People’s voices
      3. Syrens, clock tower,
      4. ...
   2. Audio scene analysis
3. Transport of the said information of the audio-video streams and of the features of 1. and 2.

**Object of standard**:

1. Audio-video coding scheme that considers the specific nature of video gaming content.
2. Audio-video feature extraction and transfer assisted by AI.

**Benefits:**

1. More efficient compression because we can better adapt compression tools to the content.
2. More efficient spatial computing functions on portable devices.
3. Better quality for the game experience.
4. More complexity for the game context.

**Bottlenecks:**

**Social aspects:**

**Success criteria:**

**Feature table:**

|  |  |
| --- | --- |
| Volume | High stream of data |
| Maturity | Research to Inception |
| AI Relevance | AI is expected to provide better performance than traditional video coding, efficiency and accuracy in feature computing. |
| Users | Video gaming service providers. |
| Participants | Academia/industry research groups, video gaming companies (as requirements providers) |
| Interest | High |

### ME.MP-10: AI- pan&scan 8K

Note: the use case is obsolete. It has been used to develop the MPAI-EVC area of work.

**Proponent:** Roberto Iacoviello (RAI)

**Description**

New camera models are available on the market, capable of capturing moving pictures with 8K resolution. The broadcaster’s end to end production chain is not ready for 8K: a render time would turn into months, drastically increasing post-production costs. Moreover, the transmission chain is not ready for that bitrate. The broadcaster could shoot in 8K to be future-proof but process and broadcast in 4K or Full HD. Sometimes downscale is not a fair job: it’s better to crop the picture. The main problem is choosing the region of interest and this is a very long job for a professional because should be done picture by picture. AI-tools capable to choose the region of interest maybe leveraging on the attention mechanism could be a good improvement for the production TV chain. Additionally, from the 8K master the broadcaster could derive many HD outputs based on different AI-driven crops, giving different points of view of the same shot.

The algorithm does not necessarily need to be applied to the whole shot, it can also be applied scene by scene and could also be assisted e.g., the director or the editor draws a circle that shape the region of interest in the first frame (a kind of assisted attention mechanism) and then the algorithm applies it to the whole scene.

**Comments:**

**Examples:**

**Requirements:**

**Object of standard:**

**Benefits:**

**Bottlenecks:**

**Social aspects:**

**Success criteria:**

**Feature table:**

|  |  |
| --- | --- |
| Volume |  |
| Maturity |  |
| AI Relevance |  |
| Users |  |
| Participants |  |
| Interest |  |

### ME.MP-11: AI-services for next generation TV

Note: Part of this use case has been taken over by ME.MP-13: Object and scene description.

**Proponent:** Roberto Iacoviello (RAI)

**Description:**

A user is watching TV. The user wants to know basic information about the main character of the movie, the location of the scenes, the audio track. Interactions can be tracked using a remote control as a pointer (or via voice-first devices like smart speakers).

Pointing the main character on the TV screen will display an enriched menu e.g. with his name and a short biography.

In a realistic scenario it can be assumed that an OTT service sends streams enriched by AI-preprocessed features on the servers. These features will be subsequently correlated on the TV/settopbox side with the video. Features sharing needs standardization to unbundle the AI at the edge/cloud from the server aspect, enabling efficient edge/cloud computing strategies.

In this scenario the receiver does not have a high computing capability but enough to implement at least the last layers of a neural network.

In addition, this scenario could enable new services in edge / cloud for opportunistic features extraction.

Let's imagine an app that receives the content of RAI channels and enhances them for the final user based on features extracted from third-party services. Moreover, let's imagine that there could be many of these services, starting a market place of "enhancer apps".

**Comments:**

**Examples:**

**Requirements:**

**Object of standard:**

**Benefits:**

**Bottlenecks:**

**Social aspects:**

**Success criteria:**

**Feature table:**

|  |  |
| --- | --- |
| Volume |  |
| Maturity |  |
| AI Relevance |  |
| Users |  |
| Participants |  |
| Interest |  |

### ME.MP-12: Tracking game player’s movements

Note: Part of this use case has been taken over by ME.MP-13: Object and scene description.

**Proponents**: Leonardo Chiariglione (CEDEO), Roberto Iacoviello (RAI)

**Description**: Game players still use oldish devices when playing a game. This manual interaction could be replaced by a system that tracks the player’s movements for faster communication of the player’s intentions to the game machine located on a server.

**Comments:**

**Examples:**

**Requirements:**

**Object of standard:** descriptors of user movements

**Benefits:**

1. More natural game player’s interaction with the games
2. Ability of a game player to access any via any network-distributed game that understands the standard game player movement representation.

**Bottlenecks:** Potential synergies with

1. ME.MP-09 Distributed multiplayer online gaming
2. TP.MP-01 AI-assisted driving

**Social aspects:**

**Success criteria:**

**Feature table:**

|  |  |
| --- | --- |
| Volume |  |
| Maturity |  |
| AI Relevance |  |
| Users |  |
| Participants |  |
| Interest |  |

### ME.MP-13: Visual Object and Scene Description

**Proponent**: Leonardo Chiariglione (CEDEO)

**Description**: This use case addresses the “object and scene description” component of several use cases considered in this use case documents. By “object and scene description” we intend the usual description (language) of objects and their attributes in a scene (for which solutions exist) and the actual semantic description of the objects (for which some proprietary solutions exist).

The Examples section of this use case shows that proprietary solutions can address the needs of the examples, but at the cost of lack of interoperability or by forcing all user to adopt a single technology or application. A standard representation of the objects in a scene allows for an easier implementation of those examples.

**Comments:**

**Examples:**

1. *Multiplayer online gaming*

In one use case considered in *ME.MP-09: Distributed multiplayer online gaming in Next Gener­ation Games* (Hide and seek)

1. Player A
   1. Points their smartphone to a scene populated by persons
   2. Sends the description of the scene to a server
   3. Sends commands to the server to animate a synthetic person with the goal to hide it among the persons
2. The server
   1. Understands the scene description received
   2. Adds the animated synthetic person to the scene
   3. Distributes the composite scene to all players
3. All players but A seek the synthetic person in the crowd.
4. *Person matching*

In *ME.MP-11: AI-services for next generation TV*, descriptors of the video program and associated metadata are transmitted next to the actual program.

When the user wants to know more about an object,

1. User points the remote control to the object on the screen
2. Set top box
   1. Computes descriptors of the object
   2. Compares the computed descriptors with those in the stream
   3. Executes as per metadata when a match is found
3. *Tracking game player’s movements*

In *ME.MP-12: Tracking game player’s movements*, the game device/application sends descrip­tions of the game player’s movements to a server. The server decodes the game player’s inten­tion from the descriptors.

1. *AI-assisted driving*

In *TP.MP-01: AI-assisted driving*, a car computes the descriptors of the scene and displays ap­propriate messages to the driver. When necessary, the car communicates the scene descriptors to neighbouring cars.

1. *Correct Posture*

In *HC.MP-02: Correct Posture*, an application analyses the video of a person to suggest how the person should correct their posture.

1. *Integrative genomic/video experiments*

*ST.OD-06: Integrative AI-based analysis of multi-source genomic/sensor experiments* considers the case there genomic data and their effects on living organisms is jointly assessed by AI. The article <https://www.theguardian.com/environment/2020/oct/08/behind-chinas-pork-miracle-how-technology-is-transforming-rural-hog-farming> shown how industrial pig farming operations in China are already partly addressing the issue. Here it is necessary to automatically identify anom­alous behaviour of a pig that may indicate that the animal is affected by an illness.

**Requirements:**

**Object of standard**: an “obvious” solution is to standardise the neural network that is capable to describe the object. In most cases this solution is less than ideal because it “freezes” the technology and prevents innovation.

A better solution that should be addressed by MPAI could be to select a set of descriptors whose performance has been considered the best under given conditions and leave it to individually-created neural networks to convert imput signals into the standard set of descriptors

**Benefits:**

**Bottlenecks:**

**Social aspects:**

**Success criteria:**

**Feature table:**

|  |  |
| --- | --- |
| Volume | High |
| Maturity | Inception/Adoption |
| AI Relevance | High |
| Users | Virtually everybody |
| Participants | Academia/research, companies with actual solutions for some objects |
| Interest | High |

## Transportation

### TP.MP-01: AI-assisted driving

Note: Part of this use case has been taken over by ME.MP-13: Object and scene description.

**Proponents**: Leonardo Chiariglione (CEDEO), Ajay Luthra (Picsel Labs)

**Description:** Fully connected and autonomously driving vehicles have the potential to revolutionize transportation mobility and are likely to become components of many peoples’ lives in future. However, vehicles that understand the environment to a certain degree, take some limited actions and commun­icate with the external environment are already becoming a reality.

Society of Automotive Engineers (SAE) (<https://www.sae.org/>) has defined 6 levels of driving autonomous (see below).

Center of Advanced Automotive Technology (<http://autocaat.org/>) has defined 5 types of connec­ted vehicles that can communicate (1) vehicle-to-infrastructure [V2I]), (2) vehicle-to-vehicle (V2V), (3) vehicle-to-cloud [V2C], (4) vehicle-to-pedestrian [V2P] and vehicle-to-everything [V2X] (see below for the next level of information on the benefits).

Depending upon its capabilities (level of autonomy and connectivity), a vehicle equipped with electromagnetic and acoustic sensors, GPS, odometry, and inertial measurement systems may be able to sense that there is an anomaly in the environment, have a level of understanding of the anomaly and take various actions, for example, changing its velocity, and/or alerting the driver and other vehicles etc.

**Comments**:

**Examples**:

1. A vehicle knows what characterises a vehicle and takes an action when, say, preceding vehicles behave in an unexpected way.
2. A vehicle knows what characterises a living body moving on the road and takes an action when the living body is likely to intercept the vehicle’s trajectory.
3. A vehicle knows what characterises a non-identified object on the road.
4. A vehicle knows the main sounds relevant on a road.
5. A vehicle knows/detects the road signs and lights and takes appropriate actions
6. [Add examples of connectivity]

**Requirements**:

**Object of standard**:

The following elements should be included in the standard

1. The types of objects considered by the standard
2. A set of primitives at the appropriate level of granularity that can be used to describe and compose the audio-visual objects
3. An object description language to describe how an object can be reconstructed from its descriptors
4. A scene description language to describe where relevant objects in the environment are placed, what are their speeds and expected trajectories.

**Benefits**: By standardising the object primitives, MPAI will create a competitive market of algorithms and hardware components exposing standard interfaces that may find use in areas beyond autonomous driving.

**Bottlenecks**:

**Social aspects**:

**Success criteria**:

**Feature table**

|  |  |
| --- | --- |
| Volume | High |
| Maturity | Inception |
| AI relevance | High |
| Users | Manufacturers of automotive components |
| Participants | Academia, Research, Automotive industry |
| Interest | High |

**Annex A**

**Additional Information**

**From SAE**

Levels of Automation

Fully automated, autonomous, or “self-driving” vehicles are defined by the U.S. Department of Transportation's National Highway Traffic Safety Administration (NHTSA) as “those in which operation of the vehicle occurs without direct driver input to control the steering, acceleration, and braking and are designed so that the driver is not expected to constantly monitor the roadway while operating in self-driving mode.” There have been multiple definitions for various levels of automation, for the sake of standardization, and to aid clarity and consistency, NHTSA has adopted the SAE International definitions for levels of automation. These definitions divide vehicles into levels based on “who does what, when.”

Level 0: The human driver does all the driving.

Level 1: An advanced driver assistance system (ADAS) on the vehicle can assist the human driver with either steering or braking/accelerating.

