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

|  |  |
| --- | --- |
|  | 2021/01/12 |
| Source | Valeria Lazzaroli, Guido Perboli, Mariangela Rosano |
| Title | Proposal for MPAI-CUI Use Cases and Functional Requirements |
| Target | MPAI Members |

# Introduction

Moving Picture, Audio and Data Coding by Artificial Intelligence (MPAI) is an [international association](http://mpai.community/) with the mission to develop *AI-enabled data coding standards*. Research has shown that data coding with AI-based technologies is *more efficient* than with existing technologies.

This document is a Call for Technologies for the MPAI Compression and understanding of industrial data (MPAI-CUI) project. The MPAI-CUI standard uses AI substantially to reduce the amount of data with a controlled loss of information and extract the most relevant information from the industrial data, with the aim of assessing company performance and predicting the risk of bankruptcy long before.

The MPAI approach to developing AI data coding standards is based on the definition of *standard interfaces* of *AI Modules (AIM).* AIMs operate on input data having a standard format to provide output data having a standard format. AIMs can be combined and *executed* in an MPAI-specified *AI-Framework* called MPAI-AIF. A [Call for MPAI-AIF Technologies](https://mpai.community/standards/mpai-aif/) is currently open.

While AIMs must expose standard interfaces to be able operate in an MPAI AI Framework, their performance may differ depending on the technologies used by implementors. MPAI believes that *competing* developers striving to provide more performing *proprietary* and *interoperable* AIMs will promote *horizontal markets* of *AI solutions* that tap from and further promote AI *innovation*.

This document calls for technologies that can specifically be used to develop specifications of input and output interfaces AIMs whose assembly provides a solution to the identified MPAI-CUI Use Case:

1. Data Compression and Understanding

It should be noted the Use Case that makes up MPAI-CUI will obviously be non-normative. The internals of the AIMs will also be non-normative. The input and output interfaces of the AIMs whose requirements have been derived to support the Use Cases will be normative.

Therefore, the scope of this Call for Technologies is restricted to the input and output interfaces of the AIMs. However, MPAI invites comments on any element of this Call for Technologies.

The content of this document is

|  |  |
| --- | --- |
| Chapter 2 | briefly introduces the AI Framework Reference Model and its six Components |
| Chapter 3 | briefly introduces the Use Case. |
| Chapter 4 | presents the MPAI-CUI Use Case with the following structure1. Reference architecture
2. Description of AI Modules and their I/O data
3. Technologies and Functional Requirements
4. Interfaces of AIM I/O Data
 |
| Chapter 5 | identifies the technologies likely to be common across MPAI-CUI and other MPAI use cases |
| Chapter 6 | gives suggested references. Respondents are advised to become familiar with the references |
| Chapter 7 | gives a basic list of relevant terms and their definition |

# The MPAI AI Framework (MPAI-AIF)

Most MPAI applications considered so far can be implemented as a set of AIMs – AI/ML and even traditional data processing-based units with standard interfaces assembled in suitable topol­ogies to achieve the specific goal of an application and executed in an MPAI-defined AI Frame­work. MPAI is making all efforts to iden­tify processing modules that are re-usable and upgradable without necessarily changing the inside logic.

MPAI plans on completing the development of a 1st generation AI Framework called MPAI-AIF in July 2021.

The MPAI-AIF Architecture is given by

*Figure 1*.



*Figure 1 – The MPAI-AIF Architecture*

Where

1. *Management and Control* manages and controls the AIMs, so that they execute in the correct order and at the time when they are needed.
2. *Execution* is the environment in which combinations of AIMs operate. It receives external inputs and produces the requested outputs both of which are application specific interfacing with Management and Control and with Communication, Storage and Access.
3. *AI Modules* (AIM) are the basic processing elements receiving processing specific inputs and producing processing specific
4. *Communication* is required in several cases and can be implemented, e.g., by means of a service bus and may be used to connect with remote parts of the framework
5. *Storage* encompasses traditional storage and is used to e.g., store the inputs and outputs of the individual AIMs, data from the AIM’s state and intermediary results, shared data among AIMs.
6. *Access* represents the access to static or slowly changing data that are required by the application such as domain knowledge data, data models, etc.