Level 2: An ADAS on the vehicle can control both steering and braking/accelerating under some circumstances. The human driver must continue to pay full attention (“monitor the driving environment”) at all times and perform the rest of the driving task.

Level 3: An automated driving system (ADS) on the vehicle can perform all aspects of the driving task under some circumstances. The human driver must be ready to take back control at any time the ADS requests the human driver to do so. In all other circumstances, the human driver performs the driving task.

Level 4: An ADS on the vehicle can itself perform all driving tasks and monitor the driving environment – essentially, do all the driving – in certain circumstances. The human need not pay attention in those circumstances.

Level 5: An ADS on the vehicle can do all the driving in all circumstances. The human occupants are just passengers and need never be involved in driving.

**From Center of Advanced Automotive Technology (http://autocaat.org/)**

Connected Vehicles

Connected vehicles are vehicles that use any of a number of different communication technologies to communicate with the driver, other cars on the road (vehicle-to-vehicle [V2V]), roadside infrastructure (vehicle-to-infrastructure [V2I]), and the “Cloud” [V2C]. This technology can be used to not only improve vehicle safety, but also to improve vehicle efficiency and commute times. Listed below are the types of communicaton, with links to more information, and some of the benefits of connected vehicles:

V2I: Vehicle to Infrastructure

V2V: Vehicle to Vehicle

V2C: Vehicle to Cloud

**Benefits of Connected Autonomous Vehicles**

In 2016, motor vehicle-related crashes on U.S. highways claimed 37,461 lives. US Dept of Transportation research shows that 94% of serious crashes are due to human error. Some of the benefits of connected and automated vehicles include:

|  |  |
| --- | --- |
| crash elimination | **Crash Elimination:** Crash-free driving and improved vehicle safety, a vehicle can monitor the environment continuously, making up for lapses in driver attention. |
| reduced infrastructure | **Reduced Need for New Infrastructure:** By managing traffic flow, self-driving can reduce the need for building new infrastructure and reduce maintenance costs |
| improved travel time | **Travel Time Dependability:** V2V, V2C, and V2I can substantially reduce uncertainty in travel times via real-time, predictive assessment of travel times on all routes |
| productivity improvements | **Productivity Improvements:** A reduction in driving tasks will allow travelers to use travel time more productively |
| energy efficiency | **Improved Energy Efficiency:** Reduced energy consumption in at least three ways: more efficient driving; lighter, more fuel-efficient vehicles; and efficient infrastructure |
| new models of vehicle usage | **New Models for Vehicle Ownership:** Self-driving vehicles could lead to a major redefinition of vehicle ownership and expand opportunities for vehicle sharing |
| new business models | **New Business Models and Scenarios:** Convergence of technologies may realign industries such that companies need to compete and collaborate at the same time |

### TP.MP-02: Highspeed Train Safety

**Proponent**: Luigi Troiano

**Description**: Recent studies and research projects have shown how AI can be used for the recognition of objects along the railway line that may endanger the safety of high-speed trains. Considering the cruising speed (e.g. 300Km/h in Italy) and the stopping space (up to 5400m on level ground) this requires a video system for monitoring the railway network, considering the integration of video streams from distributed sources (typically fixed).

**Comments:**

**Examples:**

**Requirements:**

**Object of standard:**

**Benefits:**

**Bottlenecks:**

**Social aspects:**

**Success criteria:**

**Feature table:**

|  |  |
| --- | --- |
| Volume |  |
| Maturity |  |
| AI Relevance |  |
| Users |  |
| Participants |  |
| Interest |  |

## Telco

### TC.MP-01: Videoconferencing

This use case is obsolete. It has been incorporated in MPAI-EVC Use Case

**Proponent**: Luigi Troiano

**Description**: The increasing spread of video-conferencing tools poses new challenges to telecommunications infrastructures for the transmission and recording of video. Given that video-conferencing typically captures a subject in the foreground, possibly with a blurred background, this paves the way for specific coding methods based on AI models that are particularly efficient for this type of video content, considering possible criteria such as available bandwidth, latency, infrastructure load, nature and stability of the connection.

**Comments:**

**Examples:**

**Requirements:**

**Object of standard:**

**Benefits:**

**Bottlenecks:**

**Social aspects:**

**Success criteria:**

**Feature table:**

|  |  |
| --- | --- |
| Volume |  |
| Maturity |  |
| AI Relevance |  |
| Users |  |
| Participants |  |
| Interest |  |

## Information Technology

### IT.MP-01: Source verification

**Proponent**: Luigi Troiano

**Description**: The problems related to fake news through video content that may have been altered in a misleading way, poses important requirements regarding the possibility of verifying the authenticity of a content, or the impossibility that it has been altered or artfully constructed. This may require encoding schemes capable of limiting activities aimed at altering content that are not easily verifiable.

**Comments:**

**Examples:**

**Requirements:**

**Object of standard:**

**Benefits:**

**Bottlenecks:**

**Social aspects:**

**Success criteria:**

**Feature table:**

|  |  |
| --- | --- |
| Volume |  |
| Maturity |  |
| AI Relevance |  |
| Users |  |
| Participants |  |
| Interest |  |

## Aerospace

Currently, no use case available

## Manufacturing

## Healthcare

### HC.MP-02: Correct Pose

Note: Part of this use case has been taken over by ME.MP-13: Object and scene description.

**Proponent**: Danilo Pau (STM)

**Description**: A user employs an image sensor to know how s/he should correct their posture. The sensor sends an AI description of a video clip to a remote application which suggests how and what the user should change to improve their posture.

could rely on a subset of AI descriptors for human movement selected for their specific applications.

**Comments:**

**Examples:**

**Requirements:**

**Object of standard:**

**Benefits:**

**Bottlenecks:**

**Social aspects:**

**Success criteria:**

**Feature table:**

|  |  |
| --- | --- |
| Volume |  |
| Maturity |  |
| AI Relevance |  |
| Users |  |
| Participants |  |
| Interest |  |

## Food & Beverage

### FB.MP-01: Checking the health of vines

**Proponent**: Danilo Pau (STM)

**Description**: An AI camera placed on a rover moving at 6-8 km/h along the rows of vines captures (sequences of) pictures to monitor the disease affecting vines at different times of the day. The sick vines are georeferenced.

**Comments:**

**Examples:**

**Requirements:**

**Object of standard:**

**Benefits:**

**Bottlenecks:**

**Social aspects:**

**Success criteria:**

**Feature table:**

|  |  |
| --- | --- |
| Volume |  |
| Maturity |  |
| AI Relevance |  |
| Users |  |
| Participants |  |
| Interest |  |

## Science & Technology

Currently, no use case available

## Other Domains

Currently, no use case available

# Audio

## Media & Entertainment

### ME.AU-01: Purpose-dependent Audio Reproduction

**Proponent**: Luigi Troiano

**Description:** The sound has cognitive aspects that cannot be captured solely by the time/frequency analysis of the audio track. These aspects include the origin of the source and its purpose, which determine different requirements. For example, playing music and voice poses different problems and evaluations. The use of AI in recognizing the type of audio in every single moment can pave the way for the definition of encoding schemes that are selective and adaptive in relation to it.

**Comments:**

**Examples:**

**Requirements:**

**Object of standard:**

**Benefits:**

**Bottlenecks:**

**Social aspects:**

**Success criteria:**

**Feature table:**

|  |  |
| --- | --- |
| Volume |  |
| Maturity |  |
| AI Relevance |  |
| Users |  |
| Participants |  |
| Interest |  |

### ME.AU-02: Context-based Audio Enhancement (MPAI-CAE)

Note: This use case is obsolete. It has moved MPAI-CAE Functional Requirements

**Proponents**: Michelangelo Guarise, Andrea Basso (VOLUMIO)

**Description**:

The quality of the overall user experience or content is highly dependent on the context in which audio is used, e.g.

1. Entertainment: audio in the home, in the car (automotive) on-the-go (e.g. while doing sports, running, biking, etc.)
2. Voice communications: e.g. office, car, home, on-the-go
3. Audio and video conferencing: e.g. office, car, home, on-the-go
4. (serious) gaming applications: e.g. office, home, on-the-go
5. Audio (post-)production: in the studio
6. Audio restoration: in the studio

By using context information to act on the content using AI, it is possible substantially to improve the user experience.

**Comments:**

Currently, there are solutions for some of the contexts mentioned above. However, they tend to be vertical in nature. This makes it difficult to re-use AI-based elements of the solutions for different applications. A horizontal market of re-usable components exposing standard interfaces would make the market more competitive.

**Examples**

The following examples describe and initial list of cases where MPAI-CAE can make the difference.

1. *Enhanced audio experience in a conference call*

Often, the user experience of a video/audio conference can be marginal. Too much background noise or undesired sounds can lead to participants not understanding what participants are saying. By using AI-based adaptive noise-cancellation and sound enhancement, MPAI-CAE can virtually eliminate those kinds of noise without using complex microphone systems to capture environment characteristics.

1. *Pleasant and safe music listening while biking*

While biking in the middle of city traffic, AI can process the signals from the environment captured by the microphones available in many earphones and earbuds (for active noise cancellation), adapt the sound rendition to the acoustic environment, provide an enhanced audio experience (e.g. performing dynamic signal equalization), improve battery life and selectively recognize and allow relevant environment sounds (i.e. the horn of a car). The user enjoys a satisfactory listening experience without losing contact with the acoustic surroundings.

1. *Emotion enhanced synthesized voice*

Speech synthesis is constantly improving and finding several applications that are part of our daily life (e.g. intelligent assistants). In addition to improving the ‘natural sounding’ of the voice, MPAI-CAE can implement expressive models of primary emotions such as fear, happiness, sad­ness, and anger.

1. *Efficient 3D sound*

MPAI-CAE can reduce the number of channels (i.e. MPEG-H 3D Audio can support up to 64 loudspeaker channels and 128 codec core channels) in an automatic (unsupervised) way, e.g. by mapping a 9.1 to a 5.1 or stereo (radio broadcasting or DVD), maintaining the musical touch of the composer.

1. *Speech/audio restoration*

Audio restoration is often a time-consuming process that requires skilled audio engineers with specific experience in music and recording techniques to go over manually old audio tapes. MPAI-CAE can automatically remove anomalies from recordings through broadband denoising, declicking and decrackling, as well as removing buzzes and hums and performing spectrographic 'retouching' for removal of discrete unwanted sounds.

1. *Normalization of volume across channels/streams*

Eighty-five years after TV has been first introduced as a public service, TV viewers are still strug­gling to adapt to their needs the different average audio levels from different broadcasters and, within a program, to the different audio levels of the different scenes.

MPAI-CAE can learn from user’s reactions via remote control, e.g. to a loud spot, and control the sound level accordingly.

1. *Automotive*

Audio systems in cars have steadily improved in quality over the years and continue to be integrated into more critical applications. Toda, a buyer takes it for granted that a car has a good automotive sound system. In addition, in a car there is usually at least one and sometimes two microphones to handle the voice-response system and the hands-free cell-phone capability. If the vehicle uses any noise cancellation, several other microphones are involved. MPAI-CAE can be used to improve the user experience and enable the full quality of current audio systems by reduc­ing the effects of the noisy automotive environment on the signals.

1. *Audio mastering*

Audio mastering is still considered as an ‘art’ and the prerogative of pro audio engineers. Normal users can upload an example track of their liking (possibly obtained from similar musical content) and MPAI-CAE analyzes it, extracts key features and generate a master track that ‘sounds like’ the example track starting from the non-mastered track. It is also possible to specify the desired style without an example and the original track will be adjusted accordingly.

**Requirements:**

This is a small sample of the functional requirements that will be developed in the next few weeks. When the full set of requirements will be developed, the MPAI General Assembly will decide whether an MPAI-CAE standard should be developed.

1. The standard shall specify the following natural input signals
   1. Microphone signals
   2. Inertial measurement signals (Acceleration, Gyroscope, Compass, ...)
   3. Vibration signals
   4. Environmental signals (Proximity, temperature, pressure, light, ...)
   5. Environment properties (geometry, reverberation, reflectivity, ...)
2. The standard shall specify
   1. User settings (equalization, signal compression/expansion, volume, ...)
   2. User profile (auditory profile, hearing aids, ...)
3. The standard shall support the retrieval of precomputed environment model (audio scene, domotic scene, ...)
4. The standard shall reference the user authentication standards/methods required by the specific MPAI-CAE context
5. The standard shall specify means to authenticate the components and pipelines of an MPAI-CAE instance
6. The standard shall reference the methods used to encrypt the MPAI-CAE streams and of service-related metadata
7. The standard shall specify the adaptation layer of MPAI-CAE streams to delivery protocols of common use (e.g. Bluetooth, Chromecast, DLNA, ...)

**Object of standard**:

Three areas of standardization are identified:

1. **Context type interfaces**: a first set of input and output signals with their syntax and semantics, corresponding to audio usage contexts considered of sufficient interest (e.g. audioconferencing and audio consumption on-the-go).

They shall have the following features

* 1. Input and out signals are context specific. However, they have a significant degree of commonality across contexts
  2. The way a framework operates is implementation-dependent offering implementors the way to produce the set of output signals that best fit the usage context

1. **Processing component interfaces**: with the following features
   1. Interfaces of a set of updatable and extensible processing modules (both traditional and AI-based)
   2. Possibility to create processing pipelines and the associated control (including the needed side information) required to manage them.
   3. The processing pipeline may be a combination of local and in-cloud processing
2. **Delivery protocol interfaces**
   1. Interfaces of the processed audio signal to a variety of delivery protocols

**Benefits:** MPAI-CAE will bring benefits affecting the following market players

1. Technology providers need not develop full applications to put to good use their technol­ogies. They can concentrate on making better the AI technology that improves the user exper­ience. Further, their technologies can find a much broader use in other application domains.
2. Equipment manufacturers and application vendors can tap from the set of technologies made available according to the MPAI-CAE standard from different competing sources, integrate them and satisfy their specific needs
3. Service providers can deliver complex optimizations and thus superior user experience with minimal time to market as the MPAI-CAE framework enables easy combination of 3rd party components from a technical and licensing perspective. Their services can deliver a high quality, consistent user audio experience with minimal dependency on the source by selecting the optimal delivery method
4. End users enjoy a competitive market that provides constantly improved user exper­iences and controlled cost of AI-based audio endpoints.

**Bottlenecks:** the full potential of AI in MPAI-CAE would be unleashed by a market of AI-friendly processing units and introducing the vast amount of AI technologies into products and services.

**Social aspects**: MPAI-CAE would help

1. Free users from the dependency on the context in which they operate
2. Make the content experience more personal
3. Make the collective service experience less dependent on events affecting the individual participant
4. Raise the technology level of past content to today’s expectations.

**Success criteria**:

These are some of the criteria

1. A competitive market of AI-based components exposing standard interfaces
2. A competitive market of processing units available to manufacturers
3. A competitive marlet of many types of end user devices

**Feature table**:

|  |  |
| --- | --- |
| Volume | The volume of data handles is medium |
| Maturity | The state of the technologies required is at inception level |
| AI Relevance | Very relevant because AI is the differentiating/enabling factor in the market |
| Users | Consumers at large; prosumers; examples of professional users |
| Participants | Researchers; audio equipment manufacturers and application providers; audio service providers; audio professionals |
| Interest | High |

### ME.AU-03: Context-aware audio experience enrichment

Note: This use case is obsolete. It has moved MPAI-CAE Functional Requirements

**Proponent**: Michelangelo Guarise, Andrea Basso (VOLUMIO)

**Description**: Multimedia content and in particular audio content is consumed in a variety of different environments, times, with different purposes and is often combined with ancillary experiences.

For example, within a home automation system, when a certain genre is played (or in a specific time of the day) environmental lighting is supposed to change (bright when listening high BPM music, soft when listening chillout at late night). Or, when in the car, in the morning we can safely assume some wants to hear news while in the evening talk radio.

The context-aware audio experience aims to learn preference of the user based on his past actions and combination patterns with ancillary experiences (also in combination with voice-control UI).

The role of AI is to learn and predict context-aware patterns which are both universal among a certain population (derived from specific demographics or matching criterias) or of the individual user, to invoke “listening scenes” which encompass contextual events and criteria that otherwise will require a complex setup for each scene and ancillary technology used.

AI will also create and adapt new “listening scenes” in first-time conditions, based on user histories and therefore propose coherent experiences also in non-repeatable contexts.

**Comments:**

**Examples:**

**Requirements:**

1. Ability to offer coherent experience that user likes, involving repeatable and non-repeatable contexts

**Object of standard:**

**Benefits:**

1. Automatic creation of “Listening Scenes” that involve audio, context and ancillary technologies

**Bottlenecks:**

**Social aspects:**

**Success criteria:**

**Feature table:**

|  |  |
| --- | --- |
| Volume | High |
| Maturity |  |
| AI Relevance | High |
| Users |  |
| Participants |  |
| Interest |  |

### ME.AU-04: Normalization of volume across channels/streams

Note: This use case is obsolete. It has moved MPAI-CAE Functional Requirements

**Proponent**: Leonardo Chiariglione (CEDEO)

**Description:** Eighty-five years after TV has been introduced as a public service, TV viewers are still struggling to adapt to their needs the different average audio levels from different broadcasters and, within a program, to the different audio levels of the different types of broadcast content.

Artificial Intelligence can provide solutions that learn from the user’s reactions via remote control, e.g. to a loud spot, to control the sound level accordingly.

**Comments**:

1. The learning process can very well never end because the machine may understands that it has made an error if it has changed the sound level, still the user further adapts the sound level.
2. Video could be part of the learning and sound adaptation (in the future).

**Requirements:**

**Object of standard:** these are possible elements requiring standardisation

1. A suitably selected set of sounds to be used for machine learning
2. A suitably selected set of sounds to be used to check how well the machine has learned
3. The process to test how fast and well the machine adapts to the needs of a test user

**Benefits:**

1. Much appreciated by TV viewers sensitive to the level of sound: often unneededly too loud if the user wants to hear the whispers in the program

**Bottlenecks:**

**Social aspects:**

**Success criteria:**

**Feature table:**

|  |  |
| --- | --- |
| Volume | Very high. Could be a standard component of future TV sets |
| Maturity | There is enough technology for a first standard, but progress in this area will be fast as this “personal sound adpter will be a strong differentiating factor |
| AI Relevance | AI is the only technology that can solve the problem. All other attempts made to accommodate the requests of TV viewers have failed |
| Users | TV viewers |
| Participants | TV manufacturers, service providers, machine learning experts |
| Interest | Very high |

## Transportation

### TP.AU.01: Personal communications at high-speed

**Proponent**: Luigi Troiano

**Description:** Audio communications (es. voice calls) with personal devices can be problematic when moving around at high speed, for example on trains or planes. The availability of adaptive encoding models that take into account the infrastructure and trajectory of source, could allow better use of personal communications services.

**Comments:**

**Examples:**

**Requirements:**

**Object of standard:**

**Benefits:**

**Bottlenecks:**

**Social aspects:**

**Success criteria:**

**Feature table:**

|  |  |
| --- | --- |
| Volume |  |
| Maturity |  |
| AI Relevance |  |
| Users |  |
| Participants |  |
| Interest |  |

### TP.AU.02: Traffic sound

**Proponent**: Danilo Pau (STM)

**Description**: A microphone system placed at a crossroad (or in a car) captures the sound made by a crash. The microphone computes and sends the AI descriptors to a remote traffic control centre where an application uses the descriptors to reconstruct the dynamics of the accident and trace the sound back to NCAP crashes.

**Comments:**

**Examples:**

**Requirements:**

**Object of standard:**

**Benefits:**

**Bottlenecks:**

**Social aspects:**

**Success criteria:**

**Feature table:**

|  |  |
| --- | --- |
| Volume |  |
| Maturity |  |
| AI Relevance |  |
| Users |  |
| Participants |  |
| Interest |  |

## Telco

### TC.AU-01: High-fidelity voice

**Proponent**: Luigi Troiano

**Description:** Voice services can benefit from new AI-based encoding schemes that are specifically designed for voice reproduction, eliminating compression-related distortions and thus significantly increasing the quality of the voice reproduced at the other end.

**Comments:**

**Examples:**

**Requirements:**

**Object of standard:**

**Benefits:**

**Bottlenecks:**

**Social aspects:**

**Success criteria:**

**Feature table:**

|  |  |
| --- | --- |
| Volume |  |
| Maturity |  |
| AI Relevance |  |
| Users |  |
| Participants |  |
| Interest |  |

## Information Technology

Currently, no use case available

## Aerospace

Currently, no use case available

## Manufacturing

Currently, no use case available

## Healthcare

### HC.AU-01: 3D Audio

**Proponent**:

**Description**: A user with special needs uses a HCI device to interact with a 3D audio environment.

**Comments:**

**Examples:**

**Requirements:**

**Object of standard:**

**Benefits:**

**Bottlenecks:**

**Social aspects:**

**Success criteria:**

**Feature table:**

|  |  |
| --- | --- |
| Volume |  |
| Maturity |  |
| AI Relevance |  |
| Users |  |
| Participants |  |
| Interest |  |

## Food & Beverage

Currently, no use case available

## Science & Technology

Currently, no use case available

## Other Domains

Currently, no use case available

# Audio and Video

## Media & Entertainment

Currently, no use case available

## Transportation

Currently, no use case available

## Telco

Currently, no use case available

## Information Technology

Currently, no use case available

## Aerospace

Currently, no use case available

## Manufacturing

Currently, no use case available

## Healthcare

### HC.AV-01: Vision-to-Sound Transformation

Note: Part of this use case has been taken over by ME.MP-13: Object and scene description.

**Proponents**: Michel Hospital (UFSCAR -AIM LUZ) and Guy Paillet (General Vision)

**Description**

It is possible to give a spatial representation of an image that blind people can hear with 2 headphones as a localization and description medium. It is a conversion (compression) technique from one space to a different interpretation space. Such substitution of vision with hearing has proven its worth [1].

Three issues could be of interest:

1. *Interpretation*: Are there other conversion spaces, considering the human auditive characteristics? Sound intensity is used to evaluate brightness or distance. As we know frequency is more discriminative than intensity of sound.
2. *Compression*: Is today’s conversion the best in terms of compression and ease of human audition interpretation?
3. *Spatialization*: Today the spatialization of the sound, which is the direction where the pixel stands can only be deduced [1]:

* (azimuth) by a difference in sound intensity between left and right see
* (elevation) by the sound frequency (Higher frequency sound, pixel is high)

When we use a classic spatialization, localization is done at the back of the head which is not satisfactory when the objects are located in front. A 360º spatialization of sound in azimuth and elevation and intensity (brightness, color, distance) would be preferable.

**Comments**: We are dealing with Representation of Sensorial Conversion information and sound diffusion (3D spatialization). Compression a requirement when the visual information converted to audio must be transmitted.

**Examples**: a couple of representative examples would be bery useful

**Requirements:**

**Object of standard**:

1. A standard representation of a 3D visual object with a 3D audio object
2. The syntax and a semantics of a bistream representing a 3D visual scene that is represented as a 3D audio scene.

**Benefits**:

1. Blind and Visually Impaired People can use another sense to have a cue for navigation and recognition in a 3D space.
2. Music can be generated from an image and shape representation [2].