# Use Cases

## Data Compression and Understanding

# A company may need to access the flow of internal (i.e., financial and governance data) and exter­nal data to assess and mon­itor its financial and organizational performance, as well as the impact of vertical risks (e.g., cyber, seismic, etc.), according to the current regulations (e.g., ISO 31000 on risk assessment and management).

# 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). Indeed, the company itself 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.

# At the same time, 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.

# Functional Requirements

## Data Compression and Understanding

### Reference architecture

This Use Case can be implemented as in *Figure 2*.



*Figure 2 – Compression and understanding of Industrial Data*

### AI Modules and their I/O data

The AI Modules of *Figure 2* perform the functions described in *Table 1 – AI Modules* .

*Table 1 – AI Modules of Industrial Data Compression and Understanding*

|  |  |
| --- | --- |
| **AIM** | **Function** |
| **Data Conversion (DC)** | Gather data needed for the assessment from several sources (internal and external), in different format and covert it in a unique format json. |
| **Financial assessment (FA)**  | To analyse the data generated by the companies (i.e., financial statements) to assess the preliminary financial performances in the form of indexes. To build and extract the financial features for the machine learner. |
| **Governance assessment (GA)** | To build and extract the features related to the adequacy of the governance asset for the machine learner. |
| **Risk matrix (BMR)** | To build the risk matrix to assess the impact of vertical risks (i.e., in this use case cyber and seismic). |
| **Decision tree (DT)** | To create the decision trees for making decisions according to the Random Forest algorithm. |
| **Prediction (PRF)** | To predict values of the likelihood of company default in a time horizon of 36 months and of the adequacy of the organizational model.  |
| **Perturbation (PBC)** | To perturb the value of the probability of company crisis computed before, considering the impact of vertical risks on company performances |

### I/O interfaces of AI Modules

The I/O data of Data Compression and Understanding AIMs are given in *Table 2 – I/O data of Use Case AIMs*.

*Table 2 – I/O data of Use Case AIMs*

|  |  |  |  |
| --- | --- | --- | --- |
| **AI Module** | **Input** | **Output** | **External data** |
| **Data Conversion (DC)** | Financial statement dataGovernance dataRisk assessment data | Financial statement data (converted)Governance data (converted) |  |
| **Financial assessment (FA)**  | Financial statement data | Financial features | Standards from knowledgebase |
| **Governance assessment (GA)** | Governance data | Governance features |  |
| **Risk matrix (BMR)** | Technical data from BIM, internal assessment on cyber security | Severity  | Socio-economic data from data bases Technical data from KBStandards from KB |
| **Decision tree (DT)** | Financial features, Governance features | Ranking of features importance  | Data on active and failed companies from back testing |
| **Prediction (PRF)** | Financial features, Governance features | Probability of company crisisAdequacy of organizational model (indexes) | Data on active and failed companies from back testing |
| **Perturbation (PBC)** | Probability of company crisis (index); severity from BMR | Index of business continuity | Expert-based mapping rules |

### Technologies and Functional Requirements

#### Financial statement data

The Financial statement (raw data) are produced based on a set of accounting principles driving maintenance and reporting of company accounts, so that financial statements can be consistent, transparent, and comparable across companies.

A first set of principles, identified by International Accounting Standard/International Financial Reporting Standard (IAS/IFRS), can be taken as “universal”, as common recognized across all countries are:

1. ...
2. ...
3. ...

An example of corresponding digital representations is

1. ...
2. ...
3. ...

A second set of principles (Principle B) are typically jurisdiction dependent. In the case of Europe example principles are

1. ...
2. ...
3. ...

Repondents are requested to propose

1. A set of Principles A
2. The corresponding digital representation
3. One or more sets of Principles B where applicable jurisdictions are indentified
4. The corresponding digital representation

The Financial statement (raw data) are converted to a standard format by the Data conversion (DC).