**Bottlenecks**:

**Social aspects**: Blind and Visually Impaired People can have a a cue for navigation and recognition in real time, have a better life and become more productive.

**Success criteria**: The number of people using this sensory substitution system and the ability not only to improve the life of millions of disabled people.

**References:**

[1] Peter Meijer. The Voice “Augmented Reality for the Totally Blind”

https://www.seeingwithsound.com/

[2] Iannis Xenakis, “musiques formelles”, (Revue Musicale no 253-254, 1963, 232 p. - réédition: Paris, Stock, 1981, 261 p)

https://fr.wikipedia.org/wiki/Iannis\_Xenakis

## Food & Beverage

Currently, no use case available

## Science & Technology

Currently, no use case available

## Other Domains

Currently, no use case available

# Event sequence

## Media & Entertainment

Currently, no use case available

## Transportation

### TP.ES-01: Predictive Vehicle Maintenance

**Proponent**: Luigi Troiano

**Description:** Cars transmit large amounts of streaming events concerning the management, driving, position and sensors of the car. Having a coding scheme specifically designed for this type of data, considering the nature and near/real-time constraints for their analysis, specifically for predictive maintenance, can simplify the problems of data transmission and storage.

**Comments:**

**Examples:**

**Requirements:**

**Object of standard:**

**Benefits:**

**Bottlenecks:**

**Social aspects:**

**Success criteria:**

**Feature table:**

|  |  |
| --- | --- |
| Volume |  |
| Maturity |  |
| AI Relevance |  |
| Users |  |
| Participants |  |
| Interest |  |

## Telco

## Information Technology

## Aerospace

## Manufacturing

## Healthcare

Currently, no use case available

## Food & Beverage

Currently, no use case available

## Science & Technology

Currently, no use case available

## Other Domains

### ES-OD-01: Compression and Understanding of Industrial Data

This use case is obsolete. It has moved to MPAICUI Functional Requirements

**Proponents**: Guido Perboli (POLITO), Valeria Lazzaroli (Arisk), Mariangela Rosano (POLITO)

**Description**: Most economic organizations, e.g., companies, etc., produce large quantities of data, often because these are required by regulation. Users of these data maybe the company itself or Fintech and Insurtech services who need to access the flow of company data to assess and mon­itor financial and organizational performance, as well as the impact of vertical risks (e.g., cyber, seismic, etc.). For example, nowadays companies heavily rely on the security and dependability of their Information System for all the categories of workers, including the management of Industrial Control Systems. Adding into the risk analysis process cybersecurity-related parameters will help a more precise estimation of the actual risk exposure. Cybersecurity data will help a reassessment of financial parameters based on risk analysis data.

The sheer amount of data that need to be exchanged is an issue. Analysing those data by humans is typically on­erous and may miss vitally important information. Artificial intelligence (AI) may help reduce the amount of data with a controlled loss of information and extract the most relevant information from the data. AI is considered the most promising means to achieve the goal.

Unfortunately, the syntax and semantics of the flow of data is high dependent on who has produced the data. The format of the date is typically a text file with a structure not designed for indexing, search and ex­traction. Therefore, in order to be able to apply AI technologies to meaningfully reduce the data flow, it is necessary to standardize the formats of the components of the data flow and make the data “AI friendly”.

**Comments**:

Recent regulations are imposing a constant monitoring (ideally monthly). Thus, there is the pos­sibility to have similar blocks of data in temporally consecutive sequences of data.

The company generating the data flow may need to perform compression and understanding for its own needs (e.g., to identify core and non-core data). Subsequent entities may perform further data compres­sion and transformation.

In general, compressed data should allow for easy data search and extraction.

In a first phase MPAI-CUI is addressing primarily risk identification.

**Examples**

MPAI-CUI may be used in a variety of contexts, such as:

1. To support the company’s board in deploying efficient strategies. A company can analyse its financial performance, identifying possible clues to the crisis or risk of bankruptcy years in advance. It may help the board of directors and decision-makers to make the proper decisions to avoid these situations, conduct what-if analysis, and devise efficient strategies.
2. To assess the financial health of companies that apply for funds/financial help. A financial institution that receives a request for financial help from a troubled company, can access its financial and organizational data and make an AI-based assessment of that company, as well as a prediction of future performance. This aids the financial institution to take the right decision in funding or not that company, having a broad vision of its situation.

To assess the risk in different fields considering non-core data (e.g., non-financial data). Accurate and targeted sharing of core and non-core data that ranges from the financial and organizational information to other types of risks that affect the business continuity (e.g., environmental, seismic, infrastructure, and cyber). As an example, the cybersecurity preparedness status of a company allow better estimation of the average production parameters, like the expected number of days of production lost, which are affected by cyberattacks (e.g., days the industrial plants are stopped, days personnel cannot perform their work due to unavailability of the information system). Several parameters need to be considered, which are obtained by direct acquisition of data from the target companies that perform a cybersecurity risk analysis.

1. To analyse the effects of disruptions on the national economy, e.g., performance evaluation by pre/post- pandemic analysis [1].

**Requirements:**

1. The standards of the data in the data flow should be AI friendly. In other words, the different data required to predict a crisis/bankruptcy of a company should be gathered, carefully selected and eventually completed to be suitable for automatic analysis and then treated by an AI-based algorithm.
2. The standard shall ensure efficiency of data structure, indexing and search, according to specific syntax and semantics.
3. The standard shall all the extraction of main parameters with indication of their semantics.
4. The standard shall support context-based compression (i.e., depending on the sequence of data).
5. The standard shall support lossless compression.
6. The standard shall support context-based filtering with different levels of details.

**Object of standard**:

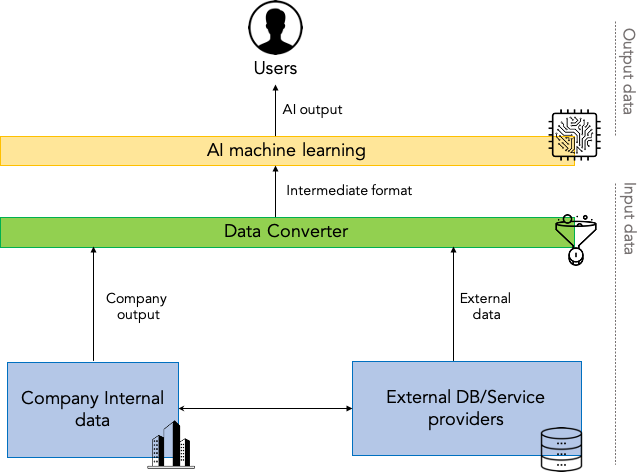
Two main areas of standardization are identified:

1. **Inputs objects**: a first set of data input related to:
   1. ***Financial data input***
      1. Financial statements and fiscal yearly reports data (usually expressed in xls or xbrl formats). Their contents follow the accounting standards defined by the Organismo Italiano Contabilità (OIC) at the Italian level and the International Accounting Standards Board (IAS/IFRS) at the international level.
      2. Invoices. In Italy, the FatturaPA format is expressed in xml, but more in general invoices have to be compliant with the European standard EN 16931-1:2017.
      3. Semantic of governance elements.
      4. Other economic data as company size uniformly recognized according to the size of employees or to the economic activities (e.g., classification elaborated by Eurostat and OECD data); imports and exports, etc…
   2. ***Vertical risks*** data input: in a preliminary phase, we will consider seismic and cyber, as vertical risks of primary interest. In the future, the object of the standard could be extended to cover other risks (e.g., related to infrastructures, sustainability). Generally, at the international level, the ISO/IEC 31000 standard defines the principles and guidelines related to the input data to consider for the risk assessment and management.
      1. *Seismic risk.* AI algorithm may help to define a socio-economic and technological model that will support companies and institution in defining reconstruction plan properly. In this direction, data input to assess seismic risk according to ASTM standards, will integrate:
      2. Technical data related to the existing/needed infrastructures (i.e., geolocation coordinates), architectural and urban planning data, as well as output data from the Building Information Modelling;
      3. Socio-demographic data, i.e., statistical data collected by certified sources (e.g., ISTAT in Italy, World Bank, International Financial Statistics, Worl Economic Outlook Databases, International Monetary Fund Statistics Data) about population figures, and their characteristics and distribution.
      4. *Cyber risks*. Considering cybersecurity-related parameters in the risk assessment will help to understand and estimate the impact of the actual risk exposure on the company performance, its financial health and business continuity.  
         As an example, an effective system to back up the sensitive data and testing periodically its effectiveness can be of help. Moreover, having well defined incident responses and a team prepared to deal with them, can help minimizing the effects of attacks and the time to recover. Therefore, an initial set of internationally recognised (ISO/IEC 27000 Information security management) inputs to consider are:
2. Data related to assessment of organizational cyber management:
   1. *organizational-level incident management* (enumerate): no, simple plans available, detailed plan + IR Team, full integrated management (e.g., with a security operations center)
   2. *backup* management: no, user requested, automatic, automatic and tested
   3. *vulnerability management* (enumerate): no, assessement, management plans, with automatic tools
   4. *enterprise patch management* (enumerate): no, manual, automatic, testing
   5. *specific cybersecurity and testing personnel* (enumerate):no cybersecurity tasks, IT personnel with cybersecurity tasks, cybersecurity-trained roles
   6. *cybersecurity procedure and mitigation testing* (enumerate): no, occasional, planned, planned&frequent
   7. *risk analysis*: no, threats identified, assessment available, some mitigations implemented, full (mitigations implemented or justified, risks quantified)
3. Data related to prevention of cyber-attacks. Being able to detect anomalies into an information system can allow preventing some attacks or discovering them before attackers, which can be estimated considering:
4. *monitoring*: no, basic detection, integrated detection, organization-level, Security Information and Event Management (SIEM) software
5. *reaction*: no, planned pre-configured responses, organized (some autom­atic), integrated tool-based & human supervised
6. Data related to training. Studies report that the personnel that has followed spec­ific training on cybersecurity aspects that are relevant for their roles are more likely to avoid errors that may compromise the information systems of their companies (e.g., use better password, less likely to click on phishing emails). Training is even more important for personnel with cybersecurity-related responsibilities, hence:
7. *awareness/training cybersecurity personnel*: no, occasional, frequent training
8. *awareness/training other categories*: no, occasional, frequent, frequent and tailored per tasks.
9. Data related to legal issues. An additional field where measuring the preparedness is the risk of losses due to legal issues (e.g., GDPR fines for
10. *availability of cybesecurity certification* (enumerate): no, certifications relevant for the company business obtained
11. *compliance to regulations* (enumerate): no, minimum, adequate
12. Data for quantifying exposure. Whenever available, the following aggregated values can help to quantify the overall exposure to cybersecurity risks:
13. *overall risk exposure*: monetization value from the risk assessment phase (number, directly obtained by the target company)
14. *mitigated risk exposure*: monetization value after the risk mitigation phase (number, directly obtained by the target company)
15. *value of assets by security requirement*: (Confidentiality-asset value/ Integrity-asset value / Availability-asset value)
16. *percentage of the value of assets-by-security-requirement on the overall value of the company assets*: (Confidentiality-asset value/ Integrity-asset value / Availability-asset value).
17. **Output objects**: represented by the outcome of the AI-based assessment in a format known by the user (format: Json)? This format is expressed in terms of a set of indexes that reflect the health of a company, the appropriateness of the governance and the impact of risks on economic-financial parameters. Some of these indexes are in response to legal requirements on business bankruptcy and crisis (e.g., DSCR), others are computed by means of a proprietary machine learning algorithm.