Respondent are invited to propose a digital representation of financial statement data that are applicable to a minimum set of financial statements whose semantics of universal and local validity. JSON is a primary example of digital representation. However, other representations are possible.

Proponents are invited to comment on this choice and possibly suggest alternative formats. Preference will be given to formats that have been standardised or are in wide use.

#### Governance data

By Governance data we mean attributes that indicate the structure of the governance structure of a company and the roles of key personnel.

The most basic roles are shareholder, manager, sole administrator, president/member of the board of directors, auditor, president/member of the statutory board of directors. They can be taken as “universal”, as common recognized across all countries.

Respondent are invited to propose a governance data ontology that captures today’s practice at the global level. How can the data from a specific company be expressed starting from the ontology?

#### Risk assessment technical data

By Risk assessment technical data, we mean attributes that indicate the internal assessment that the company performs to identify and measure potential or existing vertical risks, and their impact on business continuity.

This data contains values of likelihood, impact, gravity, residual risk and treatments. They should be encoded according to ISO 31000 – “Risk management -- Principles and guidelines”.

Proponents are invited to comment on this choice.

Respondent are invited to propose?

#### Financial features

Financial features are a set of indexes and ratios computed using financial statement data. Examples of financial features are given by *Table 3*.

*Table 3 – Financial features*

|  |  |  |
| --- | --- | --- |
| **Feature** | **Feature value** | **Feature type** |
| **1** | Absolute value | Revenue/Profit |
| **2** | Index/Percentage (%) | Revenue/Profit |
| **3** | Absolute value | Revenue/Profit |
| **4** | Absolute value | Revenue/Profit |
| **5** | Index/Percentage (%) | Revenue/Profit |
| **6** | Index/Percentage (%) | Cost/Debt |
| **7** | Absolute value | Cost/Debt |
| **8** | Index/Percentage (%) | Cost/Debt |
| **9** | Absolute value | Cost/Debt |
| **10** | Index/Percentage (%) | Cost/Debt |
| **11** | Absolute value | Production |
| **12** | Absolute value | Production |
| **13** | Index/Percentage (%) | Revenue/Profit |
| **14** | Absolute value | Production |
| **15** | Index/Percentage (%) | Cost/Debt |

Respondents are requested to propose Financial features suitable for financial assessment, e.g., those reported in *Table 3*. Financial features shall satisfy the following requirements

1. Extracted or computed from the financial statement data
2. …
3. …

#### Governance features

Governance features are a set of indexes/ parameters that are used to assess the adequacy of the organizational model. Examples are given by *Table 4*.

*Table 4 – Governance features*

|  |  |  |
| --- | --- | --- |
| **Feature** | **Feature value** | **Feature type** |
| **1** | Absolute value | Decision maker data |
| **2** | Index/Percentage (%) | Shareholder data |
| **3** | Absolute value | Shareholder data |
| **4** | Absolute value | Decision maker data |
| **5** | Absolute value | Decision maker data |

Respondents are requested to propose Governance features suitable for assessing the suitability of governance, e.g., those reported in *Table 4*. Governance features shall satisfy the following requirements

1. Extracted or computed from the Governance data

2. Numerical values

3. …

#### Severity

A set of values, each of them reflects the level of risk for that specific vertical risk evaluated by the company.

Respondents are requested to propose…

#### Decision Tree

It is a decision support tool that uses a tree-like model of decision, given the financial and governance features. It is based on the Random forest supervised learning method to predicts the value of the probability of company crisis and bankruptcy.

Respondent are requested to propose other learning methods satisfying the following requirements:

1. …
2. …

#### Expert-based mapping rules KB query

Expert-based mapping rules KB contains a set of rules established by experts in the field that express the vertical risk-financial feature assignment, i.e., an expression of the impact of a certain vertical risks on the financial performance of a company.

Respondent are requested to propose:

1. According to the new financial features proposed, the list of risks that potentially affect these features.
2. A digital representation of these mapping rules.

# References

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

# Terminology

|  |  |
| --- | --- |
| **Term** | **Definition** |
| Access |  |
| AI Module |  |
|  |  |
|  |  |
|  |  |
|  |  |