More in detail, the parameters in output are:

* Risk index of the likelihood of company default in a time horizon of 36 months. It reflects the company performances based on the financial data.
* Index of business continuity reflects the impacts of other risks on the previous measure.
* Index of adequacy of the organizational model considers the impact of the governance on the performance of the company, highlighting possible issues in terms of conflict of interest or familiarity.
* Debt service coverage ratio is a measurement of a firm's available cash flow to pay current debt obligations

This is depicted in *Figure 1* where the object of the standard is identified to be the intermediate format and the AI machine output.

**

*Figure 1 – MPAI-CUI model*

In some cases, internationally agreed input data formats exist. In several other cases a variety of formats exist. In these cases, meta formats to which existing formats can be converted should be defined.

The current framework has been made as general as possible taking into consideration the wide range of issues related to risk management. We are expecting the architecture to be enriched and extended according to inclusion of other risks and eventually the synergies with other MPAI applications.

Data confidentiality, privacy issues, etc. are for further consideration.

**Benefits:** MPAI-CUI will bring benefits positively affecting:

1. Technology providers need not develop full applications to put to good use their technol­ogies. They can concentrate on gradually introducing AI-based technologies that will allow a transition from traditionally approaches based on statistical methods, overcoming their limitations. This will enhance the accuracy of prediction and improve user experience.
2. Service providers (e.g., Fintech and Insurtech companies, advisors, banks) can deliver accurate products and services supporting the decision-making process, with minimising time to market as the MPAI-CUI framework enables easy combination of internal and external-party compon­ents.
3. End users as companies and local government can obtain an AI-decision support system to assess the financial health and deploy efficient strategies and action plans.
4. Processing modules can be reused for different risk management applications.

**Bottlenecks:** The full potential of AI in MPAI-CUI would be limited by a market of AI-friendly data providers and by the adoption of a vast amount of information and data strictly dependent from the company and its context.

**Social aspects**:

A simplified access to the technologies under the MPAI-CUI standard will offer end users AI-based products promising for predictions and supporting the decision making in different contexts, reducing the effort of user in analyzing data and improving its experience which becomes more personal, but including a wide vision (e.g., thought benchmarking).

Moreover, the MPAI-CUI standard and the introduction of AI-based technologies will allow a transition from the present systems which are human readable, to machine readable technologies and services.

At the national level, governments can simulate the effects of public interventions and deploy proper strategies and plans in supporting the companies and economy.

**Success criteria**:

MPAI CUI becomes the bridge between traditional approaches in compliance with the actual regulation on prediction of business crisis and fully AI-based systems.

**References**

[1] Perboli G., Arabnezhad E., A Machine Learning-based DSS for Mid and Long-Term Company Crisis Prediction. CIRRELT-2020-29. July 2020.

**Feature table:**

|  |  |
| --- | --- |
| Volume | High stream of data |
| Maturity | TRL 4 to 6 |
| AI Relevance | AI is expected to provide better performance than traditional statistical methods. In particular, AI is expected to provide previsions of the business interruption of a company years in advance |
| Users | Companies, Banks, Insurances |
| Participants | Academia/industry research groups, Fintech and Insurtech companies, Banks, Insurances |
| Interest | High |

# Other data

## Media & Entertainment

### OD.ME-01: Anomalous service access

**Proponent**: Steven Hawley

**Description**: A variable parameter that can be set to a baseline value or range, through machine learning. Once a baseline has been established, a background process can watch for values that are outside of the established range to identify rights infringement (account abuse)

1. To set the parameter: ML could learn what "typical" would actually be. For example, five active devices per account, 3 license requests for a movie, 3 log-in attempts for an account within a period of time. And these would become 'baseline' parameters for that particular video provider.
2. To enforce the parameter: Monitoring can then recognize attempts that are above these "typical" usage parameters. For example, if the limit is 3 log-in attempts in 15 minutes, and monitoring sees four attempts, no big worry. But if there are 1,275 attempts in 3 minutes, it is likely that a pirate has sold stolen account credentials, and its buyers are trying to access that stolen account.

**Comments**: The action that is taken when an alarm is raised is a separate question (revoke the keys, forced re-log-in, disable the account, etc). The values of these items are likely to be different, in different markets, regions, and cultures - hence they should be variables.

**Examples:**

**Requirements:**

**Object of standard**: The types and formats of these variables

**Benefits:**

1. Reduced abuse of pay TV and OTT accounts. Higher likelihood of detecting and interrupting incidents of piracy as they happen.

**Bottlenecks:**

**Social aspects:**

**Success criteria:**

**Feature table:**

|  |  |
| --- | --- |
| Volume |  |
| Maturity |  |
| AI Relevance |  |
| Users |  |
| Participants | Those operating a high-volume video service and wishing to detect infringing (or otherwise out-of-profile) behavior. The scale is so high that manual intervention is not realistic |
| Interest |  |

#### OD.ME-02: Personal language translator

This use case is obsolete. It has moved to MPAI-MMC Functional Requirements

**Proponent**: Danilo Pau (STM)

**Description:** Since humans increasingly live and move across the states and continents due to globalization, voice-based communications are of paramount important for better person to person interactions. There are about 6,500 languages spoken in the world however not all are equally used. Accordingly, to [[1]](#footnote-1) the top 12 most spoken languages in the world account for about 4.8 B people which represent a significant portion of humanity.

The AI distributed pipeline implied by a personal translator would look approximately as presented in figure 1

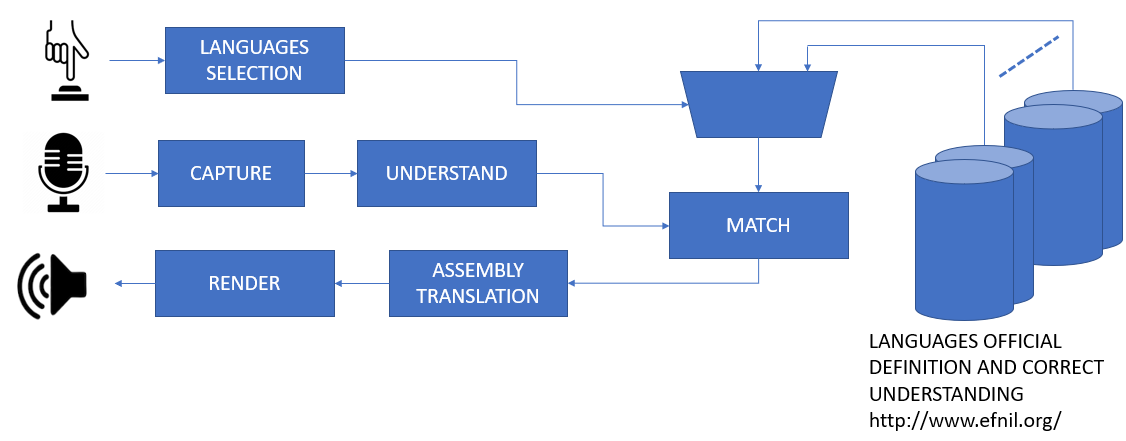


Figure 1

**Examples:**

**Requirements:**

1. Accuracy of translation

**Object of standard:**

**Benefits:**

1. Decisively improve communication between humans

**Bottlenecks:**

**Social aspects:**

**Success criteria:**

**Feature table:**

|  |  |
| --- | --- |
| Volume | Billions of devices |
| Maturity |  |
| AI Relevance | Very High |
| Users | Humanity |
| Participants | Language standardization national bodies, consumers, OEM, cloud companies, network providers |
| Interest |  |

### OD.ME-03: Multi-Modal Conversation

This use case is obsolete. It has moved to MPAI-MMC Functional Requirements

**Proponent**: Miran Choi (ETRI)

**Description:** Owing to the AI technology, natural language processing is widely used in various applications. One of the useful applications is the conversational partner which provides the user with information, entertains, chats and answers questions through the speech interface. If the application tries to provide better service to the user, it should include more than just speech interface. Emotion recognizer and gesture interpreter are among the improved multi-modal inter­faces.

The AI processing units implied by a multi-modal conversational partner would look approxim­ately as presented in figure 1

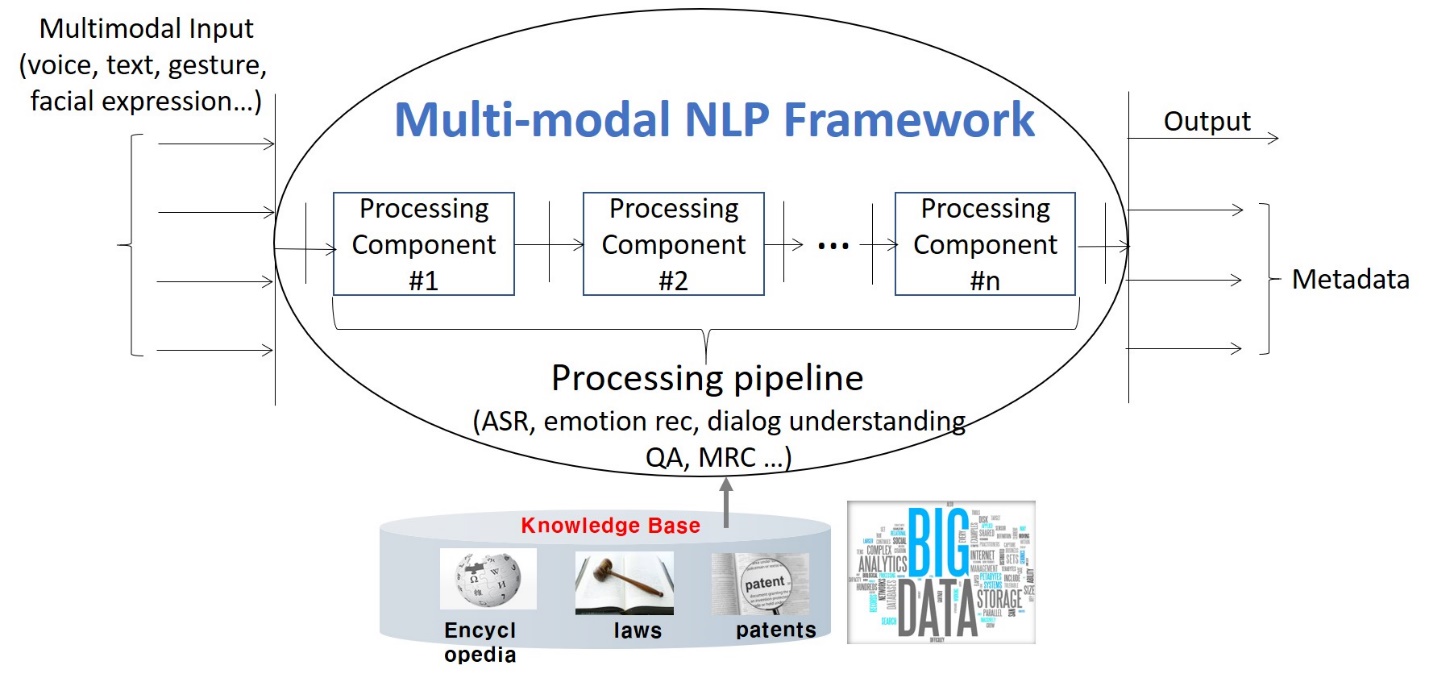


Figure 2

**Examples:**

**Requirements:**

1. Fusion of multi-modal input information
2. Natural language understanding
3. Natural language generating
4. Speech recognition
5. Speech synthesis
6. Emotion recognition
7. Intention understanding
8. Gesture recognition
9. Knowledge fusion from different sources such as speech, facial expression, gestures, etc
10. Dialog system
11. Question Answering
12. Machine Reading Comprehension (MRC)

**Object of standard:** framework for understanding human expressions by multi-modality

**Benefits:**

1. Decisively improve communication between humans and machines
2. Processing modules can be used selectively for different applications

**Bottlenecks:**

**Social aspects:**

**Success criteria:**

**Feature table:**

|  |  |
| --- | --- |
| Volume | Billions of devices |
| Maturity | Medium High |
| AI Relevance | Very High |
| Users | Humanity |
| Participants | Language standardization national bodies, consumers, OEM, cloud companies, network providers |
| Interest |  |

### OD.ME-04: Server-based Predictive Multiplayer Gaming in Next Generation Games

This use case is obsolete. It has moved to MPAI-SPG Functional Requirements

**Proponents**: Marco Mazzaglia, Andrea Basso (Synesthesia)

**Description**:

Gamers spend ludicrous amounts of cash on the latest GPUs and monitors to boost their FPS (frames per second). However, a recent study shows that higher frame rates have a noticeable impact on how well the gamer will play. A higher frame rate can improve a player’s kill-to-death ratio by as much as 90 percent. Client-side prediction is used in video games to conceal negative effects of high latency connections to keep the framerate high and the game smooth. The technique attempts to make the player's input feel more instantaneous (zero-input lag) while governing the player's actions on a remote server.

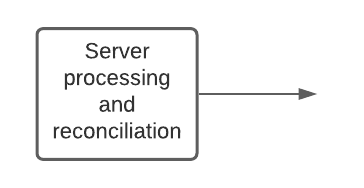
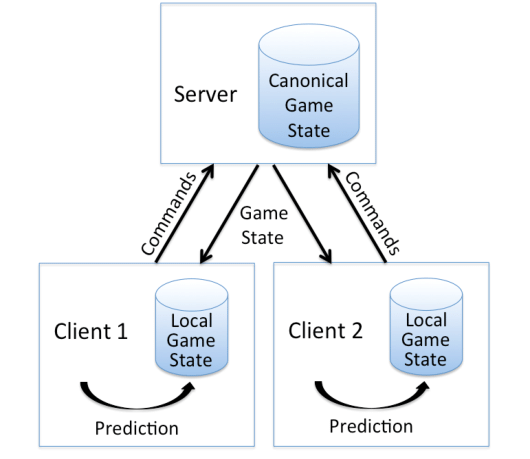


Figure 1: Client server gaming architecture with client-side prediction.

In client-side prediction the client locally reacts to user input before the server has acknowledged the input and updated the game state. So, instead of the client only sending control input to the server and waiting for an updated game state in return, the client also predicts, in parallel with this, the game state locally and gives the user feedback without awaiting an updated game state from the server. The side effect of such an approach is a potential discontinuity in the visual information that is corrected only when the information from the server is reconciled with the actual game state.

From a user input management perspective a client-server game architecture described in fig1 operates as follows:

Server

* gets inputs from all the clients, with precise timestamps
* processes inputs and updates world status
* sends regular world snapshots (i.e. 100ms updates) to all clients

Client

* sends inputs to server
* updates and presents to the gamer a local game state via client prediction models
* gets world updates from the server
* syncs predicted local state to authoritative state minimizing distortions (teleportation) and presents it to the player

From a player’s point of view, the player sees him/herself in the present and the other entities in the past which is particularly problematic for very time- and space-sensitive events.

For example with the client-server architecture of fig 1, where a fast pace shooting is taking place, a given player is aiming his weapon at the past position of the other player but the player is long gone by the time the client fires. In order to remediate this the server, once it gets the information regarding the shoot with precise timestamps, knows exactly what was the weapon’s sights and the past position of the other player body and what was the position of his body in his present. The server thus processes the shot at that past point in time, reconciles all the states and updates the clients (module on the left of figure 1). The enemy player may be the only one not entirely happy because it was shot in the past, and in the few milliseconds of difference he may have taken cover.

With the advent of cloud-based gaming, AI and ML can provide efficient solutions and different scenarios to handle latency and user experience. In particular the game rendering is handled on the cloud side and several optimizations are possible.

1. information from the game engine (inputs from clients, reconciliation info) can be used in the into the video encoding process to simplify the encoding (i.e. motion estimation)
2. AI systems can collect data from multiple clients and on that basis can perform a much better prediction of each individual move of the single participants performing complex reconciliations as described above.
3. In the same manner the representation of the audio-visual content and its coding can be optimized thanks to the prediction. This optimization will allow for higher frame rates and thus allow high quality gaming experience also on low performing hardware.

MPAI can provide several key technological components to these scenarios:

* A framework AI-powered for the prediction of the moves of a large set of players both at the server and client side.
* A user move prediction-based multi stream video encoder that can provide optimized visual experience to a large number of players.
* the associated file formats to support the framework

**Comments**:

**Examples:**

Here are a few examples of known games that illustrate how MPAI could improve feasibility and user experience

*Example 1: Racing games*

During an online racing game, players can see lags and an overall low performance on video due to network congestion or low bandwidth. Usually the information on the screen predicted by the single client is wrong if the online client information is not able to reach the server and the clients involved in the online game. In a car racing game, the player obverses at time **t0** a vehicle going straight to the wall when it is reaching a curve. After some seconds at the moment **t1** the same vehicle is “teleported” in the actual position giving an awful player experience.

The benefits of AI are able to mitigate this issue, giving to players a better game experience: the collection of the data from the different online games helps to predict to the clients a meaningful path or the correct behaviour during the missing packets time.

*Example 2: Zombie game - Left 4 Dead*

In some video games particular information is displayed differently because the physics to calculate all the outcomes from players’ action is a huge effort. An example comes from the video game “Left 4 Dead”; to show the result of killing the zombies hordes, in each client the result of the onslaught is visually different from client to client.

The benefits of a predictive input can support the online architecture to process an equal outcome on different computers. I.E. In a massive multiplayer hack&slash game, the result of the different combats among players allow to give the same live visual online experience for each player.

**Requirements:**

**AI based client prediction**

1. client input signals represented with a common set of parameters (i.e. position, speed, etc,)
2. ML components that receive these input signals and the current game state and provide updated representation of the game state (or of objects of the game state)

**AI based Server reconciliation**

client input signals with different timestamps and current game state and generate an updated reconciled game state

**Cloud gaming with thin clients**

Optimized game state video encoder that receives input from game engine and reconciliation component and generated updated game state

**Object of standard**:

1. A standard set of high-resolution time-stamped data types (e.g. define an intermediate format to provide the AI-machine with homogeneous data representation):
   1. Shooting signal from game controller
   2. ...
2. A standard format for predictions consequent to the data input: e.g.
   1. As a consequence of shooting
      1. New speed, position of bullet,
      2. recoil on the shooter,
      3. associated sound effect,
      4. effect on the scene,
      5. effect on the target
      6. motion vectors to be passed to video encoder
   2. ...
3. Possible interaction with MPAI-EVC
4. Interfaces to the delivery schemes adopted by the market (possiblly similar to the processing module defined in MPAI-CAE)
5. Possible additional level of standardisation of the AI-Machine adopting MPAI-AIF as a reference.

**Benefits:**

1. Better quality for the gameplay experience.
2. Provide an improved and detailed visual outcome for all players.
3. Enables implementation of more complex and richer games.
4. Acquired data & players’ behaviours can be used for analysis to improve the future works.

**Bottlenecks:**

Training the neural network with simulated games need further analysis (maybe follow the same approach used in music synthesis, i.e. quantization

Network congestion is still an issue; we can mitigate it, not exclude the problem at all

**Social aspects:**

Starting from the Game Industry, the system can be used in all the frameworks or the game engines for PC, console and mobile, moving more players to high quality online game experiences.

The improvement of the quality of the overall experience can benefit all the events that involve e-Sports or equivalent environments and consolidate the technology for the market.

The potential of this use case can influence common online applications that orchestrate different inputs from different users.

**Success criteria:**

Short term:improved UX on a large set of current games due to a more performant AI based architecture.

Long term: it becomes a standard used in video game engines like Unity 3D, Unreal Engine and the internal Software Development Kit (SDKs) developed by main game device manufacturers (Sony, Microsoft and Nintendo), cloud game based services (Amazon Luna, Google Stadia) and eSports platforms (FaceIt).

**Feature table:**

|  |  |
| --- | --- |
| Volume | High stream of data |
| Maturity | Research to Inception |
| AI Relevance | AI is expected to provide better predictions to allow a better UX, including a better and smoothing visual coding |
| Users | Video gaming service providers. |
| Participants | Academia/industry research groups, video gaming companies (as requirements providers) |
| Interest | High |

## Transportation

## Telco

Currently, no use case available

## Information Technology

## Aerospace

Currently, no use case available

## Manufacturing

Currently, no use case available

## Healthcare

Currently, no use case available

## Food & Beverage

Currently, no use case available

## Science & Technology

### ST.OD-01: Gravitational-Wave Interferometric Data

**Proponent**: Luigi Troiano

**Description: T**he first detection of gravitational waves, opened another important chapter in the understanding of the cosmos, known as gravitational astronomy. The LIGO (US), ViRGO (EU) and KAGRA (JP) observatories are moving towards full and operational continuity by generating observational data in the order of PB every year. Having efficient methods for compressing this type of data could significantly improve their sharing and distribution within the scientific community.

**Comments**:

**Examples:**

**Requirements:**

**Object of standard:**

**Benefits:**

**Bottlenecks:**

**Social aspects:**

**Success criteria:**

**Feature table:**

|  |  |
| --- | --- |
| Volume |  |
| Maturity |  |
| AI Relevance |  |
| Users |  |
| Participants |  |
| Interest |  |

### ST.OD-02: Multi-omics

This use case is obsolete. It has moved to MPAI-GSA Functional Requirements

**Proponent**: Paolo Ribeca

**Description:** Modern quantitative biology relies upon large datasets produced with a number of different technologies, mainly high-throughput sequencing (and derived protocols, such as RNA-sequencing, ChIP-sequencing, HiC, etc.) and mass-spectrometry. Most of those data categories needs pre-processing in order to get to a final set of numbers that represent the result of the experiment (gene expression for RNA-sequencing; signal intensity or peak location for ChIP-sequencing; connections between different point of the genome for HiC; mass-spectrometry peaks; and so on) – for instance, the representation of primary sequencing data is covered by MPEG-G parts 1-5, and the representation of the final results of analysis by MPEG-G part 6. Once pre-processing has been performed, one is left with results in a high-number of dimensions (one value for gene expression for each gene in the case of RNA-sequencing or proteomics; one value for each position in the genome for experiments producing an intensity; couples of positions in the genome for HiC; a number of clinical/phenotypic scores possibly derived from observation of the subject) for each experimental condition (identified by a biological sample). A full experiment is typically made of a number of samples, thus originating a matrix having a very large number of dimensions and as many lines as the number of experimental conditions. Such matrices are increasingly often analysed with AI techniques, in particular when integration of hetero­geneous data is required. It would be very useful to facilitate such analysis by providing a compressed format that is able to interface with AI computational libraries, such as TensorFlow, at the same time making the handling of the biological details (metadata, partial selection of experimental dimensions and conditions) transparent.

**Comments**:

**Examples:**

**Requirements:**

**Object of standard:** Compression of multi-dimensional data derived from heterogeneous quantitative experiments in biology, and the related meta-data. Access to data through AI-based libraries.

**Benefits:** There is a large infrastructural gap in the field, as no format for such kind of integrative studies exists at the moment. An MPAI format would allow homogeneous and transparent downstream access to data for all -omics studies.

**Bottlenecks:**

**Social aspects:**

**Success criteria:**

**Feature table:**

|  |  |
| --- | --- |
| Volume | From tens of GB to a few TB/experiment. Hundreds of PB/year estimated globally. |
| Maturity | Upstream analysis is mature. Integrative analysis of different datasets is not yet standardised, but a number of algorithms exist. The proposal would be algorithm-agnostic though. |
| AI Relevance | The proposal would facilitate AI-centric data analysis – so, central. |
| Users | All researchers and companies relying upon quantitative -omics data and their interpretation based on AI (includes providers of genetic tests/therapies). |
| Participants | An increasing number worldwide. Genomic is one of the fastest growing scientific disciplines at the moment, with a very large and increasing budget. |
| Interest | High. |

### ST.OD-03: Spatial metabolomics

This use case is obsolete. It has moved to MPAI-GSA Functional Requirements

**Proponent**: Paolo Ribeca

**Description:** Modern quantitative biology relies upon large datasets produced with a number of different technologies, mainly high-throughput sequencing (and derived protocols, such as RNA-sequencing, ChIP-sequencing, HiC, etc.) and mass-spectrometry. One of the most data-intensive protocols is spatial proteomics, whereby in-situ MS/metabolomics techniques are performed on “pixels”/”voxels” of a 2D/3D sample in order to obtain proteomics data at different locations in the sample, typically with sub-cellular resolution. This information can also be correlated with pictures/tomograms of the sample, in order to obtain phenotypical information about the nature of the pixel/voxel. The combined results (unprocessed/processed spectrum, and corresponding image data) are typically analysed with AI-based technique. It would be very useful to facilitate such analysis by providing a compressed format that is able to interface with AI computational libraries, such as TensorFlow, at the same time making the handling of the biological details (metadata, pixel/voxel segmentation and classification, feature identification, MS data analysis, correlation of metabolic data within neighbouring pixels/voxels, partial selection of experimental dimensions) transparent.

**Comments**:

**Examples:**

**Requirements:**

**Object of standard**: Compression of data (metabolic and images) produced by spatial metabolomics in biology, and the related meta-data. Access to data through AI-based libraries.

**Benefits:** There is a large infrastructural gap in the field, as no format for integrated studies of metabolic and image data exists at the moment. An MPAI format would allow homogeneous and transparent analysis for such data.

**Bottlenecks:**

**Social aspects:**

**Success criteria:**

**Feature table:**

|  |  |
| --- | --- |
| Volume | Typically 100 GB/experiment. A growing number of experiments worldwide. |
| Maturity | Integrative analysis of metabolic data/images is not yet standardised, but a number of algorithms exist. The proposal would be algorithm-agnostic though. |
| AI Relevance | The proposal would facilitate AI-centric data analysis – so, central. |
| Users | All researchers and companies relying upon spatial metabolomics (for instance, pharma companies studying the effect of drugs on organoids). The technique is likely to play an increasingly important role in cancer treatment in coming years. |
| Participants | An increasing number worldwide. Spatial metabolomics is one of the disciplines whose data footprint is increasing fastest. |
| Interest | High. |

### ST.OD-03: Integrated -omics and image data

This use case is obsolete. It has moved to MPAI-GSA Functional Requirements

**Proponent**: Paolo Ribeca

**Description:** During the past few years there has been an increasing interest in data-rich techniques to optimise livestock and crop production (so called “smart farming”). The range of techniques is constantly expanding, but the main ideas are to combine molecular techniques (mainly high-throughput sequencing and derived protocols, such as RNA-sequencing, ChIP-sequencing, HiC, etc.; and mass-spectrometry) and monitoring by images (growth rate under different conditions, sensor data, satellite-based imaging) for both livestock species and crops. Similar to use case 6.9.2, many of the molecular data categories needs pre-processing in order to get to a final set of numbers that represent the result of the experiment; the same happens to sensor and especially imaging data, where meaningful information (time-dependent behaviour, time-dependent phenotypic traits) need to be extracted from the raw data. Once pre-processing has been performed, one is left with results in a high number of dimensions (from both the molecular and the sensor/imaging data) for each experimental condition (identified by a biological sample). A full experiment is typically made of a number of samples, thus originating a matrix having a very large number of dimensions and as many lines as the number of experimental conditions. Such matrices are increasingly often analysed with AI techniques, in particular when integration of heterogeneous data is required. It would be very useful to facilitate such analysis by providing a compressed format that is able to interface with AI computational libraries, such as TensorFlow, at the same time making the handling of the biological details (metadata, partial selection of experimental dimensions and conditions) transparent.

**Comments**:

**Examples:**

**Requirements:**

**Object of standard:** Compression of multi-dimensional data derived from “smart farming” techniques that combine molecular and sensor/imaging data, and the related meta-data. Access to data through AI-based libraries.

**Benefits:** There is a large infrastructural gap in the field, as no format for such kind of integrative studies exists at the moment. An MPAI format would allow homogeneous and transparent downstream access to data for all studies.

**Bottlenecks:**

**Social aspects:**

**Success criteria:**

**Feature table:**

|  |  |
| --- | --- |
| Volume | From tens of GB to a few TB/experiment. An increasing number of experiments globally (typically connected with government agencies). |
| Maturity | Upstream analysis is mature. Integrative analysis of different datasets is not yet standardised, but a number of algorithms exist. The proposal would be algorithm-agnostic though. |
| AI Relevance | The proposal would facilitate AI-centric data analysis – so, central. |
| Users | All researchers and companies relying upon quantitative “smart farming” data and their interpretation based on AI (includes a number of companies and government agencies). |
| Participants | An increasing number worldwide. Smart farming has been identified by several governments as one of the future priorities, with an increasing budget. |
| Interest | High. |

### ST.OD-06: Integrative AI-based analysis of multi-source genomic/sensor experiments

This use case is obsolete. It has moved to MPAI-GSA Functional Requirements

**Proponents**: Paolo Ribeca (BioSS/James Hutton)

**Description**: Most experiment in quantitative genomics consist of a setup whereby a small amount of metadata – observable clinical score or outcome, desirable traits, observed behavior – is correlated with, or modeled from, a set of data-rich sources. Such sources can be:

1. Biological experiments – typically sequencing or proteomics/metabolomics data
2. Sensor data – coming from images, movement trackers, etc.

All those data-rich sources share the following properties:

1. They produce very large amounts of “primary” data as output
2. They need “primary”, experiment-dependent, analysis, in order to project the primary data (1) onto a single point in a “secondary” space with a high dimensionality – typically a vector of thousands of values
3. The resulting vectors, one for each experiment, are then fed to some machine or statistical learning framework, which correlates such high-dimensional data with the low-dimensional metadata available for the experiment. The typical purpose is to either model the high-dimensional data in order to produce a mechanistic explanation for the metadata, or to produce a predictor for the metadata out of the high-dimensional data.
4. Although that is not typically necessary, in some circumstances it might be useful for the statistical or machine learning algorithm to be able to go back to the primary data (1), in order to extract more detailed information than what is available as a summary in the processed high-dimensional vectors produced in (2).

Providing a uniform framework to:

1. Represent the results of such complex, data-rich, experiments, and
2. Specify the way the input data is processed by the statistical or machine learning stage

would be extremely beneficial.

**Comments**: Although this structure above is common to a number of experimental setups, it is conceptual and never made explicit. Each “primary” data source can consist of heterogeneous information represented in a variety of formats, especially when genomics experiments are considered, and the same source of information is usually represented in different ways depending on the analysis stage – primary or secondary. That results in data processing workflows that are ad-hoc – two experiments combining different sets of sources will require two different workflows able to process each one a specific combination of input/output formats. Typically, such workflows will also be layered out as a sequence of completely separated stages of analysis, which makes it very difficult for the machine or statistical learning stage to go back to primary data when that would be necessary.

MPAI-GSA aims to create an explicit, general and reusable framework to express as many different types of complex integrative experiments as possible. That would provide (I) a compressed, optimized and space-efficient way of storing large integrative experiments, but also (II) the possibility of specifying the AI-based analysis of such data (and, possibly, primary analysis too) in terms of a sequence of pre-defined standardized algorithms. Such computational blocks might be partly general and prior-art (such as standard statistical algorithms to perform dimensional reduction) and partly novel and problem-oriented, possibly provided by commercial partners. That would create an healthy arena whereby free and commercial methods could be combined in a number of application-specific “processing apps”, thus generating a market and fostering innovation. A large number of actors would ultimately benefit from the MPAI-GSA standard – researchers performing complex experiments, companies providing medical and commercial services based on data-rich quantitative technologies, and the final users who would use instances of the computational framework as deployed “apps”.

**Examples**

The following examples describe typical uses of the MPAI-GSA framework.

1. *Medical genomics – sequencing and variant-calling workflows*

In this use case, one would like to correlate a list of genomic variants present in humans and having a known effect on health (metadata) with the variants present in a specific individual (secondary data). Such variants are derived from sequencing data for the individual (primary data) on which some variant calling workflow has been applied. Notably, there is an increasing number of companies doing just that as their core business. Their products differ by: the choice of the primary processing workflow (how to call variants from the sequencing data for the individual); the choice of the machine learning analysis (how to establish the clinical importance of the variants found); and the choice of metadata (which databases of variants with known clinical effect to use). It would be easy to re-deploy their workflows as MPAI-GSA applications.

1. *Integrative analysis of ‘omics datasets*

In this use case, one would like to correlate some macroscopic variable observed during a biological process (e.g. the reaction to a drug or a vaccine – metadata) with changes in tens of thousands of cell markers (gene expression estimated from RNA; amount of proteins present in the cell – secondary data) measured through a combination of different high-throughput quantitative biological experiments (primary data – for instance, RNA-sequencing, ChIP-sequencing, mass spectrometry). This is a typical application in research environments (medical, veterinary and agricultural). Both primary and secondary analysis are performed with a variety of methods depending on the institution and the provider of bioinformatics services. Reformulating such methods in terms of MPAI-GSA would help reproducibility and standardisation immensely. It would also provide researchers with a compact way to store their heterogeneous data.

1. *Single-cell RNA-sequencing*

Similar to the previous one, but in this case at least one of the primary data sources is RNA-sequencing performed at the same time on a number (typically hundred of thousands) of different cells – while bulk RNA sequencing mixes together RNAs coming from several thousands of different cells, in single-cell RNA sequencing the RNAs coming from each different cell are separately barcoded, and hence distinguishable. The DNA barcodes for each cell would be metadata here. Cells can then be clustered together according to the expression patterns present in the secondary data (vectors of expression values for all the species of RNA present in the cell) and, if sufficient metadata is present, clusters of expression patterns can be associated with different types/lineages of cells – the technique is typically used to study tissue differentiation. A number of complex algorithms exist to perform primary analysis (statistical uncertainty in single-cell RNA-sequencing is much bigger than in bulk RNA-sequencing) and, in particular, secondary AI-based clustering/analysis. Again, expressing those algorithms in terms of MPAI-GSA would make them much easier to describe and much more comparable. External commercial providers might provide researchers with clever modules to do all or part of the machine learning analysis.

1. *Experiments correlating genomics with animal behaviour*

In this use case, one wants to correlate animal behaviour (typically of lab mice) with their genetic profile (case of knock-down mice) or the previous administration of drugs (typically encountered in neurobiology). Hence primary data would be video data from cameras tracking the animal; secondary data would be processed video data in the form of primitives describing the animal’s movement, well-being, activity, weight, etc.; and metadata would be a description of the genetic background of the animal (for instance, the name of the gene which has been deactivated) or a timeline with the list and amount of drugs which have been administered to the animal. Again, there are several companies providing software tools to perform some or all of such analysis tasks – they might be easily reformulated in terms of MPAI-GSA applications.

1. *Spatial metabolomics*

One of the most data-intensive biological protocols nowadays is spatial proteomics, whereby in-situ mass-spec/metabolomics techniques are applied to “pixels”/”voxels” of a 2D/3D biological sample in order to obtain proteomics data at different locations in the sample, typically with sub-cellular resolution. This information can also be correlated with pictures/tomograms of the sample, to obtain phenotypical information about the nature of the pixel/voxel. The combined results are typically analysed with AI-based technique. So primary data would be unprocessed metabolomics data and images, secondary data would be processed metabolomics data and cellular features extracted from the images, and metadata would be information about the sample (source, original placement within the body, etc.). Currently the processing of spatial metabolomics data is done through complex pipelines, typically in the cloud – having these as MPAI-GSA applications would be beneficial to both the researchers and potential providers of computing services.

1. *Smart farming*

During the past few years, there has been an increasing interest in data-rich techniques to optimise livestock and crop production (so called “smart farming”). The range of techniques is constantly expanding, but the main ideas are to combine molecular techniques (mainly high-throughput sequencing and derived protocols, such as RNA-sequencing, ChIP-sequencing, HiC, etc.; and mass-spectrometry – as per the ‘omics case at point 2) and monitoring by images (growth rate under different conditions, sensor data, satellite-based imaging) for both livestock species and crops. So this use case can be seen as a combination of cases 2 and 4. Primary sources would be genomic data and images; secondary data would be vectors of values for a number of genomic tags and features (growth rate, weight, height) extracted from images; metadata would be information about environmental conditions, spatial position, etc. A growing number of companies are offering services in this area – again, having the possibility of deploying them as MPAI-GSA applications would open up a large arena where academic or commercial providers would be able to meet the needs of a number of customers in a well-defined way.

**Requirements:**

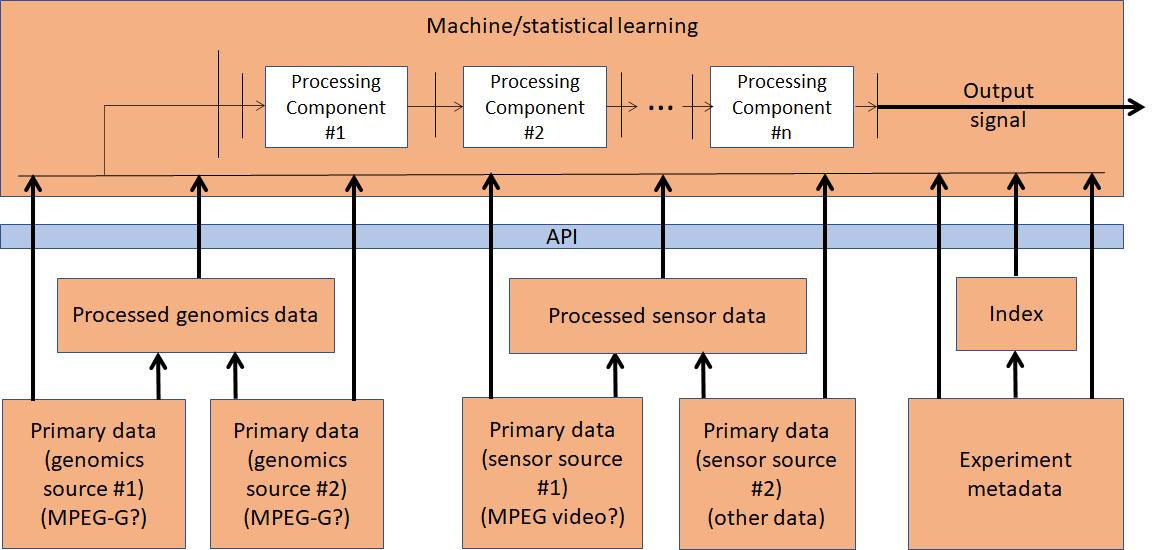
MPAI-GSA should provide support for the storage of, and access to:

* Unprocessed genomic data from the most common sources (reference sequences, short and long sequencing reads)
* Processed genomic data in the form of annotations (genomic models, variants, signal tracks, expression data). Such annotations can be produced as the result of primary analysis of the unprocessed data or come from external sources
* Video data, both unprocessed and processed (extracted features, location, movement tracking data)
* Sensor data, both unprocessed (GPS position tracking, series describing temperature/humidity/weather, general input of multi-channel sensors) and processed.
* Experiment meta-data (such as collection date and place; classification in terms of a number of user-selected categories for which a discrete or continuous value is available)
* Support for the semantic description of the ontology of all the considered sources.

MPAI-GSA should also provide support for:

* The combination into a general analysis workflow of a number of computational blocks that access processed, and possibly unprocessed, data as input channels, and produce output as a sequence of vectors in a space of arbitrary dimension. Combination would be done in terms of nodes (processing blocks and adaptors [blocks that return as output a subset of the input channels, nodes that replicate the output several times]), and a connection graph
* The possibility of defining and implementing a novel processing block from scratch in terms of either some source code or a proprietary binary codec
* A number of pre-defined blocks that implement well-known analysis methods (such as PCA, MDS, CA, SVD, NN-based methods, etc.).

**Object of standard**: A high-level, schematic description of the standard can be found in the following figure:



Currently, three areas of standardization are identified:

1. **Interface to define the machine/statistical learning in terms of basic algorithmic blocks:**
   1. Ability to define basic computational blocks that take a (sub)set of inputs and produce a corresponding set of outputs (the cardinalities of the sets can be different). More in detail:
      1. The way to define a block will be mandated by the standard (in particular, a unique ID might be issued to the implementer by a centralised authority)
      2. A basic computational block will need to define and export its computational requirements in terms of CPU and RAM
      3. Some blocks might specify their preferred execution model, for instance some specific cloud computation platform
      4. Some pre-defined blocks might be provided by the standard, for instance the implementation of a few well-known methods to perform dimensional reduction (PCA, MDS, etc.).
   2. Ability to create processing pipelines in terms of basic algorithmic blocks and to define the associated control flow required to perform the full computation
   3. A standardised output format
   4. Whenever possible, interoperability with established technologies (e.g. TensorFlow).
2. **Interface for the machine/statistical learning to access processed (secondary) data**: A first set of input and output signals, with corresponding syntax and semantics, to process secondary data (i.e., the results of processing primary data) with methods based on machine learning. They have the following shared properties:
   1. Irrespective of their source (genomic or sensor) all the inputs to the AI processor are expressed as vectors of values of different number types
   2. Information about the semantics of each input (which source produced which input, and which biological entity/feature each input corresponds to) is standardised and provided through metadata
   3. In order to provide fast access and search capabilities, an index for the metadata is provided.
3. **Interface for the machine/statistical learning to access unprocessed (primary) data** with the following features:
   1. Definition of the semantics of the primary data accessible for a large number of data types (especially genomic – see ISO/IEC 23092 part 6 for an indicative list)
   2. Information on the methods used to process the primary data into secondary with a clearly defined semantics is standardised and provided through metadata
   3. Possibly, ability to define basic computational blocks for primary analysis similar to those at (2) and their combination into complex computations, in order to re-process primary data whenever needed by (2)
   4. The processing pipeline may be a combination of local and in-cloud processing.

On the other hand, we would not like to re-standardise the representation of primary data! For instance, part 6 of the the MPEG-G format (ISO/IEC 23092) already standardizes meta-data for most of the techniques employed in the field of genomics under the unifying concept of genomic annotations. Part 3 already offers a clear API through which primary sequencing data can be accessed. In order to avoid effort replication, MPAI-GSA might represent genomic data as MPEG-G encoded data. Similar possibilities might be considered for video primary sources.

**Benefits:** MPAI-GSA will be offer a number of advantages to a number of actors:

1. Technology providers and researchers will have available a robust, tested framework within which to develop modular applications that are easy to define, deploy and modify. They will no longer need to spend time and resources on implementing access to a number of basic data formats or the mechanics of a unified access to heterogeneous resources – they will be able to focus fully on the development of their machine learning methods. The methods will be clearly defined in terms of computational modules, some of which can be provided by commercial third parties. This will drastically improve reproducibility, which is an increasing problem with the current biological research based on big data. Offering a robust, well-defined framework will also lower the amount of resources needed to enter the market and help smaller actors to generate and implement competitive ideas
2. Equipment manufacturers and application vendors can choose from the set of technologies made available through MPAI-GSA standard by different sources, integrate them and satisfy their specific needs
3. Service providers will have available a growing number of MPAI-GSA applications able to solve different categories of data analysis problems. As all such applications are clearly expressed in terms of a reproducible standard rather than being developed and hosted opaquely on some closed corporate computing infrastructure, comparing the different applications and offering different options to customers will become much simpler
4. End users will enjoy a thriving, competitive market that provides an ever-growing number of quality solutions by an ever-growing number of actors. The general availability of such a powerful technology will hopefully make widespread applications that today require research computational equipment and personnel, such as clinically oriented genetic/genomic analysis.

**Bottlenecks:** In order to fully exploit the potential of MPAI-GSA, one will need widespread availability of computing power, in particular for applications comprising steps whereby primary data is processed into secondary. That would typically require the ability to perform some of such computational steps in the cloud, as few users have access to enough resources. AI-friendly processing units able to implement and speed up secondary analysis would also help, perhaps allowing MPAI-GSA applications based only on secondary-analysis on commodity devices such as mobile phones.

**Social aspects**: Genomic applications partially based on the phone might facilitate social uses of the technology (such as receiving and exploring the results of genetic tests, or establishing genealogies).

**Success criteria**: Data-rich applications are the future of a number of disciplines, in particular life sciences, personalised medicine and the one-health approach – whereby humans, livestock, farming and ultimately the whole terrestrial ecosystem is seen as an integrated system, with each part influencing the rest. At the moment, however, creating analysis workflows able to exploit the data is a painstaking ad-hoc process which requires sizeable investments in technology and development. Most of the times the effort cannot be reused, as by definition the applications developed are problem-specific. MPAI-GSA will be successful if it can create a framework facilitating the development of modular, reusable analysis frameworks that on one hand trivialise data storage and on the other hand streamline the creation of complex methods. Its success will be defined by its ability to attract a number of actors – researchers, commercial providers of computational solutions and analytical services, end users. The thriving ecosystem of applications thus generated will be a necessary ingredient to transparently integrate data-rich technologies for the life sciences into common practice, widespread appliances, and every day’s life.

## Other Domains

### OD.OD-02: Machine vibrations

**Proponent**: Danilo Pau (STM)

**Description**: An AI inertial sensor (accelerometer with gyroscope) must learn to recognise unsupervised if the vibrations of a bearing are normal or abnormal to learn what anomalies are encountered in its life

**Comments:**

**Examples:**

**Requirements:**

**Object of standard:**

**Benefits:**

**Bottlenecks:**

**Social aspects:**

**Success criteria:**

**Feature table:**

|  |  |
| --- | --- |
| Volume |  |
| Maturity |  |
| AI Relevance |  |
| Users |  |
| Participants |  |
| Interest |  |

### OD.OD-03: Building vibrations

**Proponent**: Danilo Pau (STM)

**Description**: An inertial sensor community (accelerometer with gyroscope) must learn to recognise unsupervised if the vibrations in the house are due to the passage of the metro, an earthquake, a structural failure.

**Comments**: How does the AI inertial sensor pair with high quality seismographs? How do they talk to the databases of the national institutes? How do we prevent a sensor from giving false negatives?

**Examples:**

**Requirements:**

**Object of standard:**

**Benefits:**

**Bottlenecks:**

**Social aspects:**

**Success criteria:**

**Feature table:**

|  |  |
| --- | --- |
| Volume |  |
| Maturity |  |
| AI Relevance |  |
| Users |  |
| Participants |  |
| Interest |  |

1. <https://blog.busuu.com/most-spoken-languages-in-the-world/#:~:text=Well%2C%20roughly%206%2C500%20languages%20are%20spoken%20in%20the%20world%20today.> [↑](#footnote-ref-1